1 Introduction

1.1 Some Background and Motivation

Between two randomly selected persons in the world, roughly how many friends are there connecting them together? When searching from one webpage to another through the World Wide Web (WWW), how many clicks are needed in average? How can computer viruses propagate so fast and so wide through the Internet? How are people infected by epidemics such as AIDS, SARS, and bird flu all over the world? How do rumors spread out in human societies? How does an economic crisis occur and then being recovered regionally or even globally? How does electric power blackout emerge from local system failures through the huge power grid? How can traffic jams in metropolitan cities be regulated effectively? How can the human brain work so efficiently? … All these seemingly different issues have something to do with "networks"—Internet, WWW, social relationship networks, viruses and rumors propagation networks, economic trading and competition networks, power and traffic flow networks, wired and wireless communication networks, biological neural networks, biological metabolic networks, and so on. Most important above all, these apparently different networks have a lot in common.

Since the 1990s, the rapid growth of the Internet as an icon of the high-tech era has led our life to an age of networks. The influence of various complex and dynamical networks is currently pervading all kinds of sciences, ranging from physical to biological, even to social sciences. Its impact on modern engineering and technology is prominent and will be far-reaching. There is no doubt that we are living in a networked world today. On one hand, networks bring us with convenience and benefits, improve our efficiency of work and quality of life, and create tremendous opportunities which we never had before. On the other hand, however, networks also generate harms and damages to humans and societies, typically with epidemic spreading, computer virus propagation, and power blackout, to name just a few. Therefore, the increasing demand for networks and networking also requires a correct view and a serious investigation of the complex properties of various networks. For a long time in the history, studies of communication networks, power networks, biological networks, economic networks, social networks, etc., were carried out separately and independently. However, recently there are some rethinking of the general concept and theory of complex dynamical networks towards a better understanding of the intrinsic relations, common properties and shared features of all kinds of networks in the real world. The new intention of studying the fundamental properties and dynamical behaviors of most, if not all complex networks, both qualitatively and quantitatively, is important and timely, although very challenging technically. The current research along this line has been considered as "new science of networks" [1,2], and has become overwhelming today.

Life science is perhaps the most exciting revolutionary area of scientific research today. The main stream of research in life science in the twentieth century was the reductionism-based molecular biology. The fundamental principle of the reductionism is that, within different levels of the structure of a system, the high-level dynamical behaviors are completely determined by those at the lower levels. There was a common belief that if the individual basic ingredients of life (e.g., DNA, RNA and proteins) can be well understood then the activities and behaviors of cells at the higher level can be comprehended, while the interactions among these basic elements even among molecules can be neglected. Yet, this traditional reductionism has been seriously challenged in the beginning of the twenty-first century due to the many significant discoveries of the importance and essence of networking interactions and interactive dynamics between different levels of the life structure and among large numbers of tiny ingredients.

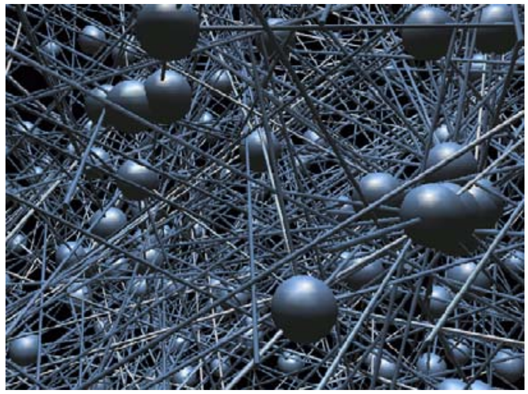
Networks are generally complex, and the complexity of networks may be viewed from different perspectives:

1. *Structural complexity*: A network usually appears structurally complicated, which may even be seemingly messy and disordered (Fig. 1-1). The network topology (i.e., structure) may vary in time (e.g., the WWW has new web-pages to join and old web-sites to be removed everyday). Moreover, the edge connections among nodes may be weighted or directed (e.g., brain cells can be stimulated or restrained and the connections between cells can be strong or weak).

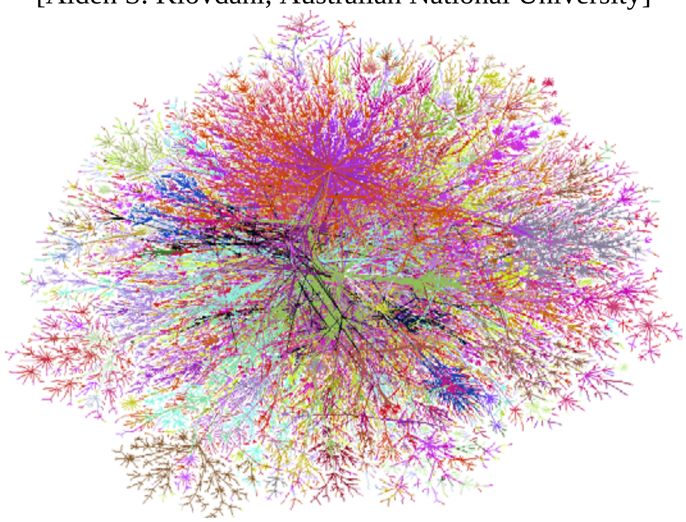
2. *Node-dynamic complexity*: A node in the network can be a dynamical system, which may have bifurcating and even chaotic behaviors (e.g., gene networks or Joshphson lattices have dynamically evolving nodes). Moreover, a network may have different kinds of nodes (e.g., a power grid has electric generators and also has loads such as motors, lights, etc.).

3. *Mutual interactions among various complex factors*: A real-world network is typically affected by many internal and external factors (e.g., if the coupled brain cells are repeatedly excited by certain stimuli then their connections will be strengthened, which is considered the basic reason for learning and memorization by a repeated process of studies). Furthermore, the close relations between networks or sub-networks make the already-complicated behaviors of each of them become much more complex and intrinsic (e.g., the blackout of a huge power grid may lead to chain reactions in human life and industrial productions and may slowdown the activities of other related networks, such as transportation and traffic, communication and information, and economic and financial networks).

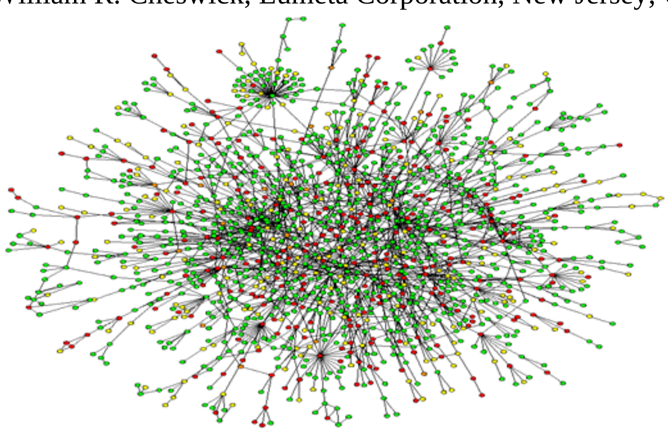
In the intensive study of nonlinear science and dynamical systems, on the other hand, networked systems have been one of the focal topics for research since the mid-twentieth century. However, most of such coupled dynamical systems were placed in a fixed and regularly connected network model for investigation, where the main interest was the complexity caused by the node dynamics but not the network topology. Typical examples of this type are coupled map lattice (CML) [4] and cellular neural (or nonlinear) networks (CNN) [5], which can generate rich spatiotemporal patterns. By assuming a network with a regular topology, one can focus on the effects of the node dynamics on the behaviors of the network in interest, setting aside the troublesome influence of the network structure. Moreover, the networked elements in a regular topology can be easily implemented by integrated circuits, which is a main concern in commercial applications of networked devices, systems and infrastructures.



(a) Illustrative graph of a social relationship structure in Canberra, Australia
[Alden S. Klovdahl, Australian National University]



(b) Illustrative graph of some IP addresses in the Internet
[William R. Cheswick, Lumeta Corporation, New Jersey, USA]



(c) Illustrative graph of interactions among proteins
[Hawoong Jeong, Korea Advanced Institute of Science and Technology]

Fig. 1-1 Three illustrative graphs with complex structures

1.2 A Brief History of Complex Network Research

1.2.1 The Königsburg Seven-Bridge Problem

Complex network research has a long history. The recent study of complex networks has directed most interests to the modeling and understanding of various complex networks, especially the relations between the complexity of the network topology and the behaviors of the network dynamics. To describe the common properties and characteristic features of different types of networks, a rigorous and efficient analytic tool is needed, which has been introduced in the form of *graphs* in mathematics. A network can be viewed as a graph consisting of *nodes* connected by *edges* according to a certain rule or law, in which the nodes and edges do not necessarily have physical meanings in the investigation of abstract networks.

