Traffic

12

The central function of a great number of networks is to provide convenient routes for the transportation of humans, objects, energy, information, and other items between their nodes. The goal is to obtain dense but free-flowing traffic. In this lecture we mainly discuss traffic and traffic congestion in complex networks.

12.1 Traffic in the Internet

The Internet is based on packet switching, which enables robust data transport.[1] In the process of transmission, (i) files are cut into packets; (ii) the packets are submitted (routed) 'independently' to their destinations through optimal routes; (iii) finally, at their destinations, the packets are reassembled into the original sequences. This technology allows networks to efficiently route information around failed nodes and links. In addition, packet switching distributes the flows of data over the network, and uses network resources fully. The header of each packet contains full addressing information, so that routers forward incoming packets by using this information and their routing tables. A routing table is the router's dynamically updated individual database of routes. Based on its routing table, a router forwards a packet to its neighbour which is expected to be on the optimal route to the destination.

[1] For a more detailed introduction to the basics of Internet traffic, see the paper of Willinger, Govindan, Jamin, Paxon, and Shenker [182] and a brief survey of concepts written by Smith [164].

The first theoretical studies of packet switching communications were made in early 1960s (Leonard Kleinrock, MIT), but the fundamental suite of communications protocols was written by V. G. Cerf and R. E. Kahn considerably later, between 1972–1974 [53]. These were the Transmission Control Protocol/Internet Protocol (TCP/IP). The basic protocol—IP—enables addressing and forwarding of packets, but IP is not sufficient for robust networking. TCP enables flow control and recovery from lost packets. The TCP/IP suite was adapted between 1980–1983.

Thus, traffic in the Internet is, in essence, a flow of packets from hosts–sources to hosts–destinations governed by Internet protocols. Of the rich set of features that this highly variable flow demonstrates, here we touch upon only two—self-similarity and fluctuations.

The self-similarity of Internet traffic is an incredibly widely discussed issue. It was believed until the first half of the 1990s that packet flow in the Internet is a *Poisson process*. This is a basic random process—a chain of uncorrelated events occurring at some average rate. For example, at each discrete moment, a random variable takes the value 0 with

probability $1-p$ and the value 1 with probability p. Then the number of ones in some given time interval is distributed according to a Poisson distribution, and the number of zeros between two subsequent intervals is distributed according to an exponential distribution. If we coarsen this Poisson flow by binning—summing the ones within sufficiently long intervals of length T, then the resulting flow looks much smoother. The relative amplitude of fluctuations, $\sqrt{pT}/pT$, quickly decreases with T, and the flow is smoothed out. In 1993, Will Leland, Murad Taqqu, Walter Willinger, and Daniel Wilson thoroughly investigated the flow of packets collected from a network link and found that the real flow is in remarkably stark contrast to the Poisson process [118]. They discovered that Internet traffic viewed over different time scales of a few (in their case, four) orders of magnitude is highly bursty.[2] Over all these time scales, the flow $f(t)$ (the numbers of packets recorded during time intervals of a given length—'bins') looks 'self-similar' in the following sense. Scaling the time variable by a factor a rescales the flow:

$$f(t) = a^{-H} f(at)\,. \tag{12.1}$$

Here H is the so-called *Hurst exponent*. For Internet traffic, typically, $H \approx 0.8$. One can easily show that this self-similarity is equivalent to a slow power-law decay of the autocorrelation function:

$$\langle [f(t) - \langle f \rangle][f(t+\tau) - \langle f \rangle] \rangle \propto \tau^{-\tilde{\beta}}\,, \tag{12.2}$$

where exponent $\tilde{\beta}$ is related to the Hurst exponent: $\tilde{\beta} = 2(1-H)$, so $\tilde{\beta} \approx 0.4$.

What is the nature of this unexpectedly slow decay of correlations? The existing explanations reveal two contrasting views of the Internet—the network engineer's view and the physicist's one. The most widely accepted engineer's (computer scientist's) explanation is based on the broad distribution of users' sessions, for example, HTTP connections [182]. Empirical data demonstrate that the sizes of sessions (the numbers of packets in sessions or the lengths of transmitted files) are distributed according to a power law. This fact (supplied by an additional assumption that the sessions arrive uniformly at random) directly leads to the self-similar traffic of packets.[3]

Physicists usually propose a quite different explanation. In their abstract models, they observe that on the verge of congestion, the flow of packets is power-law correlated, as is normal for a critical phenomenon. If a real network functions most efficiently, which corresponds to the highest possible traffic flow without congestion, then the traffic should be indeed critical and so, self-similar.

Computer scientists explain why the physicists' theories are not applicable to the Internet: 'self-similar scaling has been observed in networks with low, medium, or high loads' and not at some critical load [182].[4] The physicists' traffic models are too abstract and do not account for the real complexity of the Internet, ignoring, for example, even basic details of Internet protocols. This criticism pinpoints the common feature of

[2] 'Bursty' means 'has a relatively large variance'.

[3] Exponent $\tilde{\beta}$ in eqn (12.2) is expressed in terms of the exponent $\tilde{\alpha}$ of the size distribution of sessions: $\tilde{\beta} = \tilde{\alpha} - 2$. Typically, $\tilde{\alpha} = 2.2$–2.4.

[4] We will return to this criticism in Lecture 14.

the physicists' theories of the Internet: physicists essentially ignore that the Internet is a 'highly engineered' system. They simply believe that this fact is not of primary importance. According to Reginald Smith—a network engineer, for physicists, 'the Internet is a complex dynamical system whose macroscopic properties are based on general mechanisms of statistical mechanics', while for network engineers, 'the Internet is a complex engineering system whose behaviour is determined by its protocols, users, and functions' [164]. We suggest, however, that in some form, the 'protocols, users, and functions' can be included into physicists' models. Why not?

Traffic variance (the amplitude of fluctuations) is an important characteristic of bursty traffic. Let the average flux measured at node i of a network be $\langle f_i \rangle$ and its standard deviation be $\sigma_i = \sqrt{\langle (f_i - \langle f_i \rangle)^2 \rangle}$. The question is: how are these quantities related? In 2004, Marcio de Menezes and László Barabási analysed data on traffic at a large number of nodes in the Internet and in a few other networks, which they collected over long time periods [68]. These empirical data were plotted as σ_i versus $\langle f_i \rangle$. It turned out that for the Internet, this set of points for different nodes could be fitted by a square root function: $\sigma \propto \langle f \rangle^{1/2}$. For traffic on the other investigated networks (for the number of visits of web sites, for the stream flow of rivers, for traffic on highways, and for flows within microprocessors), they found a proportional dependence $\sigma \propto \langle f \rangle$. The researchers suggested that the square root dependence is valid if flow fluctuations have an intrinsic nature, while proportional dependence is present if the variations of traffic are induced by some fluctuating external force. Later studies showed, however, that the empirical curves $\sigma_i(\langle f_i \rangle)$ are usually between these two limiting cases, the results depend on the time interval of the measurements, and so the situation is more complicated than was believed originally [124].

12.2 Congestion

The total traffic in the Internet grows even faster than the number of hosts, which in its turn grows exponentially with time. Traffic becomes more and more dense, which increases the risk of congestion, especially for highly connected nodes.[5] Let us discuss very basic features of this phenomenon. Suppose that a network has N nodes which all permanently send packets to each other, each node injects packets at a rate λ. If we assume that the routes are along the shortest paths between the nodes, then the flow through a node is proportional to the number of shortest paths which run through this node, or actually, to the betweenness centrality B_i of this node. The node with the highest betweenness centrality $B_{\max}$ will be the first congested. So it is easy to see that congestion emerges at the rate

$$\lambda_c = \frac{N-1}{B_{\max}} \,. \tag{12.3}$$

[5] On average, traffic increases by 50–60% every year [1].

This simple relation can be generalized to more general protocols which allow routes to depart from the shortest paths [168]. Importantly, the formula relates congestion and the architecture of a network. It is clear, in particular, that homogeneous networks have higher congestion thresholds than networks with hubs.

The natural method of delaying congestion, is to send packets around hubs. For example, we can use 'a random walk protocol', that is make all the packets random walkers. (Note, however, that this protocol is extremely inefficient because of long delivery times.[6]) Furthermore, we can increase the congestion threshold, by slightly modifying these random walks. Let a node forward a packet to its neighbour i chosen with probability proportional to q_i^μ [179,95]. Here q_i is the number of connections of this neighbour and μ is a given exponent for this random walk. Recall formula (12.1) for the stationary probability of finding a walker at a node of degree q for standard random walks: $p_{\rm fin}(q) \propto q$. For the modified random walk, we similarly have $p_{\rm fin}(q) \propto q^{1+\mu}$. The highest congestion threshold is apparently realized when $p_{\rm fin}$ is independent of q, which takes place at $\mu = -1$. Thorough calculations showed that this is indeed the case [179, 95].

