11

Traffic on complex networks

The main role of many networks is to provide the physical substrate to the flow of some physical quantity, data, or information. In particular, this is the case for the transportation and technological infrastructures upon which the everyday functioning of our society relies. In this context, congestion phenomena, failures, and breakdown avalanches can have dramatic effects, as witnessed in major blackouts and transportation breakdowns. Also, in technological networks, if the single element does not have the capacity to cope with the amount of data to be handled, the network as a whole might be unable to perform efficiently. The understanding of congestion and failure avalanches is therefore a crucial research question when developing strategies aimed at network optimization and protection.

Clearly the study of network traffic and the emergence of congestion and large scale failures cannot neglect the specific nature of the system. The huge literature addressing these questions is therefore domain oriented and goes beyond the scope of this book. On the other hand, large-scale congestion and failure avalanches can also be seen as the emergence of collective phenomena and, as such, studied through simple models with the aim of abstracting the general features due to the basic topology of the underlying network. This chapter is devoted to an overview of results concerning traffic, congestions, and avalanches and is more focused on the general and ubiquitous features of these phenomena than on system-specific properties. However, some of the general results highlight properties and phenomenologies that can be applied in a wide range of phenomena and therefore provide insight on real-world systems.

11.1 Traffic and congestion

It is impossible to provide a general and complete account of the study of traffic in networks, since in each domain we find a specific research area devoted to the subject, each worth a chapter of this book.

https://doi.org/10.1017/CBO9780511791383.012 Published online by Cambridge University Press

Traffic and congestion in transportation networks have a clear impact on the economy and the environment. Understanding individuals' movements is also crucial in epidemiology. This has led to a substantial body of work that concerns the problem of how to determine the traffic on every link of the network once the outgoing and incoming flows are known for all locations (the so-called origin–destination matrix). This problem, known as the "traffic assignment problem," was addressed a long time ago by Wardrop (1952) in the equilibrium perspective. It is usually assumed that every user will try to minimize her or his travel cost computed as a function of distance, transportation cost, travel time, etc. The main assumption formulated by Wardrop is that the equilibrium is reached when no user can improve travel time by changing routes unilaterally. This user equilibrium condition is akin to the Nash equilibrium in economics. The interaction between users occurs because of congestion effects which increase the travel time when the traffic is too large. Despite the large number of studies on this equilibrium principle, it is still unclear how the different heterogeneities observed in transportation networks will affect the traffic patterns.

Indeed, several studies have reported evidence in transportation networks of complex topological properties encoded in large-scale heterogeneities and heavy-tailed statistical distributions on different systems such as the airline system (Guimerà and Amaral, 2004; Barrat *et al.* 2004a), subways or railways (Latora and Marchiori, 2002; Sen *et al.*, 2003) and roads and city streets (Cardillo *et al.*, 2006; Buhl *et al.*, 2006; Crucitti *et al.*, 2006; Scellato *et al.*, 2006; Kalapala *et al.*, 2006). Although complex networks made their appearance in transportation research through the empirical measures, the impact of complex topologies on traffic is still to be understood. In addition, even if the separation between these different processes – the evolution of transportation networks and the dynamical traffic patterns – allows for a simple presentation of this field, they are ultimately coupled (see for example the paper by Yerra and Levinson [2005]). The growth of urban centers is facilitated by their accessibility, which in turn will create high demand in terms of transportation means. This simple fact means that modeling approaches should consider the coupling between various socio-economical factors, an effort which has already been undertaken by geographers, economists, and transportation researchers, but in which complex networks still play a minor role.

In the context of information technology, the behavior of internet traffic has been analyzed since the early times of computer networks.[1] An impressive bibliographical guide to the literature in the early years can be found in Willinger, Taqqu and Erramilli (1996), but work in this field is still rapidly progressing (Crovella,

[1] This kind of analysis has been supported by the timely availability of packet traffic tools such as `tcpdump`, developed by V. Jacobson, C. Leres, and S. McCanne in 1989.