Representing a physical problem by a graph and then solving it by mathematical analysis and computation is not a new idea. This approach can be traced back to the eighteenth century when the great mathematician Leonhard Eüler (1707-1783) studied and solved the famous seven-bridge problem in a town named Königsburg, which is now in the territory of Russia. As shown in Fig. 1-2 [6], there are two small islands in the river Pregel passing through the town Königsburg, and there are seven bridges over the river. In the old days, the residents were always amazed as if someone could walk through all the seven bridges and return to the starting point without going over any bridge for more than once.

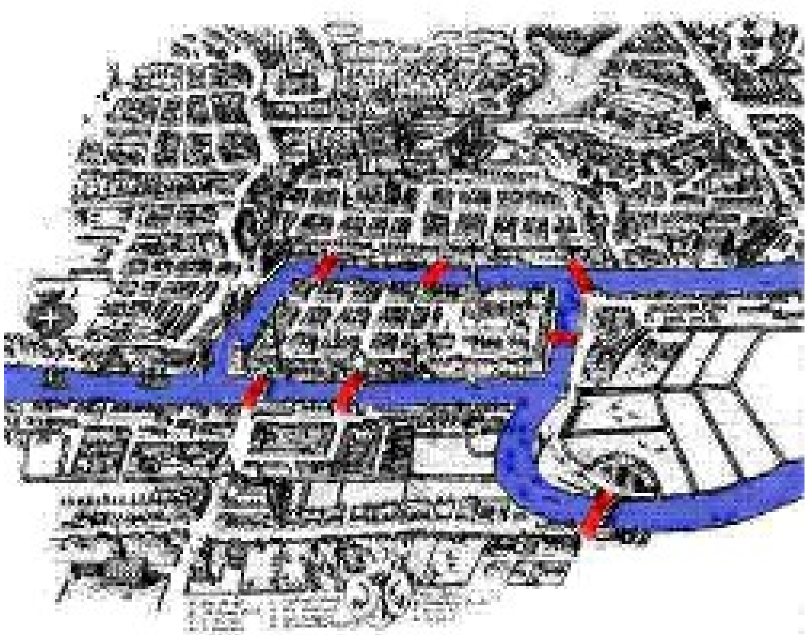


Fig. 1-2 The town Königsburg and the seven bridges in 1736 [6]

In 1736, Eüler had a good idea to describe this real-world problem with an abstract graph, using points *A, B, C, D* to represent the four pieces of lands separated by the river in town, with lines *a, b, c, d, e, f, g* to represent the seven bridges that connect the four points *A, B, C, D* together (Fig. 1-3 [6]). Thus, Eüler was able to convert the physical problem to the following mathematical problem: In the graph shown in Fig. 1-3, starting from any point, is there any possible loop leading back to the starting point such that it passes all the seven lines once and once only?

Eüler furthermore derived a necessary and sufficient condition for the existence of such a loop, thereby proving that the Königsburg seven-bridge problem has no solutions. More precisely, Eüler observed that in order to have such a loop, if a point (*A, B, C,* or *D* in Fig. 1-3) has one incoming edge then it should also have one outgoing edge; therefore, it is necessary for each point to have an even number of edges. However, the graph shown in Fig. 1-3 does not satisfy this condition. More significantly, Eüler also proved that this condition is sufficient as well.

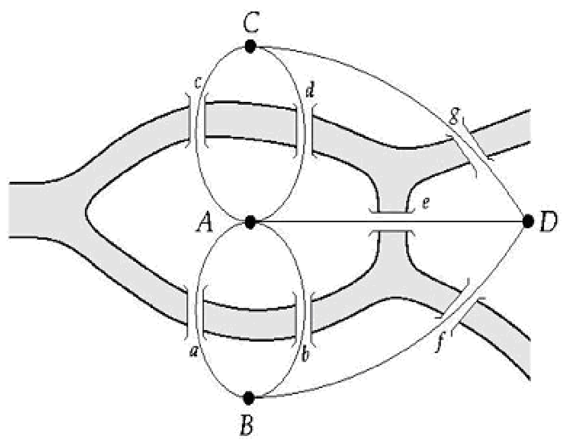


Fig. 1-3 Graph of the Königsburg seven-bridge problem [6**]**

The contribution of Eüler had gone much beyond this simple seven-bridge problem—it has opened up a new branch of mathematics: the graph theory. Thereafter, Eüler was named the father of graph theory, and the picture shown in Fig. 1-3 was called an Eülerian graph. As a matter of fact, this simple graph is a foundation stone of the modern mathematical graph theory, which has led to the extensive studies of complex networks today.

1.2.2 Random Graph Theory

The development of the mathematical graph theory had a very slow start after Eüler solved the Königsburg seven-bridge problem. The first monograph on graph theory was not published until exactly two-hundred years later, in 1936. Nevertheless, the theory was developed rather rapidly thereafter, and the foundation of the now-famous "random graph theory" was laid by two Hungarian mathematicians, Paul Erdös (1913-1996) and Alfréd Rényi (1921-1970), in the late 1950s [7], which is considered the first rigorous and complete modern graph theory.

Erdös and Rényi defined a random graph as N labeled nodes connected by n edges, which are randomly chosen from among the $n(n-1)/2$ possible edges.

A common way to generate an ER random graph is to start with N nodes, from which every possible pair of nodes are being connected with probability p. More specifically, to generate an ER random network, one may start with N isolated nodes, pick up every possible pair of nodes, once and once only from a total of $N(N-1)/2$ pairs of nodes, and then with probability p ($0 < p < 1$) connect the pair with an edge.

Here, "with probability p" can be performed as follows: run a pseudorandom number generator to generate a "1" with probability p: at a step if the generator yields a "1" then connect the pair of nodes by an edge; otherwise, if the generator yields a "0" then do not connect them at all.

Since every possible pair of nodes is picked up once and once only, there will not be multiple edges between any pair of nodes and, moreover, no node has a self-connected edge.

It can be easily seen that in such an ER random graph of N nodes, the expectation value of the total number of edges will be $N(N-1)p/2$, which is a random variable because p is so. Consequently, the probability of obtaining a graph with N nodes and M edges is equal to $p^M(1-p)^{N(N-1)/2-M}$. Erdös and Rényi systematically studied many asymptotic properties, as $N \to \infty$, of such graphs and their relations with the edge-connectivity probability p. If a graph has a property P with probability 1 as $N \to \infty$, then they consider almost every ER random graph has property P. One of their most important and also quite surprising discoveries is that many properties of ER random graphs emerge suddenly but not gradually, in the sense that for a given edge-connectivity probability p, either almost every ER random graph has a certain property P or almost every such graph does not have property P [8].

As a historical remark, Paul Erdös is one of the most distinguished leading mathematicians of the twentieth century, "the man who loves only numbers" [9]. Erdös was legendary; he published more than 1,600 research papers with more than 500 co-authors, and he had made very fundamental contributions in modern mathematics such as number theory, Diophantine equations, combinatory mathematics, probability theory, real and complex analyses, in addition to the random graph theory. Erdös was always excited when he met and worked with other mathematicians. When he met a colleague, he often said "My brain is open"; when he left one co-worker to meet with another, he used to say "Another roof, another proof" [10]. He had devoted his entire life to his beloved mathematics.

1.2.3 Small-World Experiment

A. Milgram's Small-World Experiment

A social network is a collection of individuals and communities, connecting one another according to some relationships. The relationships in a social network can be friends, coworkers, marriages, partners, etc., between people or business cooperation between companies. Fig. 1-1 illustrates one case of the social relationship in Canberra, Australia, drawn based on some real data.