[6] We do not discuss ways of making random walk protocols more efficient [175].

We can avoid congestion in another way. In *adaptive routing* protocols, instead of avoiding hubs, routed packets avoid nodes with long queues. In 2005, Pablo Echenique, Jesus Gómez-Gardeñes, and Yamir Moreno proposed a simple model of adaptive routing [84]. In this model, all nodes send packets to each other at a rate λ. Each node also functions as a router. Let n_i be the number of packets at node i at a given moment, which is a queue length. The buffers are assumed to be infinite. The nodes forward packets taking into account the queue lengths at their neighbours, but also trying to select the shortest routes to destinations. A packet is sent from a node to its neighbour j, which has the minimum value of the quantity:

$$\delta_j = h\ell_{jt} + (1-h)n_j \,. \tag{12.4}$$

Here n_j is the queue length at node j, ℓ_{jt} is the length of the shortest path from node j to the target t of the packet, and the parameter h of the model is a number between 0 and 1.

Quantitatively, congestion is characterized by the following order parameter [11]:

$$\rho = \frac{\text{\# of packets that failed to reach targets during the observation}}{\text{the total number of packets injected during this time period}} \tag{12.5}$$

or, equivalently, by

$$\rho = \frac{N_p(t+\Delta t) - N_p(t)}{N\lambda\Delta t} \,. \tag{12.6}$$

Here N is the network size, $N_p(t)$ is the number of packets in the network at time t, and Δt is the time of observation. In the jammed phase, the number of packets in a network grows linearly with time, which

leads to time-independent ρ. Figure 12.1 shows the results of numerical simulations of this model, namely the dependencies of the order parameter on injection rate. The researchers used the map of a real Internet network. At $h=1$, the model provides shortest-path routing. In this limiting case, the transition to the congested phase was found to be continuous, without a jump. The curves become discontinuous at smaller values of h, and the congestion threshold significantly increases. Note that this transition resembles the birth of a k-core, compare Figs. 12.1 and 6.9 (c).

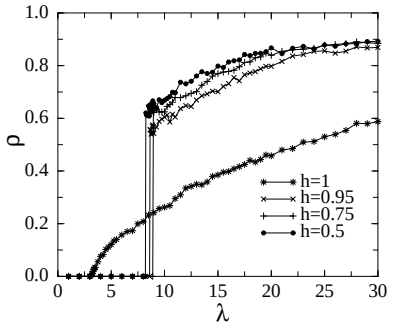

Fig. 12.1 Order parameter ρ, eqn (12.6), as a function of a packet injection rate λ in the Echenique–Gómez-Gardeñes–Moreno model of congestion. Adapted from [84].

12.3 Cascading failures

Wide-spread power outages in large electricity grids are among the most financially devastating events that occur in man-made systems. The latest large 2005 Java–Bali Blackout affected about 100 million people; the famous Northeast Blackout of 2003 affected 55 million people. The failure of a single element in a power grid triggers a cascade of failures across the entire grid or a large part of it. How can this happen? Each element in a general transportation system has a limiting load capacity. If a transportation network is fully or nearly fully loaded, then the failure of a single node or link overloads a few other elements, they fail, which leads to new overloads and failures, and so on.

In 2002, Adilson Motter and Ying-Cheng Lai proposed a simple model of these cascade processes [131]. In the Motter–Lai model, all nodes in a network permanently, at equal rate, send goods (or data packets, etc.) to each other along the shortest paths. Therefore the load of a node in this system is proportional to the number of shortest paths in the network that pass through this node. Recalling the definition of betweenness centrality, we can use this quantity as a load. Let $B_{0,i}$ be the betweenness centrality (load) of node i in an undamaged network. Motter and Lai assumed that each node has its limiting capacity, which is the maximum load the node will tolerate:

$$c_i = (1+\alpha)B_{0,i}\,. \tag{12.7}$$

Here $\alpha \geq 0$ is a 'tolerance parameter', which shows how much an initial load can be exceeded without failure. If the betweenness centrality b_i of a node exceeds its capacity c_i, this node is immediately removed. A cascading failure in this model occurs in the following way.

(i) Compute the betweenness centralities $B_{0,i}$ and so capacities c_i of all nodes in a network.

(ii) Remove a chosen node from the network.

(iii) Compute all the betweenness centralities $B_{1,i}$ in the resulting network.

(iv) Delete all the overloaded nodes where $B_{1,i} > c_i$.

(v) Repeat (iii)–(iv) until no overloaded nodes remain.

[7] Note that there is no self-averaging here. For each of the three types of initial failures, we have to average over all starters. In the simulations, it was sufficient to average over a few triggers.

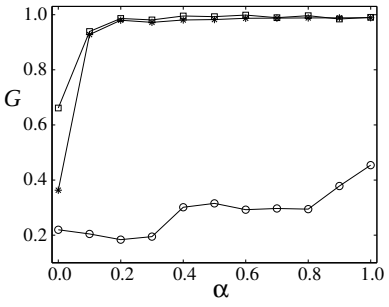

Fig. 12.2 Modelled cascading failure in the Western US power transmission grid (4941 nodes with, on average, 2.67 connections). The curves show the relative size of a giant connected component remaining after the cascading failure versus the tolerance parameter α. The cascades are triggered by the removal of a single node selected: (i) uniformly at random ($\Box$), (ii) among nodes of highest degrees ($*$), and (iii) among nodes with highest loads ($\circ$). Adapted from Motter and Lai [131].

[8] These were recursive scale-free networks with degree-distribution exponent $2 < \gamma \leq 3$. The triggers were chosen uniformly at random.

In numerical simulations, Motter and Lai studied a number of loopy undirected networks. The main quantity of interest was the relative size of a giant connected component that remains in a network after a cascading failure: $G = N'/N$, where N' is the size of the resulting giant component, and N is the total number of nodes in the original network before the failure. The researchers, in particular, investigated the dependence of this quantity on the tolerance parameter α. It is clear that zero α guarantees the complete collapse of a network, so the question is: how rapidly does G increase with α? Note that the size of a cascade failure depends significantly on a starting node. Figure 12.2 shows three distinct curves $G(\alpha)$ for three kinds of triggers—first failed nodes: for a node selected uniformly at random, for a node chosen from among the most connected nodes, and, finally, for a node chosen from among the most loaded nodes in an original network.[7] The figure shows that, as is natural, the largest cascades are triggered by failed nodes with the highest loads. Later, comprehensive numerical simulations of large networks uncovered more detailed features of cascading failures [117].[8] It has been found that there is a threshold tolerance, $\alpha_c > 0$, above which all cascading failures are finite, $G = 1$ (compare with the curve for uniformly random triggers in Fig. 12.2 where $G \approx 1$ for $\alpha > 0.2$). Importantly, at this threshold, the size distribution of cascades is power-law, which shows that this point is critical. That is, the Motter–Lai model has a phase transition to the state with giant avalanches.

Interacting systems on networks

13

In this lecture we consider systems of interacting agents placed on networks. The agents—spins, oscillators, interacting individuals, etc.—occupy the nodes of networks and interact with each other through network links. In more complicated situations, these agents, in turn, influence their network substrates, and so the pair—a network and the system of agents—co-evolve. We discuss unusual phenomena in these cooperative systems on complex networks.

13.1 The Ising model on networks

In 1924 the young scientist Ernest Ising (1900–1988) received a Ph.D. in physics from the University of Hamburg. Ising's Ph.D. thesis contained a study of the basic model of a ferromagnet, proposed to Ising by his renowned supervisor, Wilhelm Lenz (1888–1957). This famous reference model—now known as the Ising model—will help us to understand the main features of interacting systems placed on various substrates. The Ising model describes interactions between classical spins ($s_i = \pm 1$) in an idealized ferromagnet. The energy of a given configuration of spins in this model (physicists call it Hamiltonian) is

$$\mathcal{H} = -\sum_{i,j} J_{ij} s_i s_j - H \sum_i s_i \,, \tag{13.1}$$

where the sum is over all pairs of nearest-neighbouring spins, and H is an applied magnetic field. The couplings $J_{i,j}$ are positive, so the 2-degenerate ground state of the model is ferromagnetic: all spins in the system are of the same sign—either plus or minus. When the thermal energy $k_B T$ exceeds the energy of coupling of a spin to its z nearest neighbours, $z\overline{J}$, this ferromagnetic order is destroyed. Here, $\overline{J}$ is the mean coupling. The system enters the paramagnetic phase, where each spin frequently flips, so that the average value of each spin $\langle S_i \rangle = M_i$ and the macroscopic magnetic moment are zero. The critical temperature, separating ferromagnetic and paramagnetic phases is estimated as $T_c \sim z\overline{J}/k_B$, where k_B is the Boltzmann constant.

