Chapter 1

Introduction to Networks

A network is a mathematical abstraction made up of discrete points connected by line segments. The individual points are called the *nodes*, and the lines connecting the nodes are called the *links*. Fig. 1.1 shows a network.

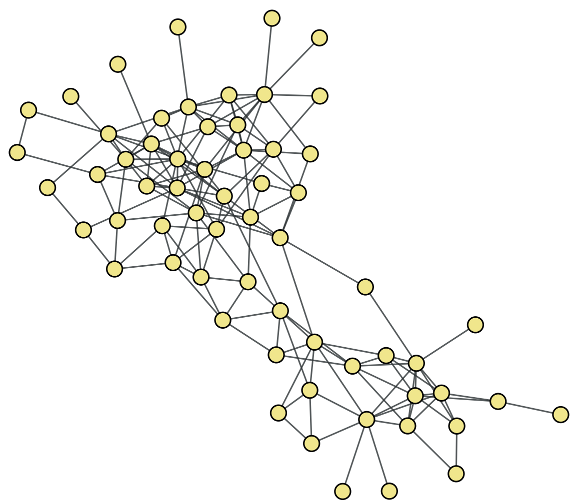

Figure 1.1: An example network. The yellow points are the *nodes* and the lines connecting the nodes are the *links*.

What good are networks? Many (in fact too many!) natural and artificial systems are in the form of a large number of elements or components that are connected to each other in an intricate fashion. Here are some examples: Human brain (neurons connected to each other by synapses), the World Wide Web (webpages connected to each other by hyperlinks), Transportation (cities connected to each other by highways, rail routes, air routes etc), Facebook (individual accounts connected to each other by friendships). A moment's reflection will make it clear that in spite of belonging to very different disciplines, each of these systems can be thought of as a network in which the individual components are nodes and the connections or interactions between the components are links. Thus, networks provide a uniform framework for studying highly diverse systems, and because of this universality, studying a networked system belonging to a particular field (say biology), often provides useful insights into the working of a networked system from another field (say social sciences).

1.1 A note on the terminology

As mentioned above, the study of networks is spread across many disciplines, and many fields have devised their own terminology to talk about networks. Here we provide a quick overview of some important terms.

1. A **Network** is also commonly called a **Graph** (especially by mathematicians). In fact, a separate branch of mathematics, called *Graph theory* exists that studies properties of graphs in their own right. Occasionally one also uses the terms **Net** and **Web**. The World Wide Web is a prominent example of the later term. Physicists many times use the term **Nontrivial Lattice**, but this is becoming increasingly rare. However, a special type of networks are still called virtually by all as **Lattices**. We will mostly use the terms Networks and Graphs.

2. A **Node** in a network is also commonly called a **Vertex** (again by mathematicians) [1]. In social sciences the term **Actor** is commonly used since in social networks the nodes are often the human beings. Physicists tend to use the term **Site** for historical reasons. We will mostly use the terms Node and Vertex.

3. A **Link** is also called an **Edge** in mathematics. In physics one commonly uses the term **Bond**. In social sciences the term **Tie** is used. Again, we will mostly use the terms Link and Edge.

4. A **Neighbour** of a vertex is a vertex that is connected to it by an edge. In Fig. 1.2, vertex number 2 is a neighbour of vertex 4. Vertex 0 has no neighbour at all.

5. The **Degree** of a vertex in a network is the total number of neighbours it has. Thus, the vertex 4 has degree 3, vertices 1, 2, 3 each have degree 1, and vertex 0 has degree 0. A vertex with degree zero is also called an **isolated vertex**.

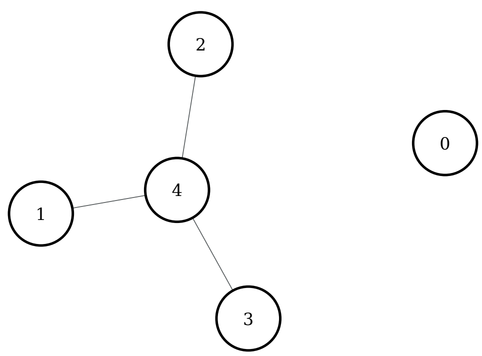

Figure 1.2: A small graph with 5 vertices.

1.2 Modeling with networks

Before we delve into the study of networks, it is important that we ask two questions to ourselves. First, which systems *should* be modelled by networks? In other words, if a system can be modelled as a network, should you always do it? Second, what, if at all, will we learn if we model a given system as a network? If we get no extra information or insight by modeling the system as a network, the whole formalism of networks is probably useless.