Take the human friendship as an example. Many people have this experience: when talking to a stranger for the first time, both sides easily found a common friend in between, and so they yell surprisingly, "What a small world!" In general, one may wonder, between two randomly selected persons in the world, how many intermediate friends are there connecting them together? In the 1960s, a psychologist of Harvard University, Stanley Milgram, did a survey in the United States and found that the average number of friends in between two randomly selected persons was only 6! This was the famous discovery of "six degree of separation." More precisely, the Milgram social experiment was carried out as follows. He first chose two targeted persons: the wife of a graduate student in the college of mythology in a small town called Sharon in Massachusetts, and a stoke broker in the big city Boston, unknown each other. He then called for two groups of volunteers in Kansas and Nebraska, respectively. He asked the volunteers each to send a letter to a friend and then ask that friend to forward the letter to another friend, and so to continue, towards one of the two targeted persons in Sharon and Boston, respectively. In the May Issue of *Psychology Today* in 1967 [11], Milgram reported his finding—it took only five times of forwarding to reach a target in average, where one surprising example is that a letter from Kansas was first mailed to a priest and the priest forwarded it to his friend in Sharon and then the letter reached the target in the third step! Although not all letters were forwarded in such a short path, Milgram was able to statistically reach a conclusion of "six degree of separation."

Whether or not this number 6 was accurate is actually not very important; what is important is that this number really is a very small number, compared to the population of the region, the country, or even the whole world. This indicates the so-called "small world" property of our very large human society. Milgram's idea and experiment have a great impact on the later studies and analyses of social science and social network research.

B. Network Game of Kevin Bacon

To further verify Milgram's "six degree of separation," several social experiments were carried out lately, among which a quite interesting one was the Game of Kevin Bacon. Bacon (1958—) is a movie actor in Hollywood who was filmed in about 60 movies to date; therefore, he has quite a large number of intervening partners in these movies.

Let us start the game with the following numbering: if an actor or an actress has a movie with Bacon together, then he or she will be given a Bacon number 1; if an actor or actress does not co-show in a movie with Bacon directly but is a partner of someone with Bacon number 1, then this person has a Bacon number 2, and so on. Thus, everyone who plays in a movie will have a Bacon number, big or small, indicating that he or she is a node in the network of movie stars starting from Kevin Bacon.

There is a database of movies and movie stars established in the Computer Science Department, University of Virginia [12], containing information of about 300,000 movies with more than 1,000,000 actors and actresses (Table 1-1). As a simple example, by inputting the name of the Hong Kong movie actor Stephen Chow, one can find that his Bacon number is 2 (as of January 1, 2010). A prominent feature of this table of statistics is that amongst the very large number of more than 1 million actors and actresses, the average Bacon number is very small—only about 3—a small world of movie stars!

Table 1-1 Bacon numbers of actors and actresses (as of January 1, 2010)

Bacon Number	Number of Actors/Actresses
0	1
1	2251
2	222314
3	713068
4	108176
5	12490
6	1075
7	158
8	17

Total number of linkable actors: 1131550
Average Bacon number: 2.984

C. Erdös Numbers

As mentioned above, Paul Erdös published more than 1,600 research papers with more than 500 co-authors. Staring from Erdös, there is a network of co-authorship: if someone (likely a mathematician) has a joint paper with Erdös, then he or she will be given an Erdös number 1; if a person does not co-author a paper with Erdös directly but is a co-author of one with Erdös number 1, then this person has an Erdös number 2, and so on [13]. Thus, sooner or latter, every scientist who has joint paper with someone in the fields of mathematics, physics, engineering and even social science will be a node in this co-authorship network, since mathematicians, scientists and engineers have been closely working together and frequently published their papers together.

As a simple example, Charles K Chui was a friend of Paul Erdös and they published one paper together, therefore Chui has an Erdös number 1 [14]. His former PhD student Guanrong Chen, the author of the present text, has an Erdös number 2 [15]. All the coauthors of Chen, like Xiaofan Wang and Xiang Li who did not publish with Erdös and Chui, have a small Erdös number 3 [16]—"What a small world" indeed!

D. Small-World Experiment over the Internet

From a statistical point of view, although the networks of Bacon and Erdös can be easily computed, they are nevertheless too small to be conclusive. Knowing this, Duncan J Watts and his group of researchers in the Department of Sociology at the University of Columbia set up a "Small World Project" on their website in 2001 [17], trying to carry out a large-scale international experiment to verify the hypothesis of "six degree of separation." They selected some targeted people, with different ages, races, professions and financial statuses as volunteers. Once logged in, a volunteer will be given a piece of information and asked to pass on this piece of information by using e-mail to a friend towards a targeted person, in a way similar to Milgram's experiment. Watts reported their findings in the *Science* magazine in August 2003, showing that within more than one year of time they had more than 60,000 volunteers from 166 countries participated in this game targeting 18 selected people in 13 countries, and 384 e-mails eventually arrived at some targets through 5-7 people in between in average. Although there were some uncontrollable factors such as discontinuity of forwarding e-mails, just like Milgram's experiment, the experimental results agree quite well with Milgram's conclusion—six degree of separation!

1.2.4 Strength of Weak Ties

How did most people find their jobs? Did they rely on their close relatives, or send out letters of application at random, or try the job fairs?

In the late 1960s, a graduate student of Harvard University, Mark Granovetter, started his research on this simple question. He interviewed about 100 people in the greater Boston area, and sent out more than 200 questionnaires, investigating a variety of technical people who either just was offered a new job or just lost the old job. Granovetter surprisingly found that in people's job hunting, usually not those close family ties but some new friends or even occasionally encountered strangers linked them to the new job positions. This means that, more often than not, weak connections lead to strong interactions (here, the results of getting jobs). In the language of network science, a long-range connection may lead to a stronger interaction between two nodes than those short connections of neighboring nodes. One can easily find similar examples from the real world. Here is a typical example provided by Granovetter in his research report [18]: Edward once met a young girl in a gathering in their high school, where he got to know the boyfriend of this girl's elder sister and this gentleman was 10 years older than Edward. Three years later, when Edward lost his job, by chance he met that gentleman again. In conversation, Edward heard that his company was looking for a graph drawer; therefore, he got that new job. This example showed, once again, "What a small world!" And, moreover, how powerful a weak tie could be!

As a side note here, Granovetter submitted his paper "Strength of weak ties" to the *American Sociology Review* in August 1969, but was rejected after four months. His paper was set aside for four years, but then was re-submitted to the *American Journal of Sociology,* consequently being accepted and published [18]. Interesting enough, this paper turns out to be one of the few important papers with the highest impact in the field of social sciences today.

1.2.5 New Era of Complex-Network Studies

Since the 1960s, the theory of ER random graphs has been dominating the academic research. However, it was also aware of that most real-world complex networks are not completely random, i.e., they are not generated by a completely random process. For instance, whether or not two persons are friends, whether or not two routers in the Internet are connected by optical fiber, whether or not two business men make a deal by signing an agreement, etc., are not totally random—they are not being determined by simply tossing a coin. Besides, most natural and manmade networks are rapidly growing networks with evolutionary dynamics, very different from the fundamental framework of the ER random graphs, which is static and non-growing.

The end of the twentieth century was a turning point in network research: networks were being re-visited from a physical rather than a purely mathematical viewpoint by scientists particularly applied physicists, computer scientists, engineers and biologists alike, with a new focus on the global behaviors, complex topologies, and evolving dynamics of the networks.

There were two fundamental research papers on complex networks published at the very end of the last century, opening up a new era of complex-network studies. In June 1998, Duncan J Watts, then a PhD student, and his advisor Steven H Strogatz at the Connell University published an article in *Nature* on the so-called small-world networks [19], followed by another article in *Science* on the so-called scale-free networks by Albert-László Barabási and his then PhD student Réka Albert from the University of Notre Dame in October 1999 [20]. There two seminal articles revealed the most fundamental characteristic of the small-world property and the defining feature of the scale-free property of various complex networks. They have stimulated a huge number of new publications in diverse areas of applied physics, mathematics, computer and biological sciences, engineering and technologies, and social and economic sciences in the following years, about what is confronting us the most today—complex networks.

Summarizing the above discussions, a very brief account of some milestones in the network study history is summarized in Table 1-2.

Table 1-2 A brief history of the network study

Time	People	Event
1736	Eüler	Seven-bridge problem
1959	Erdös and Rényi	Random graph theory
1967	Milgram	Small-world experiment
1973	Granovetter	Strong effect of weak ties
1998	Watts and Strogatz	Small-world network model
1999	Barabási and Albert	Scale-free network model

Scientific research on complex networks had very significant progress at the crossing of the two centuries, attributing to several high-tech developments in engineering and science: (i) the tremendous supercomputing power and the broad-spanned Internet with a vast volume of databases embedded, enabling researchers to collect and process huge amount of real data of different kinds; (ii) fast and wide overlapping of even seemingly unrelated fields of research, leading to many new findings in a broad spectrum of interdisciplinary areas; (iii) the recent breakthroughs in the study of complexity, from reductionism to global and structural understanding, spurs a renewal of interest and a rethinking of the network science, leading the research focus from local to global, from lower level to higher level, and from steady to dynamic.