Let us start from the Ising model on the simplest infinite-dimensional system—on the fully connected graph of N nodes, where N approaches infinity. For simplicity, assume that no external field is present. Thermal fluctuations force each spin to flip from time to time and result in

$|M_i(T)| < 1$. Variations in the average values of spins $\langle S_i \rangle = M_i(T)$ with temperature essentially characterize the cooperative behaviour of the system. We can easily obtain equations for $M_i(T)$ using a mean-field approach. Notice that each spin i is in effect in the field $H_i = \sum_{j \neq i} J_{ij} s_j \approx \sum_{j \neq i} J_{ij} M_j$. Here the second equality is what is called the mean-field approximation. In our system, thanks to the infinite number of nearest neighbours of a spin, this approximation is apparently exact. As a result, we obtain the equations for $M_i(T)$:

$$M_i = \frac{e^{\beta H_i} - e^{-\beta H_i}}{e^{\beta H_i} + e^{-\beta H_i}} = \tanh\left[\beta \sum_{j \neq i} J_{ij} M_j\right], \tag{13.2}$$

where $\beta = 1/k_B T$. These equations can easily be solved. For a homogeneous ferromagnetic system, we set $J_{ij} = J/N > 0$ to get a finite critical temperature in the limit $N \to \infty$. In this case, we easily find the critical temperature $T_c = J/k_B$ and the relative macroscopic magnetic moment $M(T) = \sum_i M_i(T)/N$ near the critical point:

$$M(T) \propto \sqrt{T_c - T}\,. \tag{13.3}$$

Singularities of this kind correspond to continuous phase transitions. This square root is one of two possible critical peculiarities for the order parameter in traditional mean-field theories of second-order phase transitions.[1]

[1] The second one was discussed in Lecture 6: for percolation problems, the traditional mean field theory gives the size of a percolation cluster $S \propto p - p_c$.

The same critical singularity and a similar critical temperature were found in the ferromagnetic Ising model on sparse classical random graphs. What about complex networks? We can apply eqn (13.3) even to these complicated situations. Our hope is that mean-field theories should work well for infinite-dimensional systems. Consider the ferromagnetic Ising model on top of a sparse uncorrelated graph with a given degree sequence $q_1, q_2, \ldots, q_N$. Make the following *annealed network approximation* [26]. Substitute the original random network with a fully weighted connected graph with link weights $q_i q_j / \langle q \rangle N$. In other words, substitute the original network with an effective medium, where the sum of couplings for each node i has the same value Jq_i as in the original graph. Equivalently, force the links of the network to jump frequently between different pairs of nodes while keeping a given set of values Jq_i. 'Frequently' here means that links jump more frequently than spins flip. In that sense, the set of links is 'annealed'. This approximation provides a surprisingly good description of cooperative models on sparse uncorrelated networks. In this lecture we focus on these complex networks. For the resulting fully connected non-random graph, eqn (13.2) gives

$$M_i = \tanh\left[\frac{\beta J q_i}{\langle q \rangle N} \sum_j q_j M_j\right] = \tanh[\beta J q_i \mathcal{M}]\,. \tag{13.4}$$

A weighted magnetic moment $\mathcal{M} = \sum_j q_j M_j / \langle q \rangle N$ is a solution of the following equation:

$$\mathcal{M} = \sum_q \frac{q P(q)}{\langle q \rangle} \tanh[\beta J q \mathcal{M}]\,. \tag{13.5}$$

Here $P(q)$ is the degree distribution of the original random network. Interestingly, it is not M but rather the weighted moment $\mathcal{M}$ that plays the role of the order parameter in this problem. In simple terms, $J\mathcal{M}$ is an effective field acting on a spin from its nearest neighbour in this network. To obtain eqn (13.5), substitute the expression for M_i from eqn (13.4) into the definition of $\mathcal{M}$ and use the equality $\sum_i f(q_i) = \sum_q P(q)f(q)$. To find how average spins and the macroscopic magnetic moment depend on temperature, we should first solve eqn (13.5) and then substitute the solution $\mathcal{M}$ into eqn (13.4).

Equation (13.5) is a key point in our problem. Recall a similar eqn (6.2) for the order parameter x for the emergence of a giant connected component in uncorrelated networks. Importantly, in both of these equations, the nearest-neighbour degree distribution $qP(q)/\langle q\rangle$ appears in the sum on the right-hand side. We explained this factor for the giant connected component problem. Moreover, for arbitrary pairwise interactions, the description should apparently involve the degree distributions for the ends of a link, the same $qP(q)/\langle q\rangle$, as in eqn (13.5). The presence of this factor crucially increases the role of highly connected nodes in these systems, especially for various critical phenomena.

13.2 Critical phenomena

From eqn (13.5), we readily find the critical temperature,

$$k_B T_c = J\frac{\langle q^2\rangle}{\langle q\rangle}\,. \tag{13.6}$$

This formula is approximate. However, agreement with a rigorously obtained expression is more than satisfactory.[2] Consequently, T_c approaches infinity, if the second moment of the degree distribution diverges. In particular, for scale-free networks, thermal fluctuations cannot destroy the ferromagnetic order at any finite temperature if the degree distribution exponent γ is equal to or less than 3 [7]. In other words, hubs support cooperative ordering in these systems. This infinite critical temperature in a cooperative model on a sparse network is in a sharp contrast to what we know for lattices. This is similar to percolation—hubs prevent the destruction of a giant connected component by randomly damaging a network.

When the fourth moment $\langle q^4\rangle$ is finite (that is $\gamma > 5$), expand the hyperbolic tangent $\tanh[\beta Jq\mathcal{M}] \approx \beta Jq\mathcal{M} - (\beta Jq\mathcal{M})^3/3$ in the sum on the right-hand side of eqn (13.5). The resulting equation readily gives the same square-root critical singularity (13.3) as the classical random graphs and lattices. A new, non-traditional kind of critical singularity emerges if exponent $3 < \gamma \le 5$. For these heavy-tailed degree distributions, we cannot expand $\tanh[\beta Jq\mathcal{M}]$ in eqn (13.5). Instead, we do the following. We add and subtract $\mathcal{M}T_c/T = \mathcal{M}\beta J\langle q^2\rangle/\langle q\rangle$ from the

[2] This approximate result is close to the exact value

$$k_B T_c = \frac{2J}{\ln[\langle q^2\rangle/(\langle q^2\rangle - 2\langle q\rangle)]}$$

obtained for uncorrelated networks (see, Refs. [75, 77]).

right-hand side of this equation:

$$\mathcal{M} = \mathcal{M} \sum_q \frac{qP(q)}{\langle q \rangle} \beta J q + \mathcal{M} \sum_q \frac{qP(q)}{\langle q \rangle} \frac{\tanh[\beta J q \mathcal{M}] - \beta J q \mathcal{M}}{\mathcal{M}} \,. \quad (13.7)$$

The second sum on the right-hand side in this equation is a singular function of $\mathcal{M}$. Indeed, we can estimate the sum as follows:

$$\beta J \frac{\langle q^2 \rangle}{\langle q \rangle} - 1 = \sum_{q \sim 1/\mathcal{M}}^{\infty} \frac{qP(q)}{\langle q \rangle} \beta J q \propto \int_{1/\mathcal{M}}^{\infty} dq\, q^{2-\gamma} \propto \mathcal{M}^{\gamma-3} \,, \quad (13.8)$$

and so the critical singularity is

$$M \propto (T_c - T)^{1/(\gamma-3)} \quad (13.9)$$

when $3 < \gamma < 5$.[3] Recall a similar critical behaviour for percolation on complex networks. The singular dependence (13.9) is actually exact [75,119]. Remarkably, we obtained it using simple mean-field equations. In principle, the exactness of the mean-field approach is not a surprise since these networks are infinite-dimensional. The real surprise is a dramatic change of the critical behaviour in networks with heavy-tailed degree distributions. For sufficiently homogeneous substrates (exponent $\gamma > 5$), we observe a standard square root critical singularity. If a network substrate for the cooperative system is strongly heterogeneous ($\gamma < 5$), we find non-traditional singularity.

