CHAPTER 5

BIOLOGICAL NETWORKS

A discussion of various networks of interest in biology, including biochemical networks, neural networks, and ecological networks

NETWORKS are widely used in many branches of biology as a convenient representation of patterns of interaction between appropriate biological elements. Molecular biologists, for example, use networks to represent the patterns of chemical reactions among chemicals in the cell, while neuroscientists use them to represent patterns of connections between brain cells, and ecologists study the networks of interactions between species in ecosystems, such as predation or cooperation. In this chapter we describe the commonest kinds of biological networks and discuss methods for determining their structure.

5.1 BIOCHEMICAL NETWORKS

Among the biological networks those attracting the most attention in recent years have been biochemical networks, networks that represent the molecular-level patterns of interaction and mechanisms of control in the biological cell. The principal types of networks studied in this area are metabolic networks, protein–protein interaction networks, and genetic regulatory networks.

5.1.1 METABOLIC NETWORKS

Metabolism is the chemical process by which cells break down food or nutrients into usable building blocks (so-called *catabolic metabolism*) and then reassemble those building blocks to form the biological molecules the cell needs to complete its other tasks (*anabolic metabolism*). Typically this breakdown and reassembly involves chains or *pathways*, sets of successive chemical reactions

that convert initial inputs into useful end products by a series of steps. The complete set of all reactions in all pathways forms a *metabolic network*.

The vertices in a metabolic network are the chemicals produced and consumed by the reactions. These chemicals are known generically as *metabolites*. By convention the definition of a metabolite is limited to small molecules, meaning things like carbohydrates (such as sugars) and lipids (such as fats), as well as amino acids and nucleotides. Amino acids and nucleotides are themselves the building blocks for larger polymerized macromolecules such as DNA, RNA, and proteins, but the macromolecules are not themselves considered metabolites—they are not produced by simple chemical reactions but by more complex molecular machinery within the cell, and hence are treated separately. (We discuss some of the mechanisms by which macromolecules are produced in Section 5.1.3.)

Although the fundamental purpose of metabolism is to turn food into useful biomolecules, one should be wary of thinking of it simply as an assembly line, even a very complicated one. Metabolism is not just a network of conveyor belts in which one reaction feeds another until the final products fall out the end; it is a dynamic process in which the concentrations of metabolites can change widely and rapidly, and the cell has mechanisms for turning on and off the production of particular metabolites or even entire portions of the network. Metabolism is a complex machine that reacts to conditions both within and outside the cell and generates a broad variety of chemical responses. A primary reason for the high level of scientific interest in metabolic networks is their importance as a stepping stone on the path towards an understanding of the chemical dynamics of the cell.

Generically, an individual chemical reaction in the cell involves the consumption of one or more metabolites that are broken down or combined to produce one or more others. The metabolites consumed are called the *substrates* of the reaction, while those produced are called the *products*.

The situation is complicated by the fact that most metabolic reactions do not occur spontaneously, or do so only at a very low rate. To make reactions occur at a usable rate, the cell employs an array of chemical catalysts, referred to as *enzymes*. Unlike metabolites, enzymes are mostly macromolecules, usually proteins but occasionally RNAs. Like all catalysts, enzymes are not consumed in the reactions they catalyze but they play an important role in metabolism nonetheless. Not only do they enable reactions that would otherwise be thermodynamically disfavored or too slow to be useful, but they also provide one of the mechanisms by which the cell controls its metabolism. By increasing or decreasing the concentration of the enzyme that catalyzes a particular reaction, the cell can turn that reaction on or off, or moderate its speed. Enzymes tend

to be highly specific to the reactions they catalyze, each one enabling only one or a small number of reactions. Thousands of enzymes are known and many more are no doubt waiting to be discovered, and this large array of highly specific catalysts allows for a fine degree of control over the processes of the cell.

The details of metabolic networks vary between different species of organisms but, amongst animals at least, large parts are common to all or most species. Many important pathways, cycles, or other subportions of metabolic networks are essentially unchanged across the entire animal kingdom. For this reason one often refers simply to "metabolism" without specifying a particular species of interest; with minor variations, observations made in one species often apply to others.

The most correct representation of a metabolic network is as a bipartite network. We encountered bipartite networks previously in Section 3.5 on social affiliation networks and in Section 4.3.2 on recommender networks. A bipartite network has two distinct types of vertex, with edges running only between vertices of unlike kinds. In the case of affiliation networks, for example, the two types of vertex represented people and the groups they belonged to. In the case of a metabolic network they represent metabolites and metabolic reactions, with edges joining each metabolite to the reactions in which it participates. In fact, a metabolic network is really a *directed* bipartite network, since some metabolites go into the reaction (the substrates) and some come out of it (the products). By placing arrows on the edges we can distinguish between the ingoing and outgoing metabolites. An example is sketched in Fig. 5.1a.[1]

This bipartite representation of a metabolic network does not include any way of representing enzymes, which, though not metabolites themselves, are still an important part of the metabolism. Although it's not often done, one can in principle incorporate the enzymes by introducing a third class of vertex to represent them, with edges connecting them to the reactions they catalyze. Since enzymes are not consumed in reactions, these edges are undirected—running neither into nor out of the reactions they participate in. An example of such a network is sketched in Fig. 5.1b. Technically this is now a *tripartite network*, partly directed and partly undirected.[2]

Correct and potentially useful though they may be, however, neither of these representations is very often used for metabolic networks. The most

[1] The metabolic network is the only example of a directed bipartite network appearing in this book, and indeed the only naturally occurring example of such a network the author has come across, although no doubt there are others to be discovered if one looks hard enough.

[2] Also the only such network in the book.

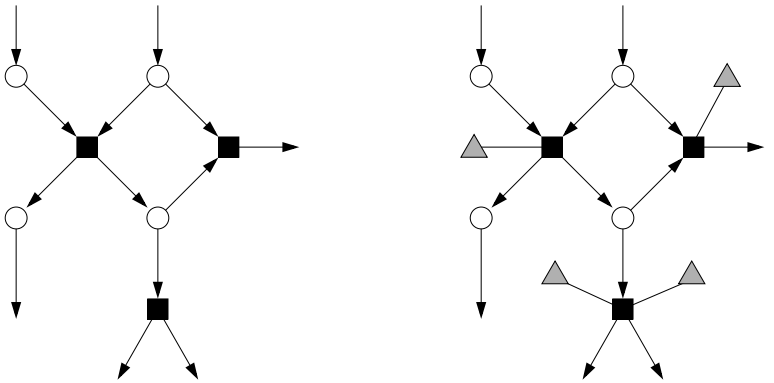


Figure 5.1: Bipartite and tripartite representations of a portion of a metabolic network. (a) A metabolic network can be represented as a directed bipartite network with vertices for the metabolites (circles) and reactions (squares) and directed edges indicating which metabolites are substrates (inputs) and products (outputs) of which reactions. (b) A third type of vertex (triangles) can be introduced to represent enzymes, with undirected edges linking them to the reactions they catalyze. The resulting network is a mixed directed/undirected tripartite network.

common representations of metabolic networks project the network onto just one set of vertices, either the metabolites or the reactions, with the former being the more popular choice. In one approach the vertices in the network represent metabolites and there is an undirected edge between any two metabolites that participate in the same reaction, either as substrates or as products. Clearly this projection loses much of the information contained in the full bipartite network, but, as we have said, it is nonetheless widely used. Another approach, probably the most common, is to represent the network as a directed network with a single type of vertex representing metabolites and a directed edge from one metabolite to another if there is a reaction in which the first metabolite appears as a substrate and the second as a product. This representation contains more of the information from the full network, but is still somewhat unsatisfactory since a reaction with many substrates or many products appears as many edges, with no easy way to tell that these edges represent aspects of the same reaction. The popularity of this representation arises from the fact that for many metabolic reactions only one product and one substrate are known

Projections of bipartite networks and the associated loss of information are discussed further in Section 6.6.

or are considered important, and therefore the reaction can be represented by only a single directed edge with no confusion arising. A number of companies produce large charts showing the most important parts of the metabolic network in this representation. An example is shown in Fig. 5.2. Such charts have become quite popular as wall decorations in the offices of molecular biologists and biochemists, although whether they are actually useful in practice is unclear.