Due to the restriction of large-scale computational ability and the non-availability of sufficiently large amount of real data, the network study in the past was typically limited to a few hundred or even just a few dozens of nodes in a physical network model. However, we are facing with networks with millions or even billions of nodes and edges such as the Internet, power grids, cash dollars in the world economics, human populations in worldwide social studies, biological neural networks at the cell level, biological systems at the gene levels, crystals in the nano-scale, and so on. Many traditional theories, methodologies and techniques, meaningful and efficient for small-scale computation and analysis, are no longer applicable to such giant networks. The interactions among nodes and between different levels of the complex structure of a huge network generate many unexpected or unpredictable behaviors, such as emergence and chaos, going much beyond the traditional thinking of networks as simple and steady graphs. Consequently, the entire approach to network research seems have to be drastically modified and advanced, or even completely changed. New viewpoints, new theories, new models, and new methods for complex networks are all needed. We have, indeed, already entered a new era of complex-network studies.

At the present time, the studies of complex networks may be roughly categorized as follows:

1) *Discovering*: Trying to reveal the global statistical properties of a network and to develop measures for these properties.
2) *Modeling*: Trying to establish a mathematical model of a given network, enabling better understanding of the network statistical properties and the causes of their appearance.
3) *Analysis*: Trying to find out the basic characteristics and essential features of nodes, edges, and the whole network in a certain topology, to develop fundamental mathematical theories that can describe and predict the network dynamical behaviors.
4) *Control*: Trying to develop effective methods and techniques that can be used to modify and improve network properties and performances, suggesting new and possibly optimal network designs and utilizations, particularly in the regards of network stability, synchronizability, and data-traffic management.
5) *Applications*: Trying to apply and utilize some special and fundamental properties and characteristics of complex networks to facilitate the design and applications of network-related problems such as data-flow congestion control on the Internet and

traffic control for city transportations, optimal integrated circuit design for chip fabrication, better decision-making of policy and strategy for commercial trading and financial management, etc.

It is clear that, accomplishment notwithstanding, complex dynamical networks as a promising and profound research subject is merely at the beginning of a foreseeable far-reaching as well as long-sustained research endeavor. New discoveries, developments, enhancements and improvements are still yet to come.

1.3 Some Basic Concepts

To describe and characterize various complex networks, many concepts and measures have been introduced, among which there are three fundamental ones: average path length, clustering coefficient, and degree distribution. The motivations of introducing these basic concepts came from the following naive considerations: when Watts and Strogatz introduced the small-world network model, they merely wanted to establish a model that has a small average path length just like the ER random graphs and meanwhile has a large cluster coefficient similar to regular networks; when Barabási and Albert introduced the scale-free network model, they intended the model to have a typical power-law type of degree distribution which pertains to most real-world networks. These three fundamental concepts are first introduced below, leaving many others to the following chapters to study.

1.3.1 Graph Representation of Networks

As seen from Eüler's work on the graph representation of the Königsburg seven-bridge problem discussed in Section 1.2.1, a given network can be presented by a set V of *nodes* and a set E of *edges*, connected together as a *graph* denoted $G=(V,E)$, where the total number of nodes is $N = |V|$ and that of edges is $M = |E|$. Clearly, each edge $v \in V$ is connected to one pair of nodes, one at each end. If any pair of nodes, denoted by (i,j) and (j,i) respectively in reverse ordering, is connected by the same edge, then the network is called an *undirected network*; otherwise, it is a *directed network*. On the other hand, if every edge (or, every node) is associated with a weight value, then the network is called a *weighted network*; otherwise, it is an *un-weighted network*. Of course, an un-weighted network can be viewed as a weighted network with all weight values identically equal to unity. In addition, a network may have different kinds of nodes; for example, in a people's relationship network, different nodes may represent individuals with different nationalities, races, ages, genders, professions, etc. Last but not least, in the mathematical graph theory, a *simple graph* requires no repeated edges and no self-loops; namely, between any two nodes there can have at most one edge connecting them, and no nodes can have an edge connecting from and to this node itself.

Figure 1-4 shows a few different types of simple graphs. Only undirected and un-weighted networks with the same type of nodes in the form of a simple graph are studied in this text, although directed and weighted networks will also be briefly discussed sometimes because they have been found very important and useful for real applications.

1.3.2 Average Path Length

The *distance* between two nodes, labeled i and j, can be defined by any metric, denoted d_{ij}. The seemingly simplest definition is adopted throughout this text: the distance between two nodes is equal to the total number of edges that connect them through shortest linkages. For example, in the network shown in Fig. 1-4 (a), the distance between the node at the top and the one at the center is obviously 1; the distance between the node at the top and the right-lowest one is 2 but not 3 nor 5 (one can easily see that there are three paths connecting them, through 2, 3, and 5 edges respectively). In other words, the distance is the shortest path length among all possible path lengths between the two nodes.

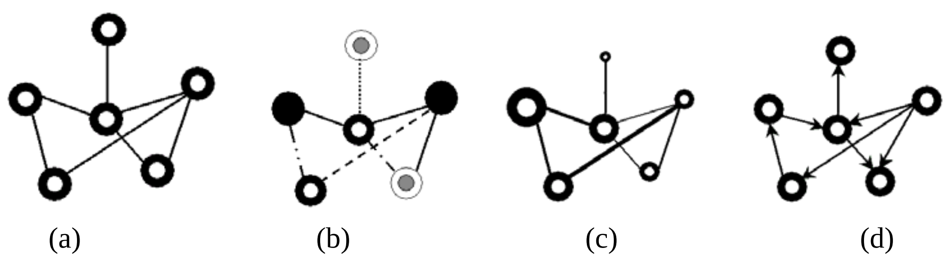


Fig. 1-4 A few examples of different types of networks [23]:
(a) an undirected and un-weighted network with same type of nodes;
(b) an undirected and un-weighted network with different types of nodes and edges;
(c) an undirected but weighted network with weights on both nodes and edges;
(d) a directed but un-weighted network with same type of nodes

The *diameter* of a network, denoted *D*, is defined to be the largest of all distances in the network; that is,

$$D = \max_{i,j} \ d_{ij} \tag{1-1}$$

The diameter of the network shown in Fig. 1-4 (a), for example, is 3 (given by the distance between the node at the top and the left-lowest one).

The average path length of a network is defined to be the average value of all distances over the network:

$$L = \frac{1}{\frac{1}{2}N(N-1)} \sum_{i>j} d_{ij} \tag{1-2}$$

Here, N is the *size* of the network, i.e., the total number of nodes in the network. Note that in this text, self-loop (a node connecting to itself by a single edge) is not considered, therefore are not counted in this formula.

It is clear that the average path length of a network with N nodes and M edges can be searched by an optimization algorithm with computational complexity $O(NM)$.

Figure 1-5 [36] shows another simple example of a network with 5 nodes and 5 edges, where $D = d_{45} = 3$ and $L = 16/10 = 1.6$, since

$$\begin{array}{llll} d_{12} = 1 & d_{13} = 1 & d_{14} = 2 & d_{15} = 1 \\ & d_{23} = 1 & d_{24} = 1 & d_{25} = 2 \\ & & d_{34} = 2 & d_{35} = 2 \\ & & & d_{45} = 3 \end{array}$$

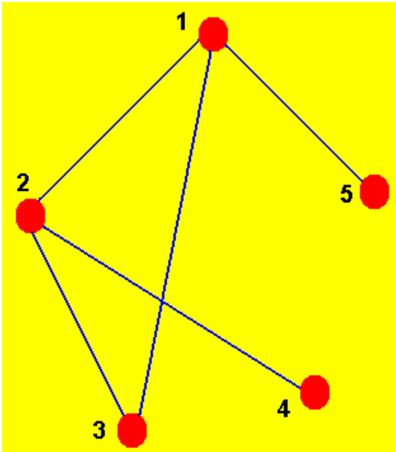


Fig. 1-5 A simple network with 5 nodes and 5 edges [36]

1.3.3 Clustering Coefficient

In a friendship network, two friends of someone may be friends themselves. This phenomenon is characterized by the concept of *clustering*.