[3] In the critical region, $M \propto \mathcal{M}$. Note that the average magnetic moment of a node increases with the number of its connections, see eqn (13.4). For hubs, this moment approaches 1 even near the phase transition.

Physicists know that in the Ising model on two- and three-dimensional lattices, the critical exponent of the order parameter also differs from the mean-field value 1/2, but is smaller than this number. According to the standard classification, this range of exponent values corresponds to second-order phase transitions. In contrast, for the Ising model on scale-free networks, the exponent increases from 1/2 to ∞ as the degree distribution exponent decreases from 5 to 3. This indicates a progressive increase of the phase transition order from second to infinite.

Strong structural correlations in a network may change this picture. Moreover, on many networks, the Ising model has no ferromagnetic ordering at all. For example, the Ising model on any tree—a network without loops—is ordered only at zero temperature. Even weak thermal fluctuations break this ordering. Similarly, eliminating a tiny fraction of nodes in a tree, we split it into a set of finite components. On the other hand, we can easily restore macroscopic ordering in these systems by adding even a relatively small number of shortcuts between randomly chosen nodes.

13.3 Synchronization

What is synchronization?

The phase transitions, which we considered in preceding sections occur only in infinite (or, in reality, very large) interacting systems. At any

temperature, in a finite spin system, thermal fluctuations flip the magnetization from time to time, and on average, the magnetic moment is zero. There is, however, a wide circle of phenomena, where even a small set of interacting units, e.g. oscillators, can enter a coherent (synchronous) state. This synchronization is possible even for a pair of coupled oscillators. A traditional example is two pendulum clocks mounted on the same wall, which swing in unison. Synchronization occurs in a wide range of dynamical systems of very different sizes. Recall swarms of synchronously flashing fireflies, cicadas, and crickets chirping in unison and so on. Moreover, it is even possible that a finite dynamical system demonstrates synchronization, but its infinite counterpart does not.

The reader will find a detailed discussion of diverse synchronization phenomena in various networks in review [12]. Many real-life systems exhibiting synchronization have really complex architectures. A wide spectrum of synchronization phenomena, for example, was observed in complex networks in the brain. The brain contains a hierarchy of networks. At the cellular level, these are neuronal networks, where waves of activation spread over neurons (activators and inhibitors) interconnected by directed links—synaptic connections.[4] At a larger scale, researchers consider weighted cortical networks, in which nodes are interconnected cortical areas. Temporal patterns of brain electrical activity show a complicated set of oscillations at various frequencies, mostly in the range 1–100 Hz. Each pattern of rhythmic activity is closely related to a specific regime of brain functioning. Scientists are still far from understanding these brain rhythms and similar oscillatory behaviours of traffic and network flows in the Internet, cellular and ecological networks. Only for very simple dynamic models could researchers find a relation between the quality of synchronization and network architecture. We will discuss two of the basic systems exhibiting synchronization.

[4] Each directed weighted link in this network is a chain: axon–synapse(s)–dendrites, where synapses can change their states recording information. The human brain contains about 10^{11} neurons and 10^{14} synapses.

What is phase synchronization in essence?[5] The phase $\theta(t)$ of a single phase oscillator with an angular frequency ω evolves linearly, $\theta = \omega t + \text{const}$. This evolution is described by a trivial differential equation, $\dot{\theta}(t) = \omega$. For a pair of independent oscillators with two different frequencies ω_1 and ω_2, we have two independent equations: $\dot{\theta}_1 = \omega_1$ and $\dot{\theta}_2 = \omega_2$. Suppose that these oscillators are coupled in some way. Then, when is it possible that θ_1 and θ_2 evolve in phase, despite the difference between ω_1 and ω_2? Here 'in phase' means that the difference between the phases of the oscillators approaches a constant value. To couple oscillators, let us add new terms to the right-hand sides of the motion equations for the phases. These interaction terms must be zero when the phases coincide (mod2π, of course) and they must increase as the phases deviate (mod2π). The simplest relevant periodic function is a sinusoid. So we arrive at a pair of coupled equations,

[5] See the popular science book of Steven Strogatz [171] for a wonderful introduction to the topic.

$$\begin{aligned} \dot{\theta}_1 &= \omega_1 + J\sin(\theta_2 - \theta_1)\,, \\ \dot{\theta}_2 &= \omega_2 + J\sin(\theta_1 - \theta_2)\,, \end{aligned} \tag{13.10}$$

where J is a coupling constant. For the difference $\varphi = \theta_1 - \theta_2$, these

equations give

$$\dot{\varphi} = \omega_1 - \omega_2 - 2J \sin \varphi \,. \tag{13.11}$$

Synchronization means $\varphi(t \to \infty) \to$ const. For this, the equation $\omega_1 - \omega_2 = 2J \sin \varphi$ must have a solution. Consequently, the oscillators synchronize if the coupling is sufficiently strong, namely, $J > |\omega_1 - \omega_2|/2$. For smaller J, oscillators run incoherently.

Oscillators with random frequencies

Yoshiki Kuramoto at Kyoto University made a major breakthrough in the 1970s. He found an exact solution for the direct generalization of this model, eqn (13.10), to the case of N coupled oscillators with natural frequencies ω_i, $i = 1, 2, \ldots, N$ [116]. Kuramoto solved the model on the infinite fully connected graph, but, generally, the *Kuramoto model* is formulated on an arbitrary graph:

$$\dot{\theta}_i = \omega_i + J \sum_{j=1}^{N} a_{ij} \sin(\theta_j - \theta_i) \,. \tag{13.12}$$

[6] For an infinite fully connected graph, $a_{ij} = 1$ and J is substituted for J/N.

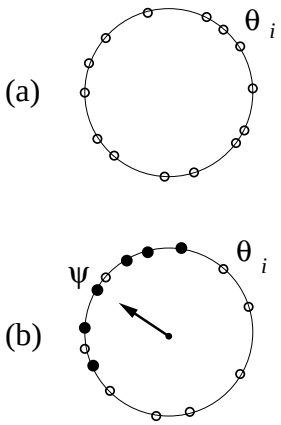

Fig. 13.1 Snapshots of the phases $\theta_i(t)$ of oscillators. (a) The incoherent state. The open dots show the drifting phases of oscillators on the unit circle in the complex plane. (b) The synchronized state. The filled dots show coherent (or locked) phases. The remaining oscillators drift incoherently. The direction of the vector and its length show the average phase $\psi(t)$ of the oscillators and the module of the complex order parameter, respectively. For more details, see the introductory review of Strogatz [172].

Here we assume that interactions between oscillators are through the links of a network with an adjacency matrix a_{ij}.[6] The frequencies ω_i in the Kuramoto model are random numbers from some distribution function, say, of a bell-shaped form, $\overline{\omega} = \sum_i \omega_i / N$. When $N > 2$, it is hard to follow the evolution of many phases $\theta_i(t)$, and synchronization becomes essentially more complicated and interesting than for two oscillators. The reader can suggest that when $N > 2$, only some fraction of the oscillators may be in phase, that is, the differences between their phases are constant, while the rest run incoherently. It turns out that this is indeed the case. To describe the collective dynamics of many individual phases and their mutual entrainment, Kuramoto used *the complex order parameter*:

$$r(t) e^{i\psi(t)} = \frac{1}{N} \sum_{k=1}^{N} e^{i\theta_k(t)} \,. \tag{13.13}$$

Here $\psi(t)$ plays the role of the average phase, and the module $r(t)$ characterizes the extent of coherence.

According to Kuramoto, the picture looks as follows, see Fig. 13.1. If interactions are sufficiently weak, the system stays in an incoherent state. In this state, the phases $\theta_i(t)$ drift more or less at random on the unit circle in the complex plane. They are homogeneously scattered over the circle, and so $r = 0$, see Fig. 13.1(a). There is a critical value of coupling constant above which this picture changes sharply. This critical coupling constant is of the order of the width of the distribution of oscillator frequencies, $\delta\omega$. Above the critical point, the phases are partially synchronized. Namely, a fraction of phases are coherent, which are locked and scattered around the average phase $\psi(t)$. These locked oscillators oscillate with the same frequency, see Fig. 13.1(a). (Their natural frequencies are in the region around $\overline{\omega}$.) The phases of the remaining oscillators drift incoherently. As interactions become stronger,

the number of locked oscillators grows, and, in addition, their phases draw nearer to the average phase. Consequently, the module of order parameter, r, increases.