[1] plural: vertices

1.2.1 When should networks be used to model a real-world system?

The first question is relatively easy to answer, although it is also somewhat subjective. Consider an example of a galaxy. Astrophysicists who study galaxy are interested in the dynamics of the galaxy as a result of the interaction between the stars that make up that galaxy. In other words, it looks like this is a perfect system to model as a network with individual stars as vertices and the gravitational interactions between them as edges. However, because of the nature of the gravitational force, all stars interact with each other. Hence, in the constructed network, all the possible edges would be present, and we will get a trivial network. For this reason, the use of networks doesn't provide any advantage over traditional methods to study a galaxy (for example fluid dynamics based methods), and probably one shouldn't model a galaxy as a network. As an another example, consider a problem of studying a crystal structure of a material. In this case, we again have individual atoms that form bonds with each other to form a crystal, and at first sight it looks like we can easily model a crystal as a network with the individual atoms as nodes and the bonds between them as links. However, in this case the system under consideration is highly regular, and its structure around any point is repeated in space. Because of this regularity, probably networks won't provide any extra advantage in studying crystal structure. (Technically, physicists model a crystal structure using what is known as a 'lattice' which is a highly regular network). To summarize, the power of networks is seen when we model systems in which interactions between the parts are sparse and irregular.

1.2.2 What is 'complex' in complex networks?

The title of the present course is 'complex networks' instead of just 'networks'. Let us briefly discuss the reason behind this. It should be clear by now that the term 'network' is a quite generic, and can encompass a spectrum of systems including those that are completely regular like lattices to completely random. However, the field of complex networks is concerned with using the framework of networks to understand the real-world networked systems. When more and more data about these systems started becoming available, it was realized that in many respects they are not completely random, and there must exist organizing principles which govern their shape. At the same time, these principles don't seem to be as rigid as those of Euclidean geometry, and allow the system to incorporate randomness in it. The result is that the real-world networks are neither completely random nor completely regular, and lie somewhere in between. The term 'complex' conveys this point succintly, and hence the field is now known by the name 'complex networks'. However, bear in mind that 'complex' does not mean 'complicated'. A complicated system may or may not be 'complex'. Perhaps a reasonable example is an intricate design created by an artist: although it might be extremely complicated, we woudn't call it complex. Unfortunately, there doesn't exist a single definition of what 'complex' means, and here we will not spend time pondering over this somewhat philosophical question.

1.2.3 What do we learn by modeling a real-world system as a network?

This is a broad question, and certainly can't be answered in a single page. However, we could look at some specific cases to illustrate how useful the networks could be.

1. Consider a situation in which a rumour is spreading over Facebook, and this is causing other problems like riots, and thus needs to be stopped or at least slowed down immediately. A standard way used by authorities to achieve this is to block some Facebook accounts temporarily. Since these blocked accounts can't forward the rumour further, the hope is that the rumour will slow down. However, this could be done only for a small fraction of the total accounts, since suspending too many accounts may negatively affect the service's reputation. Thus, the question is how do we determine which accounts to suspend so that suspension of only a small number of accounts will drastically reduce the spread of the rumour? If we look at this question from a network point of view, we immediately see that instead of randomly blocking accounts, it would be much more effective to block accounts which lie on the network paths between the accounts to which rumour has reached, and others to which it hasn't!
2. Consider a network shown in Fig. 1.3. This network is a network of friendships among the members of a karate club's members based in a north American University. A researcher Wayne Zachary constructed the network by actually observing the members over a span of two years. During the observation, a dispute arose in the club over the issue of whether to raise the fees of the club, and this eventually led

to a split of the club resulting into two separate clubs. In the figure, the different colors and shapes show the actual division of the network, and it can be seen that the members did not join the two clubs randomly. Instead, they tried to join that faction which their friends also joined. Later we will see that networks sometimes let us predict such splits, and hence provide valuable insights into such intricate social dynamics.