In a network, let i be a node with k_i edges directly connecting it to other k_i nodes, which are called *neighbors* of node i. It is easy to verify that there are at most $k_i(k_i - 1)/2$ edges among these k_i neighbors. Let E_i be the number of the actual edges existing among these k_i nodes. Then, the ratio between the actual and the possible numbers of edges in the cluster of these k_i nodes is defined to be the *clustering coefficient* of node i, denoted C_i; namely,

$$C_i = 2\frac{E_i}{k_i(k_i - 1)} \tag{1-3}$$

Consider the example shown in Fig. 1-5, where

Node-1 has 3 neighbors, $E(1) = 1$, so $C_1 = 2 \times 1/(3 \times 2) = 1/3$;

Node-2 has 3 neighbors, $E(2) = 1$, so $C_2 = 2 \times 1/(3 \times 2) = 1/3$;

Node-3 has 2 neighbors, $E(3) = 1$, so $C_3 = 2 \times 1/(2 \times 1) = 1$;

Node-4 has 1 neighbor, $E(4)=0$, so $C_4=0$;
Node-5 has 1 neighbor, $E(5)=0$, so $C_5=0$.

Therefore, their average, denoted $\overline{C}$, is

$$\overline{C}=(1/3+1/3+1+0+0)/5=1/3$$

From a geometric point of view, this is equivalent to

$$C_i=\frac{number\ of\ \ complete\ triangles\ with\ corner\ i}{number\ of\ triangular\ graphs\ with\ corner\ i} \tag{1-4}$$

as visualized by Fig. 1-6 [36], where the complete triangle (left) is counted in both the numerator and the denominator, but the open triangular graph (right) is counted only in the denominator, in the formula (1-4).

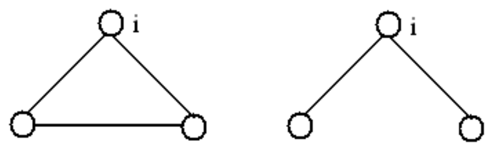


Fig. 1-6 A complete triangle and a triangular graph [36]

Once again, consider the example shown in Fig. 1-5. Since
Node-1 has 1 complete triangle and 3 triangular graphs, so $C_1=1/3$;
Node-2 has 1 complete triangle and 3 triangular graphs, so $C_2=1/3$;
Node-3 has 1 complete triangle and 1 triangular graph, so $C_3=1$;
Node-4 has 0 complete triangles, so $C_4=0$;
Node-5 has 0 complete triangles, so $C_5=0$.

Hence,

$$\overline{C}=(1/3+1/3+1+0+0)/5=1/3$$

The clustering coefficient C of the whole network is the averaged value of all such clustering coefficients C_i of node i, for all i over the entire network. Obviously, $0\le C\le 1$, and $C=0$ if and only if all neighbors are unconnected for any node in the network (i.e., a star-shaped network and a set of isolated nodes) and $C=1$ if and only if all nodes are connected each other (i.e., a fully connected regular network). For an ER random graph with N nodes, since each node has an equal probability to obtain a new edge, every node eventually has about the same number of edges, so that $C=O(N^{-1})$ for large N. It has been a common observation, however, that most real-world networks have a clustering coefficient C in the order of $C=O(1)$, satisfying $O(N^{-1})<<C<<1$. This also shows that most real-world networks are not completely random nor completely regular.

1.3.4 Degree and Degree Distribution

The concept of degree is the most fundamental character and measure of a node in a network, which may be defined in deferent ways. Here and throughout, the *degree* of node i in an undirected network is simply defined to be the number k_i of the edges directly connecting it to the other k_i neighbors. In a directed network, however, there is a distinction between *in-degree* and *out-degree*, describing the numbers of incoming edges and of outgoing edges, respectively. In most cases, a node of higher degree is "more important" than one of lower degree in a network, because it will have more significant influence on other nodes in the network, regarding dynamics, information flows, data traffic, etc. The average degree of a network is the average value of all such node degrees k_i over the entire network, and is denoted by $\langle k \rangle$.

Since in a network every node has a degree value, some large and some small while an isolated node has degree zero, the distribution of nodes with a certain degree (say, those nodes with the highest degree) in the network may be important and of great concern. This *degree distribution* is defined by a probability function, $P(k)$, which is the probability of a randomly-picked node that happens to have degree k, where each node has an equal probability to be picked:

$$\begin{aligned} P(k) = &\text{ Probability that a node has degree } k, \\ &\text{ where the node is picked at random uniformly} \end{aligned} \tag{1-5}$$

It can be easily seen that the degree distribution $P(k)$ of a fully connected regular network is a discrete delta function. For example, in a fully connected network of 100 nodes, each node has degree 99; therefore, $P(99)=1$ and $P(k)=0$ for all $k \neq 99$, i.e., for all k = 0, 1, 2, …, 98, 100, 101, 102, ...

In the ER random graph theory, it has been proved that the degree distribution of a completely random network is a Poisson distribution [21]:

Theorem 1-1 The degree distribution of a random graph satisfies

$$P(k) = \frac{\mu^k}{k!} e^{-\mu} \tag{1-6}$$

in which μ is the expectation value.

Proof. First, notice that the probability of obtaining exactly k edges, after N pairs of nodes have been randomly picked, is given by the limit of the binomial distribution

$$P(k \mid N) = \binom{N}{k} p^k (1-p)^{N-k}$$

and the expectation value is $\mu = Np$ and p is the probability that a pair of two randomly-picked nodes is connected with an edge. Viewing the above distribution as a function of μ rather than the network size and for a fixed value of p, one can rewrite the distribution as

$$P_\mu(k\mid N)=\frac{N!}{k!(N-k)!}\left(\frac{\mu}{N}\right)^k\left(1-\frac{\mu}{N}\right)^{N-k}$$

Thus, as $N\to\infty$, one has

$$P(k)=\lim_{N\to\infty}P_\mu(k\mid N)=\lim_{N\to\infty}\frac{N(N-1)\cdots(N-k+1)}{N^k}\frac{\mu^k}{k!}\left(1-\frac{\mu}{N}\right)^N\left(1-\frac{\mu}{N}\right)^{-k}=1\cdot\frac{\mu^k}{k!}\cdot e^{-\mu}\cdot 1$$

This completes the proof of the theorem. □

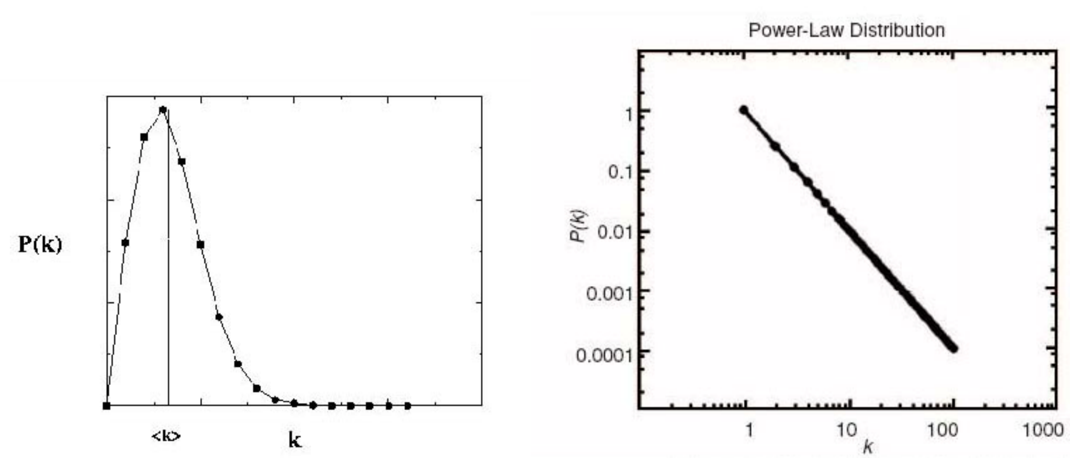
The Poisson distribution curve is shown in Fig. 1-7 (a). Clearly, the Poisson curve decays exponentially fast as k is moving away from the average value $\langle k\rangle$, at which the function attains the maximum value. Similar to the above delta distribution, this implies that a node with a too-large or a too-small degree k can hardly be found, or actually does not exist in the network. In other words, in such a random network every node has about the same degree, at least in theory. This kind of network is called a *homogeneous network*.