The experimental measurement of metabolic networks is a complex and laborious process, although it has been made somewhat easier in recent years with the introduction of new techniques from molecular genetics. Experiments tend to focus neither on whole networks nor on individual reactions but on metabolic pathways. A number of tools are available to probe the details of individual pathways. Perhaps the most common is the use of radioactive isotopes to trace the intermediate products along a pathway. In this technique, the organism or cell under study is injected with a substrate for the pathway of interest in which one or more of the atoms has been replaced by a radioisotope. Typically this has little or no effect on the metabolic chemistry, but as the reactions of the pathway proceed, the radioactive atoms move from metabolite to metabolite. Metabolites can then be refined, for example by mass spectroscopy or chromatography, and tested for radioactivity. Those that show it can be assumed to be "downstream" products in the pathway fed by the initial radioactive substrate.

This method tells us the products along a metabolic pathway, but of itself does not tell us the order of the reactions making up the pathway. Knowledge of the relevant biochemistry—which metabolites can be transformed into which others by some chemical reaction—can often identify the ordering or at least narrow down the possibilities. Careful measurement of the strength of radioactivity of different metabolites, coupled with a knowledge of the half-life of the isotope used, can also give some information about pathway structure as well as rates of reactions.

Notice, however, that there is no way to tell if any of the reactions discovered have substrates other than those tagged with the radioisotope. If new substrates enter the pathway at intermediate steps (that is, they are not produced by earlier reactions in the pathway) they will not be radioactive and so will not be measured. Similarly, if there are reaction products that by chance do not contain the radioactive marker they too will not be measured.

An alternative approach to probing metabolic pathways is simply to increase the level of a substrate or enzyme for a particular reaction in the cell, thereby increasing the levels of the products of that reaction and those downstream of it in the relevant pathway or pathways, increases that can be mea-

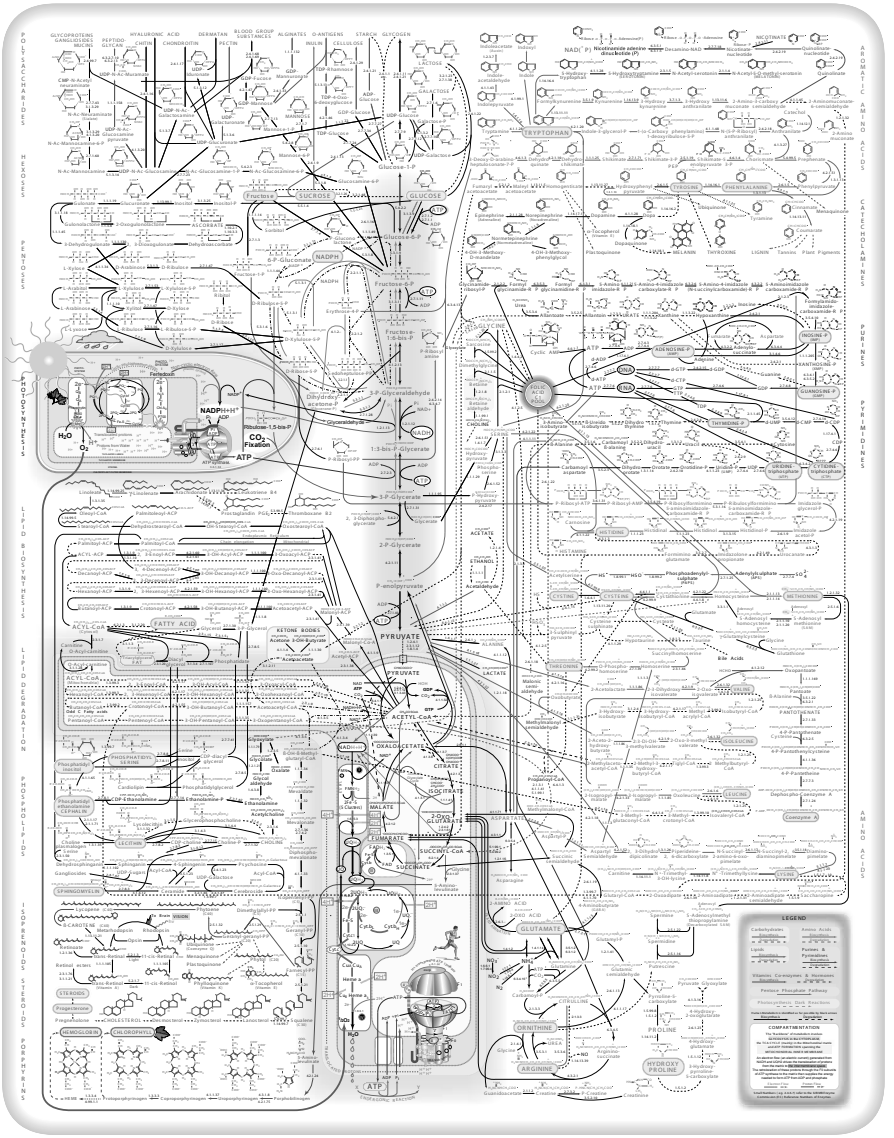


Figure 5.2: A metabolic network. (See Plate IV for color version.) A wallchart showing the network formed by the major metabolic pathways. Created by Donald Nicholson. Copyright of the International Union of Biochemistry and Molecular Biology. Reproduced with permission.

sured to determine the constituents of the pathway. This technique has the advantage of being able to detect products other than those that carry a particular radioactive marker inherited from a substrate, but it is still incapable of identifying substrates other than those produced as products along the pathway.

A complementary experimental technique that can probe the substrates of reactions is reaction inhibition, in which a reaction in a pathway is prevented from taking place or its rate is reduced. Over time, this results in a build-up in the cell of the substrates for that reaction, since they are no longer being used up. By watching for this build-up one can determine the reaction substrates. In principle the same method could also be used to determine the products of the reaction, since their concentration would decrease because they are not being produced any longer, but in practice this turns out to be a difficult measurement and is rarely done.

The inhibition of a reaction is usually achieved by disabling or removing an enzyme necessary for the reaction. This can be done in a couple of different ways. One can use *enzyme inhibitors*, which are chemicals that bind to an enzyme and prevent it from performing its normal function as a catalyst, or one can genetically alter the organism under study to remove or impair its ability to produce the enzyme (a so-called *knockout* experiment). The same techniques can also be used to determine which reactions are catalyzed by which enzymes in the first place, and hence to discover the structure of the third, enzymatic part of the tripartite metabolic network pictured in Fig. 5.1b.

The construction of a complete or partial picture of a metabolic network involves the combination of data from many different pathways, almost certainly derived from experiments performed by many different experimenters using many different techniques. There are now a number of public databases of metabolic pathway data from which one can draw to assemble networks, the best known being KEGG and MetaCyc. Assembling the network itself is a non-trivial task. Because the data are drawn from many sources, careful checking against the experimental literature (or "curation," as the lingo goes) is necessary to insure consistent and reliable inputs to the process, and missing steps in metabolic pathways must often be filled in by guesswork based on biochemistry and a knowledge of the genetics. A number of computer software packages have been developed that can reconstruct networks from raw metabolic data in an automated fashion, but the quality of the networks they create is generally thought to be poorer than that of networks created by knowledgeable human scientists (although the computers are much faster).