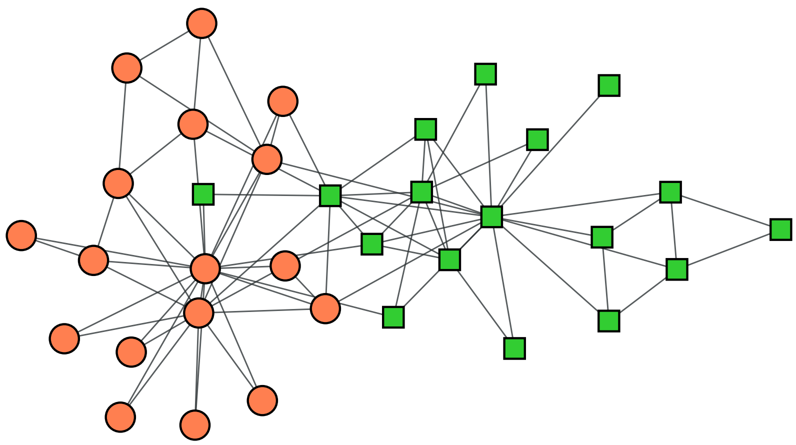

Figure 1.3: Zachary karate club network

3. Consider the workings of navigation systems like Google maps. In such systems, the user inputs the starting point and the destination, and then the system tells us which route to take to reach to the destination in the minimum amount of time. Ignoring for a moment the issues like traffic and quality of the roads, we can see that usually there are several routes that one could take. How does the navigation system decide which is the best route? The answer depends on the structure of the street network and if the structure changes, the best route would potentially change. This also means that while planning the new routes, it is important to take into account the overall network structure, and not only isolated locations and the paths connecting them. The network structure becomes even more important when we also take into account the factors like traffic since these dynamic factors are significantly affected by the underlying network.

These are only a few prototype examples that attest to the fact that networks are becoming increasingly important in as diverse fields as biology, transportation, social sciences, web security, disaster management, ecology, linguistics, medicine, and physics. Recently, the formalism of networks has even found applications in solving problems in pure mathematics and theoretical physics.

1.3 A (very) brief history of networks

Although the last two decades have seen a sudden interest in the theory of networks, the field has a long and interesting history. Especially sociologists have been working with networks to understand dynamics of groups of humans since the last hundred years or so. Many of the terms and the concepts that we use today to deal with the networks are actually borrowed from sociologists.

Almost surely, the first definitive use of a concept of a graph was made by the great mathematician Leonard Euler to solve the problem of 'Bridges of Königsberg'. The river Pregel flows through the city of Königsberg in Prussia (now the city is called Kaliningrad, and is situated in Russia). At a particular location in the city, the river divides itself into two branches which again meet forming an island in between the branches. But the branches separate again leaving another portion of land in between them. To connect the landmasses with each other, seven bridges were built over the river. See Fig. 1.4.

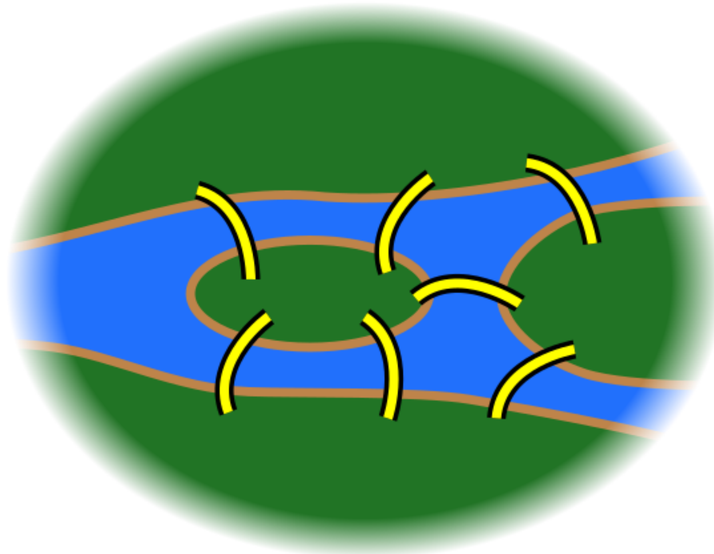

Figure 1.4: A graphic depicting the river Pregel (blue) flowing through the city of Königsberg. The landmasses are shown in green, and the seven bridges connecting the landmasses are shown as yellow segments. (*Credit: CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=990540*)

A problem with which people seem to have amused themselves was to cross each of these bridges exactly once. This means that to fulfill the challenge successfully, one must cross each bridge, and while doing so, should not cross any bridge more than once. It is said that many people over the years tried and failed to solve the problem. In 1736 Euler proved that the task was impossible by converting the problem into a problem of walking on a graph. Later we will see how exactly Euler solved this problem, but for now we just note this as a formal start of the theory of graphs.

In the last century, the progress in the field of networks was mostly divided into different disciplines with researchers from each field thinking mostly about networks occurring in their own field. The prominent fields here are Computer Science, Sociology and Mathematics. One of the reasons that these fields mostly remained isolated was that the fields have very different aims. At the start of the current century, the situation started changing rapidly thanks to the availability of huge amounts of data on the networked systems. These data showed that many common features exist in the data coming from diverse fields, and it was also becoming more and more clear that a unified approach would benefit all the fields in terms of tools and methods of analysis as well as using insights gathered in one field to apply to another one. Since then, there has been a surge of activity in the field, and the last two decades' efforts have resulted into several remarkable insights about the real-world networked systems. This in turn has helped in controlling the functioning of these systems, and making them more robust and efficient as we will see throughout the course.