It has been observed, and can be easily verified by simulation, that as the network becomes less and less random (i.e., becomes more and more regular), the shape of the Poisson distribution curve will become narrower and narrower towards the delta function (completely regular). Most such distributions in between Poisson and delta distributions are described by a power-law function of the form

$$P(k)\sim k^{-\gamma} \tag{1-7}$$

Its logarithmic curve is shown in Fig. 1-7 (b), where γ is a constant determined by the given network. Clearly, the logarithmic curve of a power-law distribution decays linearly, much slower than the Poisson curve, which decays exponentially.

The above power-law distribution is more commonly called a *scale-free* distribution, for they have the scale-independent property described by the following theorem.



(a) Poisson distribution (b) Power-law distribution (log-log plot)

Fig. 1-7 Two typical distributions

Theorem 1-2 (*Scale-free property of power-law distributions*) Consider a probability distribution function $f(x)$. If, for any given constant a, there is a constant b such that the following "scale-free" property holds:

$$f(ax) = bf(x) \tag{1-8}$$

then, with the assumption of $f(1)f'(1) \neq 0$, the function $f(x)$ is uniquely determined by

$$f(x) = f(1)x^{-\gamma}, \quad \gamma = -f(1)/f'(1) \tag{1-9}$$

Proof. Let $x = 1$ in (1-8). Then, one has $f(a) = bf(1)$, so $b = f(a)/f(1)$; therefore,

$$f(ax) = \frac{f(a)f(x)}{f(1)}$$

Since this equality holds for arbitrary a and x, one may also consider a as a variable and take a derivative of the equality with respect to a, obtaining

$$\frac{df(ax)}{d(ax)}\frac{d(ax)}{da} = \frac{f(x)}{f(1)}\frac{df(a)}{da}$$

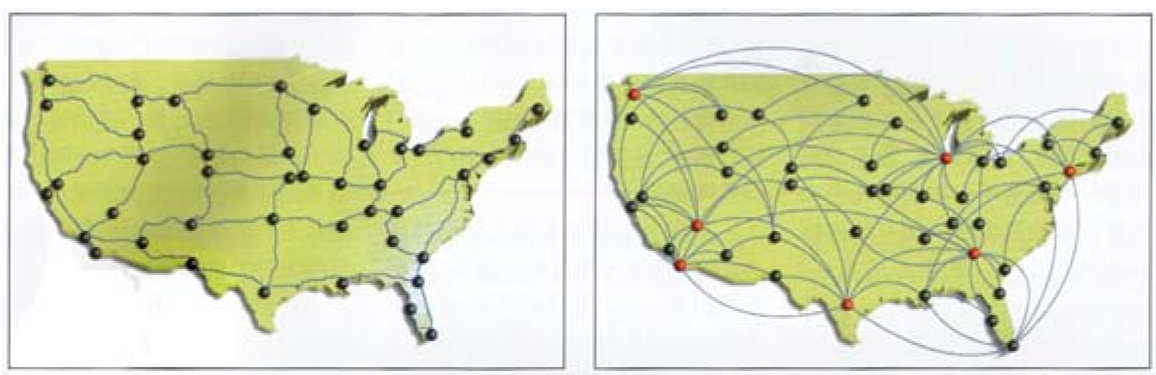
In particular, letting $a = 1$ gives

$$x\frac{df(x)}{d(x)} = \frac{f'(1)}{f(1)}f(x)$$

which has a unique solution

$$\ln f(x) = \frac{f(1)}{f'(1)}\ln x + \ln f(1)$$

This is equivalent to (1-9), completing the proof of the theorem. □



(a) Highway route map in USA (b) Airline route map in USA

Fig. 1-8 Comparison of a random network and a scale-free network [22]

In a large-scale network having a power-law degree distribution, particularly when $2 \leq \gamma \leq 3$, most nodes have very low degrees (i.e., with very few edges) and yet there are a few nodes having very high degrees (i.e., with many edges), as further verified and discussed in the following chapters of this text. This kind of complex networks is called *heterogeneous networks*, where the high-degree nodes are called *hubs*. A typical example of this kind is some USA airline route map, shown in Fig. 1-8 (b) [22], in which a few airports are hubs (e.g., New York, Chicago, Atlanta, Dallas, etc.), while the majority of airports especially those in small towns have relatively very few flights each day. In comparison, the USA highway system, shown in Fig. 1-8 (a), is *a homogeneous network*, simply because it is impossible for a particular city to have too many highways connecting it to other cities.

There is another useful degree distribution function of a network, called a *cumulative degree distribution function*, in the form of

$$P_k = \sum_{k'=k}^{\infty} P(k') \qquad (1\text{-}10)$$

This is the probability of a randomly-picked node that happens to have a degree not less than k. For a power-law distribution $P(k) \sim k^{-\gamma}$, the corresponding cumulative degree distribution function is also in a power-law form but has an exponent $\gamma - 1$ instead:

$$P_k \sim \sum_{k'=k}^{\infty} (k')^{1-\gamma} \sim k^{-(\gamma-1)} \qquad (1\text{-}11)$$

On the other hand, for an exponential distribution function $P(k) \sim e^{-k/\kappa}$, where $\kappa > 0$ is a constant, although its corresponding cumulative degree distribution function is also in an exponential form, it has the same exponent:

$$P_k \sim \sum_{k'=k}^{\infty} e^{-k'/\kappa} \sim e^{-k/\kappa} \qquad (1\text{-}12)$$

Notice that a power-law distribution function and an exponential distribution function have some similarities. To distinguish them, it amounts to noting that the logarithmic curve of the former is a straight line while the semi-logarithmic curve of the latter is a straight line. This can be easily verified mathematically.

Figure 1-9 [23] shows the cumulative degree distribution functions of six examples, among others, given in Table 1-3 [23]. In Fig. 1-9, the horizontal axes are the degree values (for the directed networks (b) and (c), they are in-degrees), and the vertical axes are cumulative degree distribution values. The curves in these figures correspond to: (a) a mathematical co-authorship network; (b) scientific citation index in 1981-1987, provided by the Institute for Scientific Information; (c) a subnet of 0.3-billion nodes of the WWW in 1999; (d) the AS-level Internet in April 1999, where AS stands for autonomous systems; (e) a power grid in the Western USA; (f) protein reactions of a yeast metabolic network. It can be seen that curves in (c), (d) and (f) follow power laws, since their logarithmic curves are straight lines; curve in (b) follows a power law only near the end; curve in (e) actually follows an exponential distribution (a semi-logarithmic curve); curve in (a) looks like a combination of two power-law distributions with different exponents.

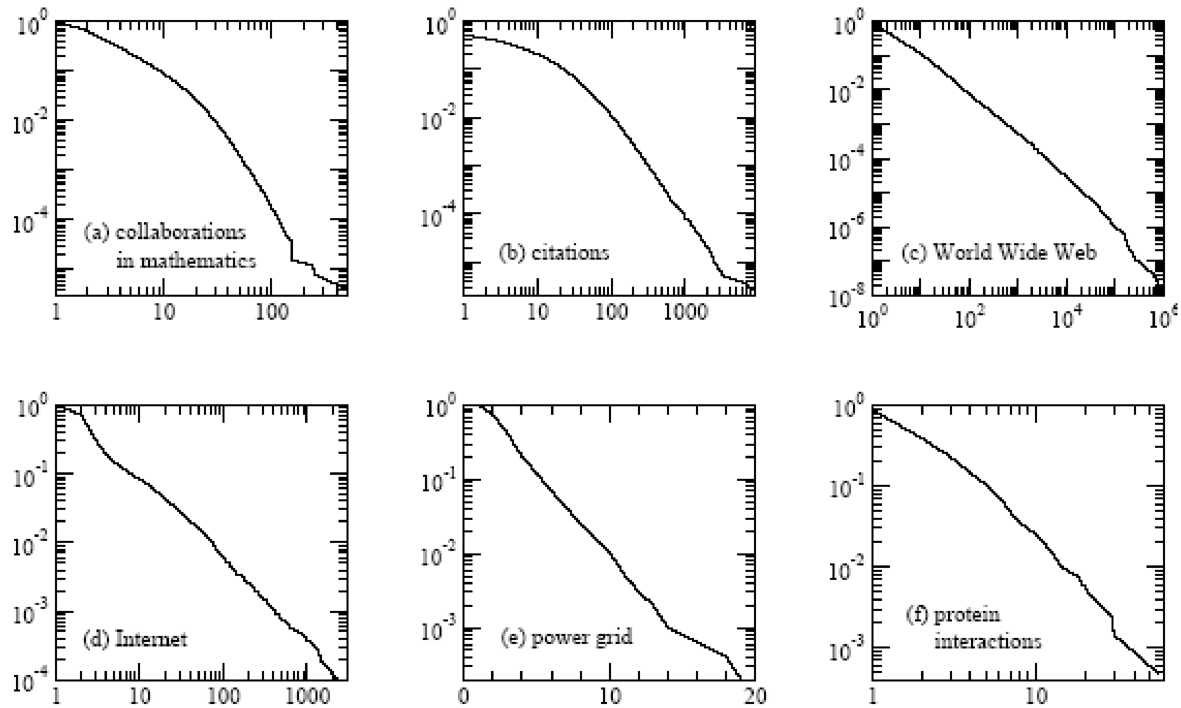


Fig. 1-9 Cumulative degree distributions of six examples given in Table 1-3 [23]

1.3.5 Statistical Properties of Some Real-World Complex Networks

Table 1-3 [23] lists the statistical properties of some real-world complex networks, where N is the number of nodes, M is the number of edges, $<k>$ is the average degree, L is the average path length, C is the clustering coefficient of a network, in which γ is the exponent of a power-law distribution function (for directed networks, both in- and out-degrees are given), and – means undefined or not applicable.