5.1.2 PROTEIN–PROTEIN INTERACTION NETWORKS

The metabolic networks of the previous section describe the patterns of chemical reactions that turn one chemical into another in the cell. As we have noted, the traditional definition of metabolism is restricted to small molecules and does not include proteins or other large molecules, except in the role of enzymes, in which they catalyze metabolic reactions but do not take part as reactants themselves.

Proteins do however interact with one another and with other biomolecules, both large and small, but the interactions are not purely chemical. Proteins sometimes interact chemically with other molecules—exchanging small subgroups, for example, such as the exchange of a phosphate group in the process known as phosphorylation. But the primary mode of protein–protein interaction—interactions of proteins with other proteins—is physical, their complicated folded shapes interlocking to create so-called *protein complexes* (see Fig. 5.3) but without the exchange of particles or subunits that defines chemical reactions.

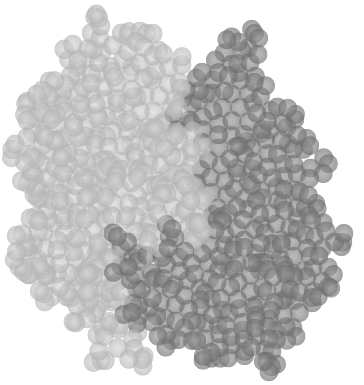
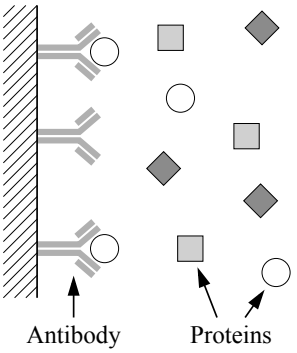


Figure 5.3: Two proteins joined to form a protein complex. Protein molecules can have complicated shapes that interlock with one another to form protein complexes.

The set of all protein–protein interactions forms a *protein–protein interaction network*, in which the vertices are proteins and two vertices are connected by an undirected edge if the corresponding proteins interact. Although this representation of the network is the one commonly used, it omits much useful information about the interactions. Interactions that involve three or more proteins, for instance, are represented by multiple edges, and there is no way to tell from the network itself that such edges represent aspects of the same interaction. This problem could be addressed by adopting a bipartite representation of the network similar to the one we sketched for metabolic networks in Fig. 5.1, with two kinds of vertex representing proteins and interactions, and undirected edges connecting proteins to the interactions in which they participate. Such representations, however, are rarely used.

There are a number of experimental techniques available to probe for interactions between proteins. One of the most reliable and trusted is *co-immunoprecipitation*. Immunoprecipitation (without the "co-") is a technique for extracting a single protein species from a sample containing more than one. The technique borrows from the immune system, which produces *antibodies*, specialized proteins that attach or *bind* to a specific other target protein when the two encounter each other. The immune system uses antibodies to neutralize proteins, complexes, or larger structures that are harmful to the body, but experimentalists have appropriated them for use in the laboratory. Immunopre-



In immunoprecipitation, antibodies attached to a solid surface bind to a specific protein, represented here by the circles, pulling it out of the solution.

cipitation involves attaching an antibody to a solid surface, such as the surface of a glass bead, then passing a solution containing the target protein (as well as others, in most cases) over the surface. The antibody and the target protein bind together, effectively attaching the protein to the surface via the antibody. The rest of the solution is then washed away, leaving the target protein to be recovered from the surface.

There are known naturally occurring antibodies for many proteins of scientific interest, but researchers also routinely create antibodies for specific proteins by injecting those proteins (or more often a portion of a protein) into an animal to provoke its immune system to generate the appropriate antibody.

Co-immunoprecipitation is an extension of the same method to the identification of protein interactions. An antibody is again attached to a suitable solid surface and binds to a known protein in a sample. If that protein is attached to others, forming a protein complex, then the entire complex will end up attached to the surface and will remain after the solution is washed away. Then the complex can be recovered from the surface and the different proteins that make it up individually identified, typically by testing to see if they bind to other known antibodies (a technique known as a *Western blot*).

Although well-established and reliable, co-immunoprecipitation is an impractical approach for reconstructing entire interaction networks, since individual experiments, each taking days, have to be performed for every interaction identified. If appropriate antibodies also have to be created the process would take even longer; the creation of a single antibody involves weeks or months of work, and costs a considerable amount of money too. As a result, the large-scale study of protein–protein interaction networks did not really take off until the adoption in the 1990s and early 2000s of so-called *high-throughput* methods for discovering interactions, methods that can identify interactions quickly and in a semi-automated fashion.

The oldest and best established of the high-throughput methods for protein interactions is the *two-hybrid screen*, invented by Fields and Song in 1989 [119].[3] This method relies on the actions of a specialized protein known as a *transcription factor*, which, if present in a cell, turns on the production of another protein, referred to as a *reporter*. The presence of the reporter can be detected by the experimenter by any of a number of relatively simple means. The idea of the two-hybrid screen is to arrange things so that the transcription factor is created when two proteins of interest interact, thereby turning on the reporter, which tells us that the interaction has taken place.

Transcription factors are discussed in more detail in Section 5.1.3.

[3] Also called a *yeast two-hybrid screen* or Y2HS for short, in recognition of the fact that the technique is usually implemented inside yeast cells, as discussed later.

The two-hybrid screen relies on the fact that transcription factors are typically composed of two distinct parts, a so-called *binding domain* and an *activation domain*. It turns out that most transcription factors do not require the binding and activation domains to be actually attached to one another for the transcription factor to work. If they are merely in close enough proximity production of the reporter will be activated.

In a two-hybrid screen, a cell, usually a yeast cell, is persuaded to produce two proteins of interest, each with one of the domains of the transcription factor attached to it. This is done by introducing *plasmids* into the cell, fragments of DNA that code for the proteins and domains. Then, if the two proteins in question interact and form a complex, the two domains of the transcription factor will be brought together and, with luck, will activate production of the reporter.

See Section 5.1.3 for a discussion of DNA coding of proteins.

In a typical two-hybrid experiment, the protein attached to the binding domain of the transcription factor is a known protein (called the *bait* protein) whose interactions the experimenter wants to probe. Plasmids coding for a large number of other proteins (called *prey*) attached to copies of the activation domain are created, resulting in a so-called *library* of possible interaction targets for the bait. The plasmids for the bait and the library of prey are then introduced into a culture of yeast cells, with the concentration of prey carefully calibrated so that at most one prey plasmid enters each cell in most cases. Cells observed to produce the reporter are then assumed to contain plasmids coding for prey proteins that interact with the bait and the plasmids are recovered from those cells and analyzed to determine the proteins they correspond to.

The two-hybrid screen has two important advantages over older methods like co-immunoprecipitation. First, one can employ a large library of prey and hence test for interactions with many proteins in a single experiment, and second, the method is substantially cheaper and faster than co-immunoprecipitation per interaction detected. Where co-immunoprecipitation requires one to obtain or create antibodies for every protein tested, the two-hybrid screen requires only the creation of DNA plasmids and their later sequence analysis, both relatively simple operations for an experimenter armed with the machinery of modern genetic engineering.

One disadvantage of the two-hybrid screen is that the presence of the two domains of the transcription factor attached to the bait and prey proteins can get in the way of their interacting with one another and prevent the formation of a protein complex, meaning that some legitimate protein–protein interactions will not take place under the conditions of the experiment.