In the Kuramoto model on the infinite fully connected graph, r is proportional to the square root of the deviation of the coupling constant from its critical value, $\sqrt{J - J_c}$. The singularity is the same as for a ferromagnetic second-order phase transition. On the other hand, in sharp contrast to ferromagnets, the Kuramoto model on infinite finite-dimensional lattices does not show synchronization at any coupling strength.[7]

[7] Here the distribution of natural frequencies is supposed to rapidly decrease. Long-range shortcuts restore synchronization in these infinite systems.

Instead of the fully connected graph, as a substrate for the Kuramoto model, we could use a classical graph. The results are similar. What about more complex and realistic networks? Suppose, a network is sparse and uncorrelated. For this network, one can use an effective medium approach similar to the one that we described for the Ising model [104]. This theory shows that for synchronization, the coupling constant must exceed the critical value

$$J_c \sim \delta\omega \langle q \rangle / \langle q^2 \rangle \,. \tag{13.14}$$

Here $\delta\omega$ is the width of the distribution of oscillator frequencies. Therefore, if the degree distribution of a network is heavy-tailed then there is synchronization even at a very small interaction strength. In that sense, hubs in a network significantly improve the synchronization of Kuramoto oscillators. On the other hand, when interaction J is weak, only a small fraction of the oscillators turn out to be synchronized. In this case, the cluster of coherent oscillators includes the hubs and, mostly, their neighbours. The cluster gradually increases with J.

This scenario is similar to what takes place in percolation problems on this network and what we observed for the Ising model, and so these theoretical predictions look natural. Unfortunately, it was hard to verify these predictions in time-consuming numerical simulations. The maximal sizes of studied networks did not exceed 10^5 nodes. These sizes, however, were sufficient to observe a surprisingly strong deviation from the results of the effective medium approach. Yamir Moreno and Amalio Pacheco simulated the Kuramoto model on the Barabási–Albert network [130] and found an unexpected synchronization transition. In the version of the Barabási–Albert model, that they used, each new node had several connections, so that the network was loopy. In this network, the second moment $\langle q^2 \rangle$ diverges (the degree distribution decays as q^{-3}). So, in accordance with eqn (13.14), one could expect partial synchronization at any coupling constant. Nonetheless, Moreno and Pacheco observed that the critical coupling constant is not zero, $J_c > 0$, and below this value any coherence is absent.

The first idea was that the observed transition could be a finite size effect. Indeed, the second moment of the degree distribution slowly varies with network size N, $\langle q^2 \rangle \sim \ln N$.[8] Consequently, if eqn (13.14) works for this network, then $J_c \propto 1/\ln N$. The logarithmic function is so slow that this estimate gives a noticeable J_c even if the network is very

[8] In this network, the cut-off of the degree distribution is $q_{\rm cut} \sim \sqrt{N}$, so $\langle q^2 \rangle \sim \int^{q_{\rm cut}} dq\, q^2 q^{-3} \sim \ln q_{\rm cut} \sim \ln N$.

large. For example, $1/\ln 10^4 \approx 0.11$, $1/\ln 10^6 \approx 0.07$, $1/\ln 10^8 \approx 0.05$. In situations of this kind, researchers study a set of networks of various sizes. Moreno and Pacheco investigated the networks of 10^3, 10^4, and 5×10^4 nodes. For each of these networks they measured the variation of the order parameter r with interaction strength J. All of these curves were rather smooth, but it was possible to evaluate the dependence $J_c(N)$. It turned out that J_c does not decrease with N in contrast to our estimate. On the contrary, $J_c(N)$ increases, apparently approaching some constant value. Then, why does formula (13.14) fail in this case? At present nobody can give a definitive answer. The failure is probably because the growing Barabási–Albert network significantly differs from the uncorrelated networks.

Identical oscillators

There is a wide class of synchronization systems which are fundamentally different from the Kuramoto model. In these systems, interacting oscillators are identical. This is why the synchronization is not partial, as in the Kuramoto model, but full. So the structure of the coherent state is quite definite and clear. The natural question then is: what are the conditions of this synchronization? Namely, what is the range of coupling constant values and network architectures that provide this synchronization?

Physicists formulate models for this kind of synchronization in a very general form. The state of an individual oscillator is described by some vector $\mathbf{x}$. The equation of motion is the same for each oscillator, $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$, when the oscillators are uncoupled. Here $\mathbf{F}(\mathbf{x})$ is a given vector function. It is easy to make these oscillators interacting. For this, we couple the equations of motion for the oscillators, $i = 1, 2, \ldots, N$, together in the following way:

$$\dot{\mathbf{x}}_i = \mathbf{F}(\mathbf{x}_i) + J \sum_j a_{ij} [\mathbf{H}(\mathbf{x}_j) - \mathbf{H}(\mathbf{x}_i)] \,. \tag{13.15}$$

Here $\mathbf{H}(\mathbf{x})$ is a so called 'output function' and a_{ij} are the elements of the adjacency matrix of the underlining network. The functions $\mathbf{F}(\mathbf{x})$ and $\mathbf{H}(\mathbf{x})$ are given characteristics of the model.[9] The sum on the right-hand side of eqn (13.15) is usually represented in another form, namely

$$\sum_j (q_i \delta_{ij} - a_{ij}) \mathbf{H}(\mathbf{x}_j) \,,$$

where q_i is the degree of node i. In mathematics, the matrix $q_i \delta_{ij} - a_{ij}$ is called Laplacian. Therefore, to evaluate the influence of the network architecture on synchronization, we should study the Laplacian matrix of a given network. These matrices play a paramount role in graph theory and have numerous applications. The spectra of Laplacian matrices (Laplacian spectra) determine the dynamics of various processes on networks, for example, a random walk of a particle and, more generally, a wide range of diffusion processes. Furthermore, we do not need to

[9] Note that eqn (13.15) and corresponding eqn (13.10) in the Kuramoto model have similar structures, although the methods of coupling the oscillators are different. Compare the terms: $\sin(\theta_j - \theta_i)$ and $\mathbf{H}(\mathbf{x}_j) - \mathbf{H}(\mathbf{x}_i)$.

know a full Laplacian spectrum, see Fig. 13.2. Two physicists, Mauricio Barahona and Louis Pecora showed that the presence of this specific synchronization is determined by only two numbers in the Laplacian spectrum of a network—the minimum non-zero eigenvalue and the maximum one [17].

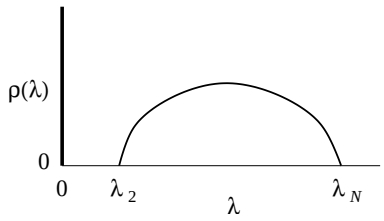

Fig. 13.2 Typical Laplacian spectrum of a large network. Let a graph of N nodes consist of a single connected component. Its Laplacian spectrum has N eigenvalues: λ_1=0 (see the peak at λ=0 in the plot) and $0<\lambda_2<\ldots<\lambda_N$. In the fully connected graph, $\lambda_2 = \ldots = \lambda_N$. In an infinite lattice, $\lambda_2 \to 0$.

According to Barahona and Pecora, a network is synchronizable (in other words, its synchronization state is stable) if the coupling strength J is in the interval $(\alpha_1/\lambda_2, \alpha_2/\lambda_N)$. Here the numbers α_1 and α_2 depend only on the specific functions $\mathbf{F}(\mathbf{x})$ and $\mathbf{H}(\mathbf{x})$. These numbers are positive, $0 < \alpha_1 < \alpha_2$. Typically, α_2/α_1 is in the range from 5 to 100. Consequently a network is synchronizable if

$$\frac{\lambda_N}{\lambda_2} < \frac{\alpha_2}{\alpha_1}\,. \tag{13.16}$$

The left-hand side is determined by the architecture of the network, while the right-hand side is determined by the nature of the oscillators and their interactions. When we consider the role of the network structure, we treat the ratio α_2/α_1 as a given parameter in the problem. Thus the smaller the ratio λ_N/λ_2, the better is synchronizability.

Criterion (13.16) is applicable to any network and lattice. In the infinite lattices, there is no gap between the first eigenvalue and second, that is $\lambda_2 = 0$. So we readily see that synchronization is absent at any coupling strength if a lattice is infinite. In the finite lattices, $\lambda_2 > 0$, and synchronization is possible. It was found that in the small-world networks, shortcuts diminish the ratio λ_N/λ_2, see Fig. 13.3, and so with shortcuts, infinite lattices become synchronizable. The same is true for phase synchronization in the Kuramoto model. Thus in both classes of interacting systems which we have discussed, the transition from a lattice to a small world results in synchronization. In the beginning of this section we listed a number of real-world systems exhibiting synchronization. Some of them are large. It is the small-world topology of these systems that makes synchronization possible.