Table 1-3 Statistic data of some real networks [23]

	Network	Type	N	M	$< k >$	L	γ	C
Social science	Film actors	undirected	449913	25516482	113	3.48	2.3	0.78
	Company directors	undirected	7673	55392	14.4	4.6	–	0.88
	Math coauthorship	undirected	253339	496489	3.92	7.57	–	0.34
	Physics coauthouship	undirected	52909	245300	9.27	6.19	–	0.56
	Biology coauthorship	undirected	1520251	11803064	15.5	4.92	–	0.6
	Telephone call graph	undirected	47000000	80000000	3.16			
	E-mail messages	undirected	59912	86300	1.44	4.95	1.5/2.0	0.16
	E-mail addresses books	undirected	16881	57029	3.38	5.22	–	0.13
	Student relationships	undirected	573	477	1.66	16	–	0
	Sexual contacts	undirected	2810				3.2	
Information Science	WWW nd.edu	directed	269504	1497135	5.55	11.3	2.1/2.4	0.29
	WWW Altavista	directed	203549046	2.13E+09	10.5	16.2	2.1/2.7	
	Citation network	directed	783339	6716198	8.57		3.0/–	
	Roget's Thesaurus	directed	1022	5103	4.99	4.87	–	0.15
	Word co-occurrence	undirected	460902	1.7E+07	70.1		2.7	0.44
Technology	Internet (AS-level)	undirected	10697	31992	5.98	3.31	2.5	0.39
	Power grid	undirected	4941	6594	2.67	19	–	0.08
	Train routes	undirected	587	19603	66.8	2.16	–	0.69
	Software packages	directed	1439	1723	1.2	2.42	1.6/1.4	0.08
	Software classes	directed	1377	2213	1.61	1.51	–	0.01
	Electric circuits	undirected	24097	53248	4.34	11.1	3	0.03
	Peer-to-peer network	undirected	880	1296	1.47	4.28	2.1	0.01
Biology	Metabolic network	undirected	765	3686	9.64	2.56	2.2	0.67
	Protein network	undirected	2115	2240	2.12	6.8	2.4	0.07
	Marine food web	directed	135	598	4.43	2.05	–	0.23
	Freshwater food web	directed	92	997	10.8	1.9	–	0.09
	Neural network	directed	307	2359	7.68	3.97	–	0.28

Problems

1-1 Consider the simple network shown in Fig. 1-10. Compute its average path length L, diameter D, cluster coefficient C, and average degree $\langle k \rangle$.

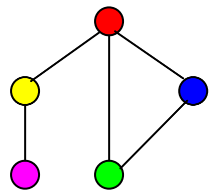


Fig. 1-10 A simple graph

1-2

(i) A network with a large clustering coefficient may have a long or a short average path length. Show (a) an example of a network with a *large* clustering coefficient (i.e., $C = 1$) and a *short* average path length (i.e., $L = 1$); (b) an example of a network with a relatively *large* clustering coefficient and also a relatively *long* average path length.

(ii) A network with a small clustering coefficient may have a short or a long average path length. Show (i) an example of a network with a *small* clustering coefficient (i.e., $C = 0$) and a *short* average path length (i.e., $L = 1$); (b) an example of a network with a relatively *small* clustering coefficient and a relatively *long* average path length.

1-3 In walking through the graph shown in Fig. 1-11, can one start from a certain node to go through all edges once and once only, and finally return to the starting point? Why?

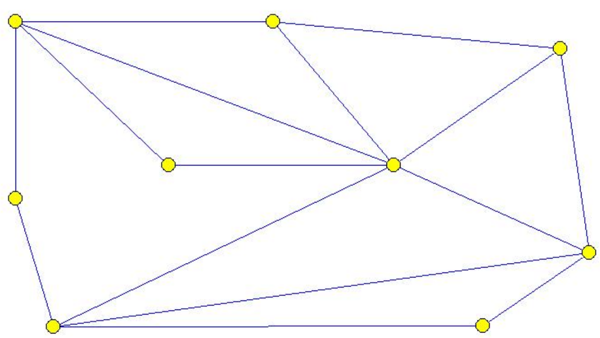


Fig. 1-11 A simple graph

1-4 In a laboratory shown by Fig. 1-12, can you walk through every door once and once only, and finally return to the starting point? If so, how? If not, why?

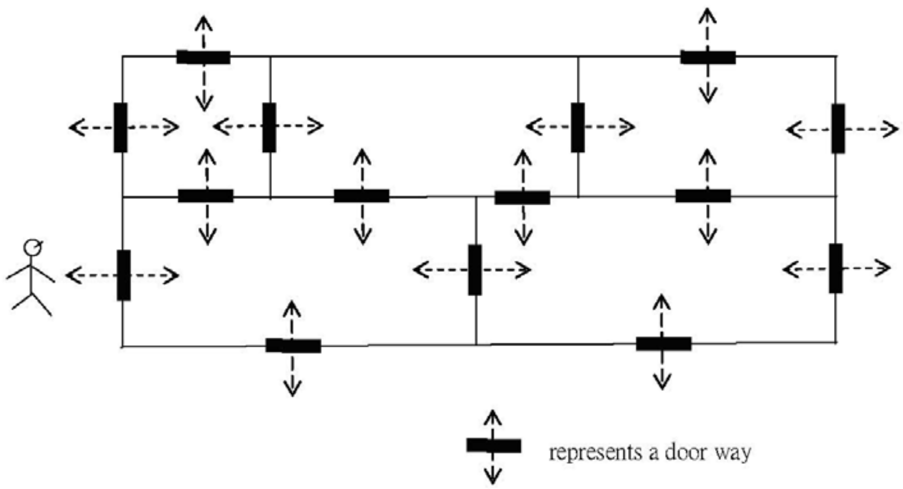


Fig. 1-12 A lab with many doorways

1-5 Write a computer program to simulate the generation of an ER random graph with 100 nodes, and plot several graphs so obtained (you may stop the program after 5,000 steps in each run).

1-6 Verify that the probability of obtaining an ER random graph with N nodes and M edges is equal to $p^M (1-p)^{N(N-1)/2-M}$, where p is the probability that two randomly picked nodes will be connected by an edge.

1-7 Verify that the total possibility of obtaining exactly k edges, after N pairs of nodes have been randomly picked, is given by the following binomial distribution:

$$P(k \mid N) = \binom{N}{k} p^k (1-p)^{N-k}$$

1-8 Name and briefly describe a few real-world examples of random graphs, small-world networks, and scale-free networks.

1-9 Give some real-world examples of strong interactions generated by weak connections, namely, a long-range connection leads to a stronger interaction between two nodes than those short connections of neighboring nodes.

1-10 **(i)** A network with a large clustering coefficient may have a long or a short average path length. Show an example of a network that has a large clustering coefficient (i.e., $C = 1$) and a short average path length (i.e., $L = 1$). Also show an example of a network that has a large clustering coefficient with a long average path length.

(ii) A network with a small clustering coefficient may have a long or a short average path length. Show an example of a network that has a small clustering coefficient (i.e., $C = 0$) and a short average path length (i.e., $L = 1$). Also show an example of a network that has a small clustering coefficient with a long average path length.

1-11 Consider the simple network shown in Fig. 1-11. Compute the average path length L, diameter D, cluster coefficient C, and average degree $\langle k \rangle$ of this network.