The principal disadvantage of the method, however, is that it is simply unreliable. It produces high rates of both false positive results—apparent in-

teractions between proteins that in fact do not interact—and false negative results—failure to detect true interactions. By some estimates the rate of false positives may be as high as 50%, meaning that fully half of all interactions detected by the method are not real. This has not stopped a number of researchers from performing analyses on the interaction networks reconstructed from two-hybrid screen data, but the results should be viewed with caution. It is certainly possible that many or even most of the conclusions of such studies are substantially inaccurate.

An alternative and more accurate class of methods for high-throughput detection of protein interactions are the *affinity purification* methods (also sometimes called *affinity precipitation* methods). These methods are in some ways similar to the co-immunoprecipitation method described previously, but avoid the need to develop antibodies for each protein probed. In an affinity purification method, a protein of interest is "tagged" by adding a portion of another protein to it, typically by introducing a plasmid that codes for the protein plus tag, in a manner similar to the introduction of plasmids in the two-hybrid screen. Then the protein is given the opportunity to interact with a suitable library of other proteins and a solution containing the resulting protein complexes (if any) passed over a surface to which are attached antibodies that bind to the tag. As a result, the tag, the attached protein, and its interaction partners are bound to the surface while the rest of the solution is washed away. Then, as in co-immunoprecipitation, the resulting complex or complexes can be analyzed to determine the identities of the interaction partners.

The advantage of this method is that it requires only a single antibody that binds to a known tag, and the same tag–antibody pair can be used in different experiments to bind different proteins. Thus, as with the two-hybrid screen, one need only generate new plasmids for each experiment, which is relatively easy, as opposed to generating new antibodies, which is slow and difficult. Some implementations of the method have a reliability comparable to that of co-immunoprecipitation. Of particular note is the method known as *tandem affinity purification*, which combines two separate purification stages and generates correspondingly higher-quality results. Tandem affinity purification is the source for some of the most reliable current data for protein–protein interaction networks.

As with metabolic reactions, there are now substantial databases of protein interactions available online, of which the most extensive are IntAct, MINT, and DIP, and from these databases interaction networks can be constructed for analysis. An example is shown in Fig. 5.4.

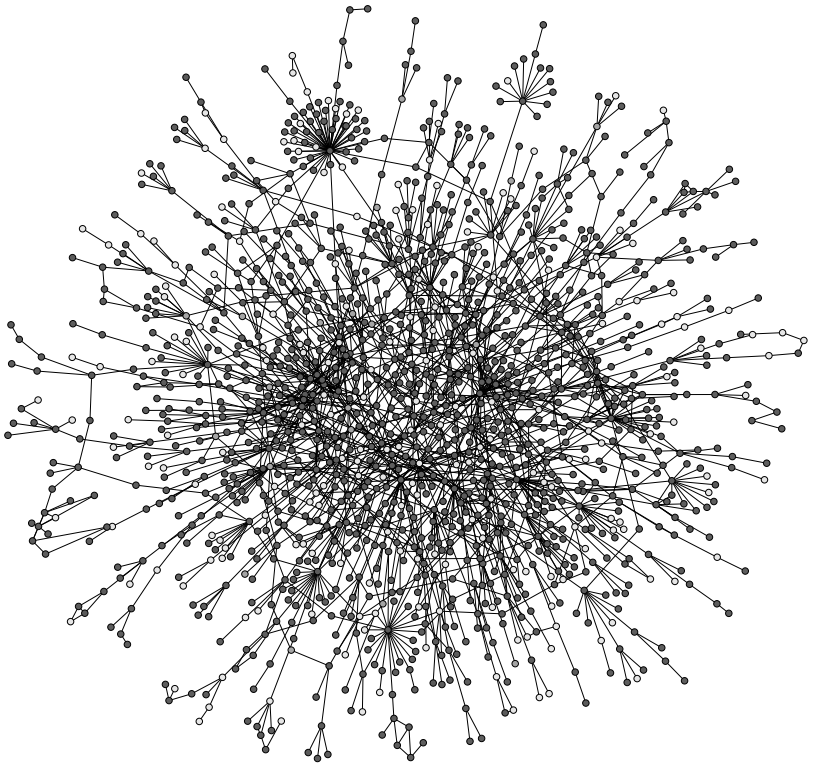


Figure 5.4: A protein–protein interaction network for yeast. A network of interactions between proteins in the single-celled organism *Saccharomyces cerevisiae* (baker's yeast), as determined using, primarily, two-hybrid screen experiments. From Jeong *et al.* [164]. Copyright Macmillan Publishers Ltd. Reproduced by permission.

5.1.3 GENETIC REGULATORY NETWORKS

As discussed in Section 5.1.1, the small molecules needed by biological organisms, such as sugars and fats, are manufactured in the cell by the chemical reactions of metabolism. Proteins, however, which are much larger molecules, are manufactured in a different manner, following recipes recorded in the cell's genetic material, DNA.

Proteins are biological polymers, long-chain molecules formed by the concatenation of a series of basic units called *amino acids*. The individual amino acids themselves are manufactured by metabolic processes, but their assembly into complete proteins is accomplished by the machinery of genetics. There are

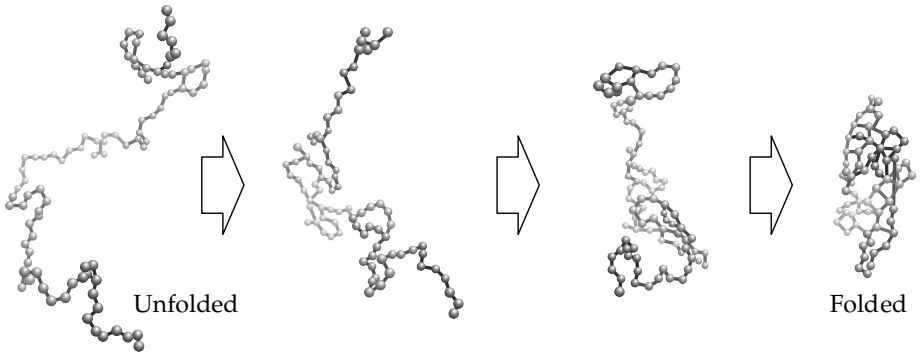


Figure 5.5: Protein folding. Proteins, which are long-chain polymers of amino acids, do not naturally remain in an open state (left), but collapse upon themselves to form a more compact folded state (right).

20 distinct amino acids that are used by all living organisms to build proteins, and different species of proteins are distinguished from one another by the particular sequence of amino acids that make them up. Once created, a protein does not stay in a loose chain-like form, but folds up on itself under the influence of thermodynamic forces and mechanical constraints, reliably producing a specific folded form or *conformation* whose detailed shape depends on the amino acid sequence—see Fig. 5.5. A protein's conformation dictates the physical interactions it can have with other molecules and can expose particular chemical groups or active sites on the surface of the protein that contribute to its biological function within the organism.

A protein's amino acid sequence is determined by a corresponding sequence stored in the DNA of the cell in which the protein is synthesized. This is the primary function of DNA in living matter, to act as an information storage medium containing the sequences of proteins needed by the cell. DNA is itself a long-chain polymer made up of units called *nucleotides*, of which there are four distinct species, adenine, cytosine, guanine, and thymine, commonly denoted A, C, G, and T, respectively.[4] The amino acids in proteins are encoded in DNA as trios of consecutive nucleotides called *codons*, such as ACG

[4]Technically, DNA is a double-stranded polymer, having two parallel chains of nucleotides forming the famous double helix shape. However, the two strands contain essentially the same sequence of nucleotides and so for our purposes the fact that there are two is not important (although it is very important in other circumstances, such as in the reproduction of a cell by cellular division and in the repair of damaged DNA).

or TTT, and a succession of such codons spells out the complete sequence of amino acids in a protein. A single strand of DNA can code for many proteins—hundreds or thousands of them—and two special codons, called the start and stop codons, are used to signal the beginning and end within the larger DNA strand of the sequence coding for a protein. The DNA code for a single protein, from start codon to stop codon, is called a *gene*.