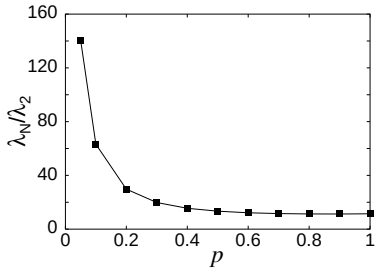

Fig. 13.3 Dependence of the ratio λ_N/λ_2 on the fraction p of shortcuts for the infinite Watts-Strogatz network. Adapted from Hong, Kim, Choi, and Park [103].

Criterion (13.16) shows that the fully connected graph ($\lambda_2 = \lambda_N$) provides the best synchronizability. For a sparse network with a given mean degree of a node, the narrowest Laplacian spectrum is for a random regular graph and then for a classical random graph. The minimum ratio λ_N/λ_2 provides the best synchronizability. For networks with heavy-tailed degree distributions, we have wide Laplacian spectra. It turns out that λ_N is determined by the cut-off of a degree distribution, $q_{\rm cut}(N)$, while λ_2 is independent of the cut-off. In particular, for uncorrelated networks, $\lambda_N = q_{\rm cut} + 1$ [108]. This crucially impairs the synchronizability of scale-free networks.

It is possible to arrive at the same conclusions in another way. Donetti, Hurtado, and Muñoz addressed the problem: what is the best synchronizable network for given numbers of nodes and links? To find the answer, they used a numerical optimization procedure [72]. They started from an arbitrary connected graph, having given numbers of nodes and links, and successively rewired uniformly randomly chosen links to ran-

domly chosen nodes. For each rewiring, they checked whether it improved synchronizability. If it improves, then rewire; if it impairs, then abort this rewiring and try the next one. This time-consuming optimization allowed them to study only very small networks of a few hundred nodes. All the resulting optimal networks turned out to be cage graphs, see Fig. 1.2, that is networks with all nodes of equal degree, having loops (cycles) of the maximal possible length. Among non-random graphs of a given size, cage graphs are the closest to random regular graphs, so, indeed, homogeneous architectures provide the best synchronizability.

13.4 Games on networks

Let us pass now from interacting spins and oscillators on networks to specifically interacting individuals, namely, to players. The game which we choose for our players is the famous *prisoner's dilemma*, which is one of the standard problems in game theory. Each of two players ('prisoners') independently decides for himself which is better: to cooperate (remain silent), C, or to defect (betray), D? This decision should be based on the following set of payoffs for the players:

Players:	Their payoffs:
(C, C)	$(1, 1)$
(C, D)	$(0, b)$
(D, C)	$(b, 0)$
(D, D)	$(0, 0)$

Here we present the simplest non-trivial set. Importantly, b is assumed to be greater than 1. If both players decide to cooperate, then they will receive equal payoffs, 1. If the first player cooperates, and the second defects, then the first will get nothing, 0, while the second will receive b, and so on. Then, what is better: to cooperate or to defect? What is more profitable in terms of payoff? In this simple game, a strategy providing the highest payoff is obvious. Since $b > 1$, the best action for a player is to defect. Indeed, suppose first that his opponent defects. Then, irrespective of the strategy of the first player, he will receive zero payoff. On the other hand, if the second player cooperates, then the first will get a bigger payoff by defecting, b against 1. Then it is better to always defect. The paradox is that if both the players use the defector strategy, then, together, they score less than cooperating: $0 + 0 < 1 + 1$.

The game becomes much more interesting when players can adopt the strategies of their opponents, and when there are many players [173]. For a network of players, at a given moment, each node (player) i can be in one of two states, $\sigma_i = C$ or $\sigma_i = D$, depending on the strategy that the player uses at this moment. Each pair of nearest neighbours independently play the game and receive their payoffs determined by the states of these pairs, see the list above. To make this system evolve we must allow individual players to change their strategies from time to time, $\sigma_i{=}C \longleftrightarrow \sigma_i{=}D$. Suppose that the players have information

only about the results of their opponents—neighbours. Then a natural idea for an adaptive player is to adopt the most successful strategy in his close environment (the player himself and his nearest neighbours).

This idea was realized in the so-called *spatial prisoner's dilemma* of Martin Nowak and Robert May [142]. Without going into detail, in this mathematical model, originally formulated for a regular graph, the evolution of the players' strategies appears as follows. Initially, there is a random configuration of defectors and cooperators. For example, the initial concentration of cooperators c may be set to $1/2$. Then all of the pairs of nearest neighbours play the game independently, and after this round, each player i accumulates his payoffs as P_i. After that, for each player i, choose at random one of its immediate neighbours, j, and compare the scores P_i and P_j.[10] If $P_i > P_j$, then leave the state (strategy) of i unchanged. Otherwise, let player i accept the strategy of j with some probability. After this update, pass to the next round, recalculate all the scores and so on. As a result of this evolution, the concentration c of cooperators approaches a stationary value. This value depends on the parameter b of the prisoner's dilemma.

This evolutionary model was generalized to other, non-regular networks and simulated [158]. Figure 13.4 shows the resulting final concentrations of cooperators in contrasting networks (regular graphs and Barabási–Albert networks). The reader can see that scale-free architectures stimulate and support cooperation. It turns out that players occupying the hubs tend to be cooperators, which makes cooperation dominant in these networks. These observations may, at least partially, explain why cooperation is widespread in heterogeneous populations. One can go even further and couple two evolutionary processes in these systems, namely, the evolutions of players and of a network [146]. In this generalization, not only does the network influence the course of the game, but also the dynamics of the game changes the structure of the network.

[10] The original version of the model is deterministic: each player adopts the best strategy in his close environment [142]. Here we describe a stochastic version.

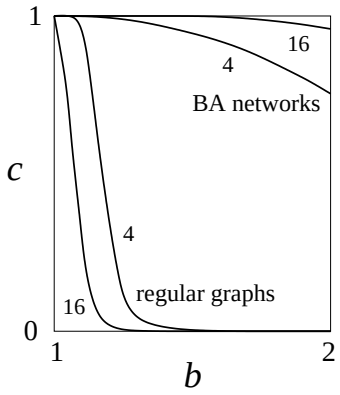

Fig. 13.4 Schematic plot of a final concentration c of cooperators versus parameter b for two contrasting networks according to Santos and Pacheco [158]. The numbers labelling the curves indicate the mean degrees of the networks.

13.5 Avalanches as branching processes

Immediately before complex networks, so-called *self-organized criticality* was the hottest topic in non-equilibrium statistical mechanics.[11] Most of the numerous researchers who studied self-organized criticality in the 1980s—1990s, later switched over to complex networks. This partially explains why many concepts and models from self-organized criticality have been extensively applied to networks. A simple example illustrates this important notion [159]. Let the nodes of an infinite lattice be occupied or empty, so that we have a set of clusters on a lattice like those in percolation problems. Consider the following process. In parallel,

(i) infinitely slowly, fill new and new uniformly randomly selected nodes, and

(ii) also infinitely slowly, choose at random nodes and empty them together with the clusters to which these nodes belong, see Fig. 13.5.

[11] See the fascinating popular science book of Per Bak, a pioneer in this research field [13].

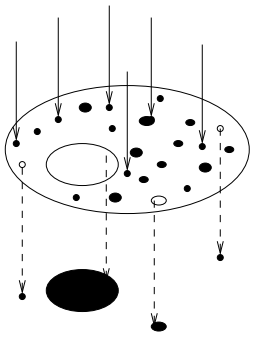

Fig. 13.5 A forest-fire model demonstrating self-organized criticality. The solid arrows show the slow filling of randomly selected nodes (growing trees). The dashed arrows the parallel infinitely slow process: random selection of nodes and elimination clusters (forests) to which these nodes belong (forest fires).

If we start from an empty lattice, the concentration of occupied nodes will slowly grow until it reaches some value. The maximal size of clusters will grow but will never approach infinity. The point is that a giant connected component cannot emerge in this system since process (ii) efficiently eliminates any cluster containing a finite fraction of nodes. Therefore the combination of slow processes (i) and (ii) drives a system exactly to the point of the emergence of a giant component, that is to a critical state. The supposed infinite slowness of the processes is crucially important. If we added nodes and removed clusters at a finite rate, the system would stay away from criticality. Remarkably, the system enters this state by itself, without fitting any control parameter, which explains the term 'self-organized criticality'. We can interpret this demonstrative example as a forest-fire model. In this interpretation, in parallel,

(i) new trees grow up in random places and form forests (clusters—connected components), and

(ii) random fires burn forests.