1-12 Consider the simple networks shown in Fig. 1-13 (a)-(f). Compute the average path length L, diameter D, cluster coefficient C, and average degree $\langle k \rangle$ of each network.

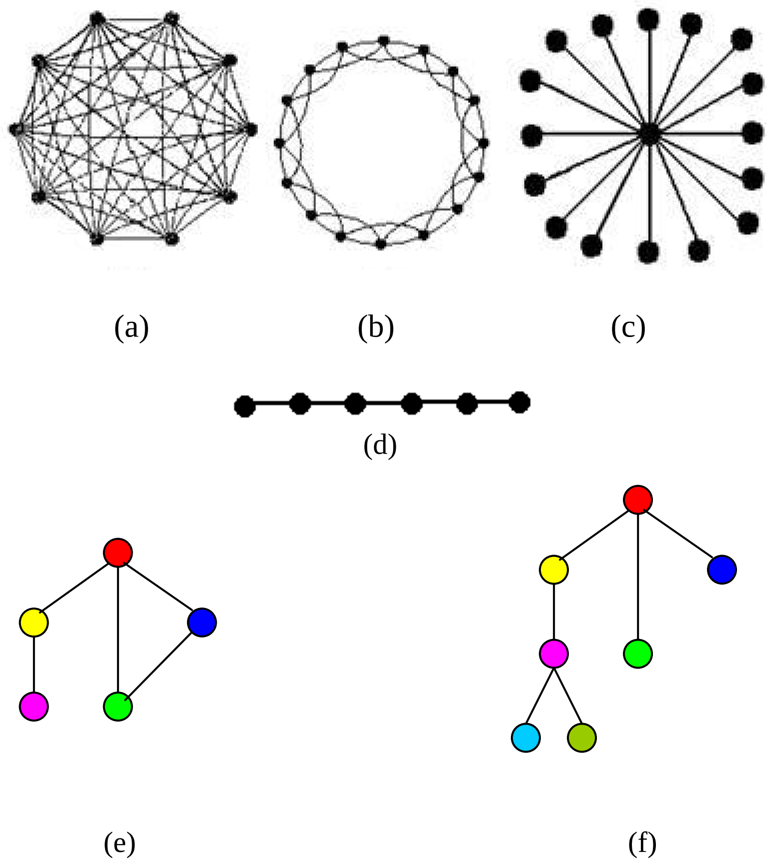
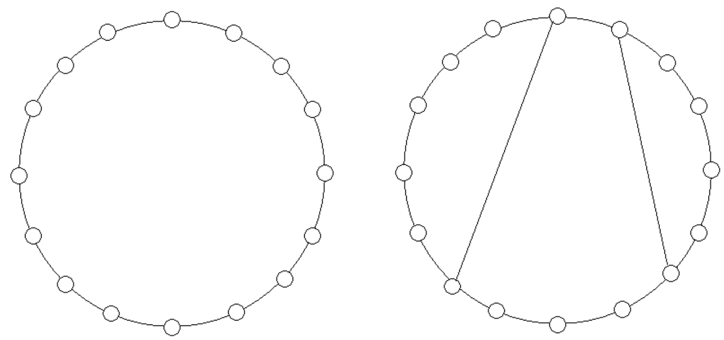


Fig. 1-13 Four simple examples of networks

1-13 Verify by some simulations that as the network becomes less and less random (i.e., becomes more and more regular), the shape of the Poisson distribution curve will become narrower and narrower towards the delta function (completely regular).

1-14 Consider the two networks shown in Fig. 1-14 (a) and (b), respectively. Compute their average path lengths L, diameters D, cluster coefficients C, and average degrees $\langle k \rangle$. Explain why the figure shown in (b), with only a few long-range edges, can gain a more prominent small-world feature than the one shown in (a).



(a) (b)

Fig. 1-14 A few long-range edges may enhance the small-world network feature of a network

References

[1] Barabási A-L. *Linked: The New Science of Networks.* Massachusetts: Persus, 2002
[2] Watts D J. The 'new' science of networks. *Annual Review of Sociology*, 2004, 30: 243-270
[3] Strogatz S H. Exploring complex networks. *Nature*, 2001, 410: 268-276
[4] Kaneko K. *Coupled Map Lattices.* Singapore: World Scientific, 1992
[5] Chua L O. *CNN: A Paradigm for Complexity.* Singapore: World Scientific, 1998
[6] http://www-history.mcs.st-andrews.ac.uk/Extras/Konigsberg.html
[7] Erdös P, Rényi A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*,1960, 5, 17-60
[8] Bollobás B, Riordan O M. Mathematical results on scale-free random graphs, in *Handbook of Graphs and Networks: From the Genome to the Internet*, Bornholdt S and Schuster H G (eds.), Weinheim: Wiley-VCH, 2003, pp. 1-34.
[9] Hoffman P. *The Man Who Loved Only Numbers*, New York: Hyperion, 1998
[10] Schechter B. *My Brian is Open: The Mathematical Journeys of Paul Erdös*, New York: Touchstone, 1998
[11] Milgram S. The small world problem. *Psychology Today*, May 1967, 60 – 67
[12] http://www.cs.virginia.edu/oracle/
[13] http://personalwebs.oakland.edu/~grossman/erdoshp.html
[14] Borosh I, Chui C K, Erdös P. On changes of signs in infinite series. *Anal. Math.,* 1978, 4(1): 3-12
[15] Chui C K, Chen G. *Kalman Filtering with Real-Time Applications.* Berlin: Springer-Verlag, 1st ed. 1987, 4th ed. 2009.
[16] Li X, Wang X F, Chen G. Pinning a complex dynamical network to its equilibrium. *IEEE Transactions on Circuits and Systems—I*, 2004, 51: 2074-2087
[17] http://smallworld.columbia.edu/
[18] Granovetter M. The strength of weak ties. *American Journal of Sociology*, 1973, 78 (6): 1360-1380
[19] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks. *Nature*, 1998, 393(6684): 440-442
[20] Barabási A-L, Albert R. Emergence of scaling in random networks. *Science*, 1999, 286(5439): 509-512
[21] Bollobas B. *Random Graph*. 2nd ed. Cambridge: Cambridge University Press, 2001.
[22] Barabási A-L, Bonabeau E. Scale-free networks. *Scientific American*, May 2003, 50-59
[23] Newman M E J. The structure and function of complex networks. *SIAM Review,* 2003, 45: 167-256
[24] Albert R, Barabási A-L. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 2002, 74: 47-97
[25] Dorogovtsev S N, Mendes J F F. Evolution of networks. *Advances in Physics*, 2002, 51: 1079-1181
[26] Wang X F. Complex networks: Topology, dynamics and synchronization. *Int. J. Bifurcation and Chaos*, 2002, 21(5): 885-916
[27] Wang X F, Chen G. Complex networks: Small world, scale free and beyond. *IEEE Circuits and System Magazine*, 2003, 3(1): 6-21
[28] Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D-U. Complex networks: Structure and dynamics. *Physics Reports*, 2006, 424: 175-308
[29] Da L, Costa F, Rodrigues F A, Travieso G, Villas Boas P R. Characterization of complex networks: A survey of measures. *Advances in Physics*, 2007, 56(1): 167-242
[30] Newman M. The physics of networks. *Physics Today,* Nov 2008
[31] Huberman B A. *The Laws of the Web: Patterns in the Ecology of Information.* Massachusetts: MIT Press, 2001

[32] Buchanan M. *Nexus: Small Worlds and the Groundbreaking Science of Networks.* New York: Norton, 2002
[33] Watts D J. *Six Degrees: The Science of a Connected Age.* New York: Norton, 2003
[34] Watts D J. *Small Worlds: The Dynamics of Networks between Order and Randomness.* New Jersey: Princeton University Press, 1999
[35] Bornholdt S, Schuster H G. (ed.) *Handbook of Graphs and Networks: From the Genome to the Internet.* Germany: Wiley-VCH, 2003
[36] Dorogovtsev S N, Mendes J F. *Evolution of Networks: From Biological Nets to the Internet and WWW.* New York: Oxford University Press, 2003
[37] Pastor-Satorras R, Vespignani A. *Evolution and Structure of the Internet: A Statistical Physics Approach* Cambridge University Press, 2004
[38] Wang X F, Li X, Chen G. *Complex Networks: Theories and Applications* (in Chinese), Beijing: Tsinghua University Press, 2006