Proteins are created in the cell by a mechanism that operates in two stages. In the first stage, known as *transcription*, an enzyme called *RNA polymerase* makes a copy of the coding sequence of a single gene. The copy is made of RNA, another information-bearing biopolymer, chemically similar but not identical to DNA. RNA copies of this type are called *messenger RNAs*. In the second stage, called *translation*, the protein is assembled, step by step, from the RNA sequence by an ingenious piece of molecular machinery known as a *ribosome*, a complex of interacting proteins and RNA. The translation process involves the use of *transfer RNAs*, short molecules of RNA that have a region at one end that recognizes and binds to a codon in the messenger RNA and a region at the other end that pulls the required amino acid into the correct place in the growing protein. The end result is a protein, assembled following the exact prescription spelled out in the corresponding gene. In the jargon of molecular biology, one says that the gene has been *expressed*.

The cell does not, in general, need to produce at all times every possible protein for which it contains a gene. Individual proteins serve specific purposes, such as catalyzing metabolic reactions, and it is important for the cell to be able to respond to its environment and circumstances by turning on or off the production of individual proteins as required. It does this by the use of *transcription factors*, which are themselves proteins and whose job is to control the transcription process by which DNA sequences are copied to RNA.

Transcription is performed by the enzyme RNA polymerase, which works by attaching to a DNA strand and moving along it, copying nucleotides one by one. The RNA polymerase doesn't just attach spontaneously, however, but is aided by a transcription factor. Transcription factors are specific to particular genes or sets of genes and regulate transcription in a variety of ways, but most commonly by binding to a recognized sub-sequence in the DNA, called a *promoter region*, which is adjacent to the beginning of the gene. The binding of the transcription factor to the promoter region makes it thermodynamically favorable for the RNA polymerase to attach to the DNA at that point and start transcribing the gene. (The end of the gene is marked by a stop codon and upon encountering this codon the RNA polymerase automatically detaches from the DNA strand and transcription ends.) Thus the presence in the cell of the transcription factor for the gene turns on or enhances the expression of that

gene. We encountered an example of a transcription factor previously in our discussion of the two-hybrid screen in Section 5.1.2.

There are also transcription factors that inhibit expression by binding to a DNA strand in such a way as to prevent RNA polymerase from attaching to the strand and hence prevent transcription and the production of the corresponding protein.

But now here is the interesting point: being proteins, transcription factors are themselves produced by transcription from genes. Thus the protein encoded in a given gene can act as a transcription factor promoting or inhibiting production of one or more other proteins, which themselves can act as transcription factors for further proteins and so forth. The complete set of such interactions forms a *genetic regulatory network*. The vertices in this network are proteins or equivalently the genes that code for them and a directed edge from gene A to gene B indicates that A regulates the expression of B. A slightly more sophisticated representation of the network distinguishes between promoting and inhibiting transcription factors, giving the network two distinct types of edge.

The experimental determination of the structure of genetic regulatory networks involves identifying transcription factors and the genes that they regulate. The process has several steps. To begin with, one first confirms that a given candidate protein does bind to DNA roughly in the region of a gene of interest. The commonest technique for establishing the occurrence of such a binding is the *electrophoretic mobility shift assay*.[5] In this technique one creates strands of DNA containing the sequence to be tested and mixes them in solution with the candidate protein. If the two indeed bind, then the combined DNA/protein complex can be detected by *gel electrophoresis*, a technique in which one measures the speed of migration of electrically charged molecules or complexes through an agarose or polyacrylamide gel in an imposed electric field. In the present case the binding of the DNA and protein hinders the motion of the resulting complex through the gel, measurably reducing its speed when compared with unbound DNA strands. Typically one runs two experiments side by side, one with protein and one without, and compares the rate of migration to determine whether the protein binds to the DNA. One can also run parallel experiments using many different DNA sequences to test which (if any) bind to the protein.

An alternative though less sensitive technique for detecting binding is the *deoxyribonuclease footprinting assay*. Deoxyribonucleases (also called DNases

[5] "Assay" is biological jargon for an experimental test.

for short) are enzymes that, upon encountering DNA strands, cut them into shorter strands. There are many different DNases, some of which cut DNA only in particular places according to the sequence of nucleotides, but the footprinting technique uses a relatively indiscriminate DNase that will cut DNA at any point. If, however, a protein binds to a DNA strand at a particular location it will often (though not always) prevent the DNase from cutting the DNA at or close to that location. Footprinting makes use of this by mixing strands of DNA containing the sequence to be tested with the DNase and observing the resulting mix of strand lengths after the DNase has cut the DNA samples into pieces in a variety of different ways. Repeating the experiment with the protein present will result in a different mix of strand length if the protein binds to the DNA and prevents it from being cut in certain places. The mix is usually determined again by gel electrophoresis (strands of different lengths move at different speeds under the influence of the electric field) and one again runs side-by-side gel experiments with and without the protein to look for the effects of binding.

Both the mobility shift and footprinting assays can tell us if a protein binds somewhere on a given DNA sequence. To pin down exactly where it binds one typically must do some further work. For instance, one can create short strands of DNA, called *oligonucleotides*, containing possible sequences that the protein might bind to, and add them to the mix. If they bind to the protein then this will reduce the extent to which the longer DNAs bind and visibly affect the outcome of the experiment. By a combination of such experiments, along with computer-aided guesswork about which oligonucleotides are likely to work best, one can determine the precise sub-sequence to which a particular protein binds.

While these techniques can tell us the DNA sequence to which a protein binds, they cannot tell us which gene's promoter region that sequence belongs to (if any), whether the protein actually affects transcription of that gene, or, if it does, whether the transcription is promoted or inhibited. Further investigations are needed to address these issues.

Identification of the gene is typically done not by experiment but by computational means and requires a knowledge of the sequence of the DNA in the region where the protein binds. If we know the DNA sequence then we can search it for occurrences of the sub-sequence to which our protein binds, and then examine the vicinity to determine what gene or genes are there, looking for example for start and stop codons in the region and then recording the sequence of other codons that falls between them. Complete DNA sequences are now known for a number of organisms as a result of sequencing experiments starting in the late 1990s, and the identification of genes is as a result a

relatively straightforward task.

Finally, we need to establish whether or not our protein actually acts as a transcription factor, which can be done either computationally or experimentally. The computational approach involves determining whether the subsequence to which the protein binds is indeed a promoter region for the identified gene. (It is possible for a protein to bind near a gene but not act as a transcription factor because the point at which it binds has no effect on transcription.) This is a substantially harder task than simply identifying nearby genes. The structure of promoter regions is, unfortunately, quite complex and varies widely, but computer algorithms have been developed that can identify them with some reliability.

Alternatively, one can perform an experiment to measure directly the concentration of the messenger RNA produced when the gene is transcribed. This can be achieved for example by using a *microarray* (colloquially known as a "DNA chip"), tiny dots of DNA strands attached in a grid-like array to a solid surface. RNA will bind to a dot if a part of its sequence matches the sequence of the dot's DNA and this binding can be measured using a fluorescence technique. By observing the simultaneous changes in binding on all the dots of the microarray, one can determine with some accuracy the change in concentration of any specific RNA and hence quantify the effect of the transcription factor. This technique can also be used to determine whether a transcription factor is a promoter or an inhibitor, something that is currently not easy using computational methods.