We can also treat each removed cluster as an avalanche initiated by the removal of a single node. It is easy to estimate the size distribution of removed clusters (avalanches or burned forests). Let us suppose for the sake of simplicity that our lattice is high-dimensional. As a first approximation, it is natural to assume that the statistics of clusters in this state is similar to that at a percolation threshold in high-dimensional lattices. Then we can use the fact that for percolation problems in this situation, the size distribution of clusters to which uniformly randomly chosen nodes belong is decreased as a power law, $s^{-3/2}$. This is the size distribution of avalanches in our model. In general, the exponent of the size distribution of avalanches in various high-dimensional self-organized criticality systems and models is exactly 3/2.

Generally, the notion 'self-organized criticality' implies self-organization of an open non-equilibrium cooperative system into a critical state.[12] This state features power-law decaying correlations, power-law distributed avalanches, and other critical phenomena. Note the contrast: (i) in standard phase transitions, a system is driven to a critical point by setting temperature (concentration of defects, pressure, etc.) to a special critical value; (ii) in self-organized criticality, the system itself, spontaneously enters a critical state. Typically, self-organized criticality sports abrupt phenomena, and so scientists mostly study the statistics of avalanches. Among real-world self-organized criticality systems and phenomena are sandpiles, earthquakes, and many others.

After this brief insight into self-organized criticality we return to networks and place a self-organized criticality system on the top of a complex network. For the sake of brevity, we only touch upon a sandpile problem on a network [98]. Suppose that each node in a network can store grains but not more than some threshold number.[13] Let us begin slowly to add grains to randomly selected nodes.[14] The rule is that if the number of grains on a node exceeds the threshold, then some of these grains jump to randomly chosen nearest neighbours of the node.

[12] Note that in our example, the system was open. Specifically, we added nodes and removed connected components.

[13] In principle, this threshold number may depend on the number of connections of a node.

[14] As is usual in self-organized criticality, we must also incorporate a grain outflow into this scheme. For example, perpetually remove all grains from some of the nodes.

The number of toppled grains depends on a specific sandpile model. The thresholds of some of these neighbours may be exceeded, and their grains will also topple and so on. Thus the first toppling has a chance to trigger an avalanche. What is the size distribution of these avalanches? Since our networks are infinitely dimensional, we guess that this distribution is of the same form as in our demonstrative forest-fire example, namely $s^{-3/2}$. It turns out that this is indeed the case if the degree distribution of a network is a rapidly decreasing function. In the case of heavy-tailed degree distributions, the problem was resolved for infinite uncorrelated networks [98]. In these networks, thanks to their local tree-like organization, the structure of avalanches is particularly simple. Clearly, the avalanches on a network of this kind must be subgraphs of the mother network, that is trees, see Fig. 13.6. Similarly to what we observed in percolation problems, these trees–subgraphs have different degree distributions, correlations, and even dimensionality than their mother networks. Nonetheless, the similarity is not complete. In networks with heavy-tailed degree distributions, the structure and statistics of these sandpile avalanches was found to differ markedly from those of connected components in percolation problems. Furthermore, the statistics of avalanches depends significantly on a specific sandpile model. The qualitative result, valid for a wide range of self-organized criticality models, was that, counterintuitively, the presence of hubs in a network makes the size distributions of avalanches more rapidly decaying. One may say that hubs hamper large avalanches.[15] For scale-free networks with a degree distribution decaying as $q^{-\gamma}$, the size distribution of avalanches becomes exponential as exponent γ approaches 2. In this limiting situation, hubs eliminate self-organized criticality.

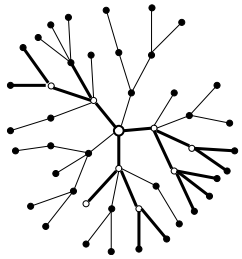

Fig. 13.6 A sandpile avalanche on a tree-like network. The large dot indicates a node which started the avalanche. Grains on white nodes were toppled, and the toppled grains jumped along the bold links to the neighbouring nodes inducing further topplings.

[15] Recall that the degree distribution of the nearest neighbour of a node is $qP(q)/\langle q\rangle$, so hubs (nodes with many connections) participate in avalanches with high probability and strongly influence their statistics.

Optimization

14

The architectures of real-world networks have one major thing in common. All of these complex architectures are optimal or nearly optimal for the function of these networks. Furthermore, the optimality of network design can be regarded as a driving force of network evolution, its engine. In this lecture we discuss how a universal requirement for optimality leads to the complex structural organization of a network.

14.1 Critique of preferential attachment

We have described the preferential attachment mechanism as a strict and convenient way for generation of heavy-tailed distributions. This explanation of scale-free distributions is particularly popular among physicists. Theoretical physicists implement preferential attachment in their constructions because it allows easy analytical treatment. In addition, preferential attachment enables researchers to rapidly generate extremely large networks for numerical simulations. Nonetheless, despite all these advantages, preferential attachment has a weak point. No satisfactory explanation of this mechanism has been proposed in numerous situations. In particular, there are no convincing quantitative arguments for specific linear forms of a preference function, necessary for scale-free distributions. Thus, preferential attachment explains power laws, but there is no good explanation for preferential attachment. A long-running criticism of the preferential attachment mechanism actually began before the complex networks boom, in the 1950s. During the 1950s—1960s, Benoît Mandelbrot, an outstanding applied mathematician who invented the fractality concept, and Herbert A. Simon, one of the proponents of what is now called the preferential attachment concept, exchanged a series of rather scaring comments on the nature of power laws. Mandelbrot repeatedly criticized the preferential attachment concept as too mechanistic and lacking in explanatory power. As an alternative to preferential attachment, he proposed consideration of optimization mechanisms [121].

In 2002, a group of network engineers unleashed another wave of criticism of preferential attachment concepts [182].[1] Their criticism was primarily aimed at a wide range of models of the organization and function of the Internet, proposed by physicists. The computer scientists stressed that these models 'are only evocative; they are not explanatory'. In their definition, an evocative model 'can reproduce the phenomenon of interest but does not necessarily capture and incorporate the true under-

[1] See also our discussion of self-similar traffic in Lecture 12.

lying cause'. On the other hand, an explanatory model 'also captures the causal mechanisms (why and how, in addition to what)'. These researchers believed that only optimization driven evolution could explain the architecture of the Internet in a natural and convincing way.

Unfortunately, this criticism from Mandelbrot and the computer scientists itself has a weak point. The problem is that optimization based models are typically far more difficult to deal with than models implementing preferential attachment. Optimization driven evolution is far less studied than preferential attachment. Mandelbrot showed very schematically how a requirement for optimality can result in power-law distributions. His work was not directly applied to networks. As for the previously mentioned group of network engineers, they did not propose 'explanatory' optimization models of networks. In principle, one can formulate plenty of models of this kind. Unfortunately, they are hardly treatable analytically or numerically. Numerical simulations of optimization driven network evolution are so time consuming that only very small networks can be generated. Their sizes (typically, a few hundred nodes) are not sufficient to arrive at solid conclusions about the structures of these networks and of their bigger counterparts. In particular, it is impossible to distinguish scale-free and non-scale-free small networks. This is why only the simplest optimization based network models were explored. In the next sections, we touch upon a few of these networks.

14.2 Optimized trade-offs

Optimization based models of networks are usually organized in the following way. New connections or rewirings must optimize some combination of network characteristics. That is, a new connection is made to the 'best place' in a network that provides the highest or smallest value of some function of characteristics of the network after this reconstruction. This function to optimize is called a *cost function*. For example, the cost function may be a linear function of the mean internode distance and other global characteristics of a network. Minimization of this function leads to more compact architectures. In another optimization scheme, the variables of a cost function are the characteristics of a node to which we link. For example, a variable can be the shortest path distance between these node, and some other one.

In 2002, three computer scientists, Fabrikant, Koutsoupias, and Papadimitriou, proposed a remarkably simple optimization driven model of a growing tree [89]. The absence of loops allowed a strict analytical treatment of the model. The growth starts from a single node (root with label 1) placed at some point of a restricted two-dimensional area. Nodes are added one by one. Each new node is placed within this area and attached to a selected existing node, which results in a tree. We can characterise completely the position of each node i in this tree by the set of distances: the shortest-path network distance $\ell_{1,i}$ between node

i and the root ('the depth' of node i) and the Euclidean distances $d_{i,j}$ between node i and other nodes, $j \neq i$, in the network, see Fig. 14.1. At each time step,

(i) place a new node t at a random point of the area and

(ii) attach it to that node i which gives the minimum cost function $\ell_{1,i} + \alpha(t)d_{i,t}$.