As with metabolic pathways and protein–protein interactions, there now exist electronic databases of genes and transcription factors, such as EcoCyc, from which it is possible to assemble snapshots of genetic regulatory networks. Current data on gene regulation are substantially incomplete and hence so are our networks, but more data are being added to the databases all the time.

5.2 NEURAL NETWORKS

A completely different use of networks in biology arises in the study of the brain and central nervous system in animals. One of the main functions of the brain is to process information and the primary information processing element is the *neuron*, a specialized brain cell that combines (usually) several inputs to generate a single output. Depending on the animal, an entire brain can contain anywhere from a handful of neurons to more than a hundred billion, wired together, the output of one cell feeding the input of another, to create a *neural network* capable of remarkable feats of calculation and decision making.

Figure 5.6 shows a sketch of a typical neuron, which consists of a cell body

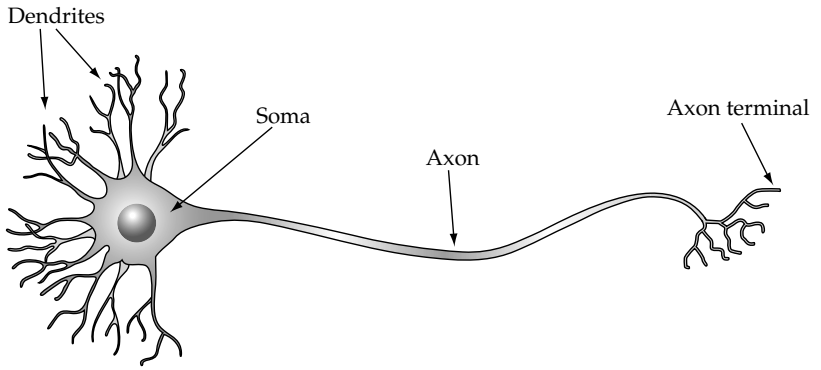


Figure 5.6: The structure of a neuron. A typical neuron is composed of a cell body or soma with many dendrites that act as inputs and a single axon that acts as an output. Towards its tip, the axon branches to allow it to connect to the inputs of several other neurons.

or *soma*, along with a number of protruding tentacles, which are essentially wires for carrying signals in and out of the cell. Most of the wires are inputs, called *dendrites*, of which a neuron may have just one or two, or as many as a thousand or more. Most neurons have only one main output, called the *axon*, which is typically longer than the dendrites and may in some cases extend over large distances to connect the cell to others some way away. Although there is just one axon, it usually branches near its end to allow the output of the cell to feed the inputs of several others. The tip of each branch ends at an *axon terminal* that abuts the tip of the input dendrite of another neuron. There is a small gap, called a *synapse*, between terminal and dendrite across which the output signal of the first (presynaptic) neuron must be conveyed in order to reach the second (postsynaptic) neuron. The synapse plays an important role in the function of the brain, allowing transmission from cell to cell to be regulated by chemically modifying the properties of the gap.[6]

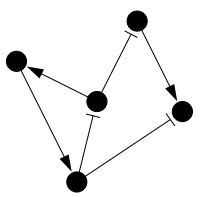
The actual signals that travel within neurons are electrochemical in nature. They consist of traveling waves of electrical voltage created by the motion of positively charged sodium and potassium ions in and out of the cell. These

[6]Neurons do sometimes have direct connections between them without synapses. These direct connections are called *gap junctions*, a confusing name, since it sounds like it might be a description of a synapse but is in reality quite different. In our brief treatment of neural networks, however, we will ignore gap junctions.

waves are called *action potentials* and typically consist of voltages on the order of tens of millivolts traveling at tens of meters per second. When an action potential reaches a synapse, it cannot cross the gap between the axon terminal and the opposing dendrite and the signal is instead transmitted chemically; the arrival of the action potential stimulates the production of a chemical neurotransmitter by the terminal, and the neurotransmitter diffuses across the gap and is detected by receptor molecules on the dendrite at the other side. This in turn causes ions to move in and out of the dendrite, changing its voltage.

These voltage changes, however, do not yet give rise to another traveling wave. The soma of the postsynaptic neuron sums the inputs from its dendrites and as a result may (or may not) send an output signal down its own axon. The neuron is stable against perturbations caused by voltages at a small number of its inputs, but if enough inputs are excited they can collectively drive the neuron into an unstable runaway state in which it "fires," generating a large electrochemical pulse that starts a new action potential traveling down the cell's axon and so a signal is passed on to the next neuron or neurons in the network. Thus the neuron acts as a switch or gate that aggregates the signals at its inputs and only fires when enough inputs are excited.

As described, inputs to neurons are excitatory, increasing the chance of firing of the neuron, but inputs can also be inhibiting—signals received at inhibiting inputs make the receiving neuron less likely to fire. Excitatory and inhibiting inputs can be combined in a single neuron and the combination allows neurons to perform quite complex information processing tasks all on their own, while an entire brain or brain region consisting of many neurons can perform tasks of extraordinary complexity. Current science cannot yet tell us exactly how the brain performs the more sophisticated cognitive tasks that allow animals to survive and thrive, but it is known that the brain constantly changes the pattern of wiring between neurons in response to inputs and experiences, and it is presumed that this pattern—the neural network—holds much of the secret. An understanding of the structure of neural networks is thus crucial if we are ever to explain the higher-level functions of the brain.



A wiring diagram for a small neural network.

At the simplest level, a neuron can be thought of as a unit that accepts a number of inputs, either excitatory or inhibiting, combines them, and generates an output result that is sent to one or more further neurons. In network terms, a neural network can thus be represented as a set of vertices—the neurons—connected by two types of directed edges, one for excitatory inputs and one for inhibiting inputs. By convention, excitatory connections are denoted by an edge ending with an arrow "$\longrightarrow$", while inhibiting connections are denoted by an edge ending with a bar "$\dashv$".

In practice, neurons are not all the same. They come in a variety of differ-

ent types and even relatively small regions or circuits in the brain may contain many types. This variation can be encoded in our network representation by different types of vertex. Visually the types are often denoted by using different shapes for the vertices or by labeling. In functional terms, neurons can differ in a variety of ways, including the number and type of their inputs and outputs, the nature and speed of their response to their inputs, whether and to what extent they can fire spontaneously without receiving inputs, and many other things besides.

Experimental determination of the structure of neural networks is difficult and the lack of straightforward experimental techniques for probing network structure is a major impediment to current progress in neuroscience. Some useful techniques do exist, however, although their application can be extremely laborious.

The basic tool for structure determination is microscopy, either optical or electronic. One relatively simple approach works with cultured neurons on flat dishes. Neurons taken from animal brains at an early stage of embryonic development can be successfully cultured in a suitable nutrient medium and will, without prompting, grow synaptic connections to form a network. If cultured on a flat surface, the network is then roughly two-dimensional and its structure can be determined with reasonable reliability by simple optical microscopy. The advantage of this approach is that it is quick and inexpensive, but it has the substantial disadvantage that the networks studied are not the networks of real living animals and their structure is probably not very similar to that of a functional brain circuit.

In this respect, studies of real brains are much more satisfactory and likely to lead to greater insight, but they are also far harder, because real brains are three-dimensional and we do not currently have any form of microscopy suitable for probing such three-dimensional structures. Instead, therefore, researchers have resorted to cutting suitably preserved brains or brain regions into thin slices, whose structure is then determined by electron microscopy. Given the structure of an entire set of consecutive slices, one can, at least in principle, reconstruct the three-dimensional structure, identifying different types of neurons by their appearance, where possible. In the early days of such studies, most reconstruction was done by hand but more recently researchers have developed computer programs that can significantly speed the reconstruction process. Nonetheless, studies of this kind are very laborious and can take months or years to complete, depending on the size and complexity of the network studied.

Figure 5.7 shows an example of a "wiring diagram" of a neural network, reconstructed by hand from electron microscope studies of this type. The net-

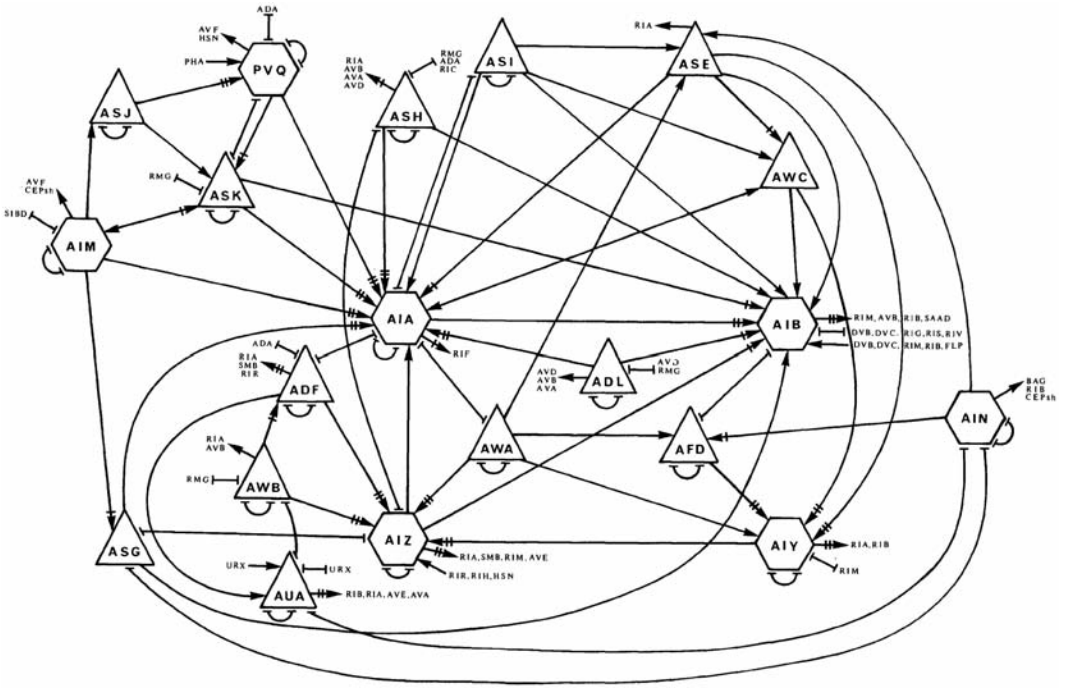


Figure 5.7: A diagram of a part of the brain circuitry of a worm. A portion of the neural circuitry of the nematode *Caenorhabditis elegans*, reconstructed by hand from electron micrographs of slices through the worm's brain. Reproduced from White *et al.* [328]. Copyright of the Royal Society. Reproduced by permission.

work in question is the neural network of the worm *Caenorhabditis elegans*, one of the best studied organisms in biology. The brain of *C. elegans* is simple—it has less than 300 neurons and essentially every specimen of the worm has the same wiring pattern. Several types of neuron, denoted by shapes and labels, are shown in the figure, along with a number of different types of connection, both excitatory and inhibiting. Some of the connections run out of the figure or enter from somewhere off the page. These are connections that run to or from other parts of the network not shown. The original experimenters determined the structure of the entire network and presented it as set of interconnected wiring diagrams like this one [328].

5.3 ECOLOGICAL NETWORKS

The final class of biological network that we consider in this chapter is networks of ecological interactions between species. Species in an ecosystem can interact in a number of different ways. They can eat one another, they can parasitize one another, they can compete for resources, or they can have any of a variety of mutually advantageous interactions, such as pollination or seed dispersal. Although in principle the patterns of interactions of all of these types could be represented in a combined "interaction network" with several different edge types, ecologists have traditionally separated interaction types into different networks. Food webs, for example—networks of predator–prey interactions (i.e., who eats whom)—have a long history of study. Networks of hosts and parasites or of mutualistic interactions are less well studied, but have nonetheless received significant attention in recent years.

5.3.1 FOOD WEBS

The biological organisms on our planet can be divided into *ecosystems*, groups of organisms that interact with one another and with elements of their environment such as sources of material, nutrients, and energy. Mountains, valleys, lakes, islands, and larger regions of land or water can all be home to ecosystems composed of many organisms each. Within ecological theory, ecosystems are usually treated as self-contained units with no outside interactions, although in reality perfect isolation is rare and many ecosystems are only approximately self-contained. Nonetheless, the ecosystem concept is one of significant practical utility for understanding ecological dynamics.

A *food web* is a directed network that represents which species prey on which others in a given ecosystem.[7] The vertices in the network correspond to species and the directed edges to predator–prey interactions. Figure 5.8 shows a small example, representing predation among species living in Antarctica. There are several points worth noticing about this figure. First, notice that not all of the vertices actually represent single species in this case. Some of them do—the vertices for sperm whales and humans, for instance. But some of them represent collections of species, such as birds or fish. This is common practice

[7] In common parlance, one refers to a *food chain*, meaning a chain of predator–prey relations between organisms starting with some lowly organism at the bottom of the chain, such as a microbe of some kind, and working all the way up to some ultimate predator at the top, such as a lion or a human being. Only a moment's reflection, however, is enough to convince us that real ecosystems cannot be represented by single chains, and a complete network of interactions is needed in most cases.

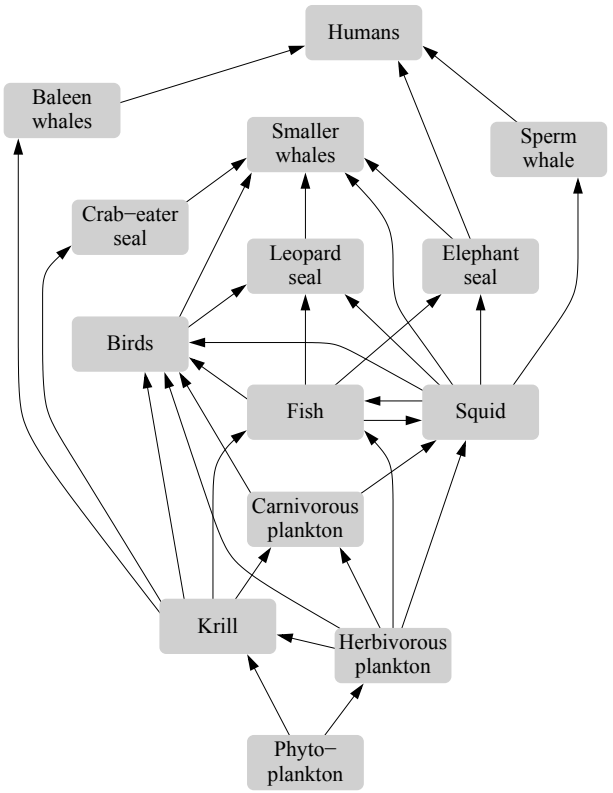


Figure 5.8: A food web of species in Antarctica. Vertices in a food web represent species or sometimes, as with some of the vertices in this diagram, groups of related species, such as fish or birds. Directed edges represent predator–prey interactions and run in the direction of energy flow, i.e., from prey to predator.

in the network representation of food webs. If a set of species such as birds all prey upon and are preyed on by the same other species, then the network can be simplified by representing them as a single vertex, without losing any information about who preys on whom. Indeed, even in cases where a set of species only have mostly, but not exactly, the same predators and prey we still sometimes group them, if we feel the benefits of the resulting simplification are worth a small loss of information. A set of species with the same or similar predators and prey is sometimes referred to as a *trophic species*.