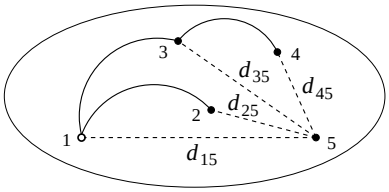

Fig. 14.1 The Fabrikant–Koutsoupias–Papadimitriou network after the first four nodes are placed within a restricted two-dimensional area. The arks show the links. d_{15}, d_{25}, d_{35}, and d_{45} are the geographic distances between the new (fifth) node and others. Find the minimum number out of αd_{15}, $1+\alpha d_{25}$, $1+\alpha d_{35}$, and $2+\alpha d_{45}$, and attach the fifth node to the corresponding node.

Here the coefficient $\alpha(t)$ is a given function of the network size t. This model is based on 'optimized trade-offs' between two conflicting objectives: a network objective versus a geographic one. The network objective is 'connect to the root (the centre of a network) and make the entire network more compact'. The geographic objective is 'connect to the geographically closest node', which usually means the cheapest connection. If $\alpha = 0$, then all attachments are to the root, and the network is a star, which is the most compact tree. If α is large, the attachments are mostly to the geographically closest nodes. The structure of the resulting network depends significantly on a given $\alpha(t)$. The authors of this model found that if $\alpha(t)$ is a power law, then the resulting degree distribution has a power-law part. This finding immediately made the model very popular, mostly among computer scientists. The hope was that this idea might explain scale-free of the Internet and other real networks. Unfortunately, this hope was short-lived. Already in 2003, it had been shown that the power-law region of the resulting degree distribution is narrow if the network is large [24].[2] In that sense, the Fabrikant–Koutsoupias–Papadimitriou model does not produce a really scale-free tree in the large-size limit. Only sufficiently small networks grown in this way can be perceived as scale-free.

[2] More precisely, depending on a form of the function $\alpha(t)$, either the power-law region is quite narrow or the great majority of vertices are leaves.

14.3 The power of choice

In optimization algorithms which we discussed above, we selected the optimal node for attachment among all nodes in a network. This leads to a heavy computational task. In practice, this optimization is impossible even for moderately large networks. We can, however, do partial optimization. Namely, choose at random a number of nodes and find the optimal one among them. If this number is much smaller than the network size, then we facilitate our work dramatically. On the other hand, this optimization is only partial, and its results may be far from those of complete optimization.

Raissa D'Souza, Pavel Krapivsky, and Cristopher Moore considered a simple growing network which illustrates the essence of these algorithms [83]. This network is a growing tree, where each new node finds a node to be attached to in the following way. Select uniformly at random $k > 1$ nodes and choose the node with the highest degree among them.[3] Importantly, the number k of choices in this model must stay constant as the network grows. If the number of choices is infinite, then all attachments are to the root, which gives a star graph. In the case of

[3] The reader can see that if $k = 1$, then this network is a random recursive tree.

a finite k, the resulting degree distribution has an exponential cut-off at degrees of the order of k, which severely restricts a region, where a slower decay can be observed. For example, if $k = 2$, then this network has practically an exponential degree distribution. When k is sufficiently large, there is a range of degrees where these authors found a power-law decaying distribution. However, for reasonable numbers, say $k = 20$ or 50, this region is so narrow that it is hard to reliably observe the power law and to measure its exponent. Theoretically, the degree distribution is $P(q) \sim 1/q$ in the range $1 \ll q \ll k$. The reader can see that, indeed, the resulting architecture significantly differs from a star graph.

This choice-driven process somewhat resembles preferential attachment since highly connected nodes are preferentially selected in both of these algorithms. Nonetheless, these two mechanisms produce contrasting degree distributions with different power laws and different cut-offs. In particular, the preferential attachment mechanism cannot provide the $P(q) \sim 1/q$ degree distribution.

In 2009, Dimitris Achlioptas, Raissa D'Souza, and Joel Spencer exploited the power of choice in another problem [2]. They studied the emergence of a giant connected component in networks. To explain their problem, let us return to classical random graphs, particularly to the $G_{N,p}$ (Gilbert) model. In this notation, N is the number of nodes in the random graph, and p is the probability that two nodes in the graph are interlinked. We showed that this equilibrium random graph has a giant connected component when the mean degree of a node, $\langle q \rangle = pN$, exceeds 1, see Fig. 2.5.[4] In other words, to obtain a giant connected component, each pair of nodes must be interlinked with a probability p exceeding $1/N$. There is another view of this network. Let us start with a large set of isolated nodes. Choose uniformly at random a pair of nodes and connect them by a link. Then choose and connect another pair of nodes and so on. That is, the number of links grows, while the number of nodes is fixed. A giant connected component will emerge when the number of links reaches $N/2$, which exactly corresponds to $p = 1/N$. Achlioptas and coauthors used a similar network-formation process in their construction.

Their network is also built starting with N isolating nodes, by adding links one by one. Each new link is created in the following way, see Fig. 14.2.

(i) Pick uniformly at random two pairs of nodes, say, i, j and k, l. Each node in a network belongs to some connected components. For each of the two pairs compute the products of the sizes of connected components to which these nodes belong, $s_i s_j$ and $s_k s_l$.

(ii) Create a link between that pair of nodes which has the smallest product.[5]

These rules lead to quite different statistics of connected components than in classical random graphs. We can treat the forming of this network as a specific aggregation process in which selected pairs of components merge into single ones. Importantly, the rules hamper the creation

[4] Recall that the notion of a giant connected component is meaningful for infinite networks, so N tends to infinity.

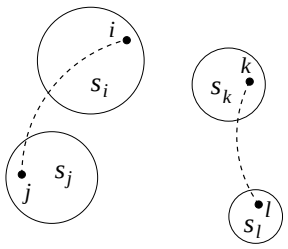

Fig. 14.2 The choice of nodes to connect by a link in the scheme of Achlioptas, D'Souza, and Spencer. Of these two pairs of nodes, the right one must be chosen since the product of the sizes of connected components to which its nodes belong, $s_k s_l$, is smaller than $s_i s_j$.

[5] In another version of this process, the sums $s_i + s_j$ and $s_k + s_l$ are compared instead of the products, which produce similar phenomena.

of large connected components and ultimately delay the emergence of a giant connected component. As a result, a giant connected component emerges at a higher concentration of connections than in classical random graphs. Moreover, according to Achlioptas and coauthors, this component emerges abruptly, as a sudden jump, so that this phenomenon was named *explosive percolation*, see Fig. 14.3. At the end of 2009, it is still unclear if this jump is supplemented by a singularity as for a hybrid transition (recall the emergence of a k-core). We can increase the number of choices in this scheme. For example, instead of comparing two pairs of nodes, we can compare three or more pairs. Then a giant component will emerge at a higher concentration of connections and a critical jump will be bigger.

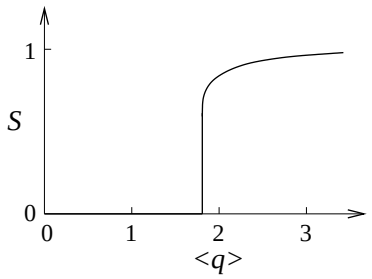

Fig. 14.3 Schematic plot of the dependence of the relative size of a giant connected component on the mean degree $\langle q \rangle$ in this construction according to Achlioptas and coauthors [2].

Outlook

15

Three major milestones mark the history of the exploration of networks: Leonhard Euler's work (1735), the introduction of random graphs (1950s), and the launch of the large-scale study of complex networks (the end of the 1990s, although isolated works on the topic appeared earlier). Each of the new milestones was related to a wider range of real-world network systems and to a wider range of sciences involved. The boom in networks, starting in the 1990s, has already involved and affected practically all natural and human sciences. Surely, the emergence of the Internet and WWW have played a pivotal role in the latest advance. So, the science of networks in its currents state is an ultimately multidisciplinary research field. We can hardly expect such dramatic breakthroughs in the near future, but the incredible rate of progress in the understanding of networks apparently does not slow down. Below we list a few particularly hot and prospective topics in complex networks.

(1) Networks with rich community structures, community detection.

(2) Complex hypergraphs and multi-partite networks.

(3) Loops and motifs in networks.

(4) Flows in complex networks.

(5) Optimization driven design of networks.

(6) Coevolving networks and interacting systems.

(7) Dynamical systems placed on networks.

(8) Searching, retrieving, and indexing information from complex networked environments.

In the long run, the task is to translate the general understanding of networks into working methods and strategies. The ultimate aim is to control, manipulate, and use the structure and function of real-world networks. Despite recent spectacular advances, we are still a long way from this challenging goal.