Second, notice the direction of the edges in the network. One might imagine that the edges would point from predators to prey, but ecologists conven-

tionally draw them in the opposite direction, from prey to predator. Thus the edge representing the eating of fish by birds runs *from* the fish vertex *to* the bird vertex. The reason for this apparently odd choice is that ecologists view food webs as representations of the flow of energy (or sometimes carbon) within ecosystems. The arrow from fish to birds indicates that the population of birds gains energy from the population of fish when the birds eat the fish.

Third, notice that almost all the arrows in the figure run up the page. Directed networks with this property—that they can be drawn so that the edges all run in one direction—are called *acyclic networks*. We encountered acyclic networks previously in our discussion of citation networks in Section 4.2. Food webs are usually only approximately acyclic. There are usually a few edges that do not run in the right direction,[8] but it is often a useful approximation to assume that the network is acyclic.

Acyclic networks are discussed in more detail in Section 6.4.2.

The acyclic nature of food webs indicates that there is an intrinsic pecking order among the species in ecosystems. Those higher up the order (which means higher up the page in Fig. 5.8) prey on those lower down, but not vice versa. A species' position in this pecking order is called by ecologists its *trophic level*. Species at the very bottom of the food web, of which there is just one in our example—the phytoplankton—have trophic level 1. Those that prey on them—krill, herbivorous plankton—have trophic level 2, and so forth all the way up to the species at the top of the web, which have no predators at all. In our antarctic example there are two species that have no predators, humans and small whales. (Note however that although such species are all, in a sense, at "the top of the food chain" they need not have the same trophic level.)

Trophic level is a useful general guide to the roles that species play in ecosystems, those in lower trophic levels tending to be smaller, more abundant species that are prey to other species higher up the food web, while those in higher trophic levels are predators, usually larger-bodied and less numerous. Calculating a species' trophic level, however, is not always easy. In principle, the rule is simple: a species' trophic level is 1 greater than the trophic level of its prey. Thus the herbivorous plankton and krill in our example have trophic level 2, because their prey has trophic level 1, and the carnivorous plankton have trophic level 3. On the other hand, the squid in our example prey on species at two different levels, levels 2 and 3, so it is unclear what level they belong to. A variety of mathematical definitions have been proposed to resolve this issue. One strategy is to define trophic level to be 1 greater than the mean of the trophic levels of the prey. There is, however, no accepted standard

[8] In Fig. 5.8, for example, there are edges in both directions between the fish and squid vertices, which makes it impossible to draw the network with all edges running in the same direction.

definition, and the only indisputable statement one can make is that in most food webs some species have ill-defined or mixed trophic level.

The food webs appearing in the ecological literature come in two basic types. *Community food webs* are complete webs for an entire ecosystem, as in Fig. 5.8—they represent, at least in theory, every predator–prey interaction in the system. *Source food webs* and *sink food webs* are subsets of complete webs that focus on species connected, directly or indirectly, to a specific prey or predator. In a source food web, for instance, one records all species that derive energy from a particular source species, such as grass. Our food web of antarctic species is, in fact, both a community food web and a source food web, since all of the species in the network derive their energy ultimately from phytoplankton. Phytoplankton is the source in this example, and the species above it (all of the species in this case) form the corresponding source web. A sink food web is the equivalent construct for a particular top predator in the network. In the antarctic example, for instance, humans consume the sperm and baleen whales and elephant seals, which in turn derive their energy from fish, squid, plankton, krill, and ultimately phytoplankton. This subset of species, therefore, constitutes the sink food web for humans—the web that specifies through which species or species groups the energy consumed by humans passes.

The experimental determination of the structure of food webs is typically done in one of two different ways, or sometimes a mixture of both. The first and most straightforward method is direct measurement. Having settled on the ecosystem to be studied, one first assembles a list of the species in that ecosystem and then determines their predator–prey interactions. For large-bodied animals such as mammals, birds, or larger fish, some predation can be established simply by observation in the field—we see a bird eating a fish and the presence of the corresponding edge is thereby established. More often, however, and particularly with smaller-bodied animals, interactions are established by catching and dissecting the animals in question and examining the contents of their stomachs to determine what they have been eating.

The second primary method of constructing food webs is by compilation from existing literature. Many predator–prey interactions are already known and have been recorded in the scientific literature, but not in the context of the larger food web, and one can often reconstruct a complete or partial picture of a food web by searching the literature for such records. Many of the currently available food web data sets were assembled in this way from pre-existing data, and some others were assembled by a combination of experimental measurement and literature searches.

In some cases attempts have also been made to measure not merely the presence (or absence) of interactions between species but also the strength of

those interactions. One can quantify interaction strength by the fraction of its energy a species derives from each of its predators, or by the total rate of energy flow between a prey species and a predator. The result is a weighted directed network that sheds considerably more light on the flow of energy through an ecosystem than the more conventional unweighted food web. Measurements of interaction strength are, however, time-consuming, difficult, and yield uncertain results, so the current data on weighted food webs should be treated with caution.

Food web data from a variety of sources have been assembled into publicly available databases, starting in the late 1980s. Examples include the Ecoweb database [73] and the web-based collection at `www.foodwebs.org`.

5.3.2 OTHER ECOLOGICAL NETWORKS

Two other types of ecological network have received significant attention in the scientific literature (although less than has been paid to food webs). *Host–parasite networks* are networks of parasitic relationships between organisms, such as the relationship between a large-bodied animal and the insects and microorganisms that live on and inside it. In a sense parasitic relations are a form of predation—one species eating another—but in practical terms they are quite distinct from traditional predator–prey interactions. Parasites, for example, tend to be smaller-bodied than their hosts where predators tend to be larger, and parasites can live off their hosts for long, sometimes indefinite, periods of time without killing them, where predation usually results in the death of the prey.

Parasitic interactions, however, do form networks that are somewhat similar to traditional food webs. Parasites themselves frequently play host to still smaller parasites (called "hyperparasites"), which may have their own still smaller ones, and so forth through several levels.[9] There is a modest but growing literature on host–parasite networks, much of it based on research within the agriculture community, a primary reason for interest in parasites being their prevalence in and effects on livestock and crop species.

The other main class of ecological networks is that of *mutualistic networks*, meaning networks of mutually beneficial interactions between species. Three

[9] One is reminded of the schoolhouse rhyme by Augustus de Morgan:

> Great fleas have little fleas upon their backs to bite 'em,
> And little fleas have lesser fleas, and so ad infinitum.

specific types of mutualistic network that have received attention in the ecological literature are networks of plants and the animals (primarily insects) that pollinate them, networks of plants and the animals (such as birds) that disperse their seeds, and networks of ant species and the plants that they protect and eat. Since the benefit of a mutualistic interaction runs, by definition, in both directions between a pair of species, mutualistic networks are undirected networks (or bidirectional, if you prefer), in contrast with the directed interactions of food webs and host–parasite networks. Most mutualistic networks studied are also bipartite, consisting of two distinct, non-overlapping sets of species (such as plants and ants), with interactions only between members of different sets. In principle, however, non-bipartite mutualistic networks are also possible.

See Section 6.6 for a discussion of bipartite networks.