Lindemann and Reiser, 2000; Crovella and Krishnamurthy, 2006). The largest part of these studies generally focuses on a *local* view of the network, by analyzing a single link or a small collection of them, providing just the properties of the local traffic. On the other hand, the vision obtained from a single link is often the outcome of a global dynamic in which very distant regions of the network also cooperate. Indeed, the characterization of traffic in technological networks has contributed enormously to the development of the theory of self-similar processes that traces back to the early work of Mandelbrot (1969), and is obviously related to fractal and scale-free phenomena (Mandelbrot, 1982). In general, traffic time series on internet connections are scale-free, a property which has correspondences in many different research fields, ranging from surface growth (Barabási and Stanley, 1995) to economics (Mantegna and Stanley, 1999) or geophysics (Rodriguez-Iturbe and Rinaldo, 1997). Moreover, the scale-free behavior is not only observed in traffic series. The traffic self-similarity also affects network performance and queueing time (Park, Kim and Crovella, 1997), inter-arrival time, and end-to-end delay (Huffaker *et al.*, 2000; Percacci and Vespignani, 2003) which displays statistical distributions with heavy tails and power spectrum of $1/f$ type (Csabai, 1994; Paxson and Floyd, 1995; Fowler, 1999; claffy, 1999). Finally, self-similarity is a general property of the large majority of traffic generated on the Internet, including WWW traffic from file requests on web servers (Crovella and Bestavros, 1997).

A global characterization of internet traffic and performance adds a new dimension to the problem that requires a large amount of resources and the solutions to several technical problems. Traffic and performance analysis on a global scale implies the collection of traffic traces and performance estimators which, even on a single link, represent a noticeable volume of data. It is easy to realize that gathering data on hundreds or thousands of links and routers poses great problems in data handling and storage, as well as in the traffic generated by the measurement process itself. Additionaly, on the global level, these data must be correlated with the detailed physical and geographical structure (bandwidth, link length, etc.) of the Internet. Despite these technical difficulties, an increasing body of work focuses on the Internet as a whole specially aimed to forecast future performance trends. For instance, interdomain traffic can be studied at a global level by looking at data representing the whole traffic received by specific service providers (Uhlig and Bonaventure, 2001). Modern measurement infrastructure also allows the construction of traffic matrices representing the traffic flow between pairs of service providers (claffy, 1999; Huffaker *et al.*, 2000). In this case, traffic flows are aggregated on the basis of the sources and destination addresses. These traffic matrices can also be correlated with the geographical location of traffic sources and destinations in order to obtain information on regional or national policies

and connectivity. Finally, measurements also focus on internet performance at the global level (Huffaker *et al.*, 2000; Lee and Stepanek, 2001; Percacci and Vespignani, 2003) and on its resilience to failures (Paxson, 1997; Labovitz and Malan, 1998; Labovitz, Ahuja and Jahanian, 1999; Magnasco, 2000).

While these previous investigations constitute a large body of work, the investigation of network traffic with the aim of abstracting general properties independent of the detailed system properties is less frequent. This approach, which focuses on *universal* properties, naturally finds its roots in statistical physics and has been pursued in several cases by explicitly considering the complex topological properties of many networks. A series of works along these lines focused on the study in five different large-scale natural and technological networks of the relationship between average flux or traffic $\langle f_i \rangle$ on a node and the corresponding fluctuations σ_i (de Menezes and Barabási, 2004a; Barabási *et al.*, 2004). Data were obtained for the daily traffic on internet routers, for the activity of logic gates in a microprocessor, for the number of daily visits of websites, for the traffic on highways, and also for the stream flow of a river network. Strikingly, power-law relations $\sigma \sim \langle f \rangle^{\alpha}$ were observed, with $\alpha = 1/2$ for the Internet and the microchip, and $\alpha = 1$ for the other cases. While the value $\alpha = 1/2$ is obtained in the case of a very simple model of independent random walkers visiting the sites of the network, the value $\alpha = 1$ arises when *externally induced* fluctuations are large (i.e., in the case of the simple model of random walkers, if the number of these walkers fluctuates strongly). As a confirmation of this prediction, de Menezes and Barabási (2004b) and Barabási *et al.* (2004) propose a method of separating the internal and external contributions to the dynamics. Namely, if $f_i(t)$ is the signal measured on node i ($i = 1, \ldots, N$) at time t, during a time window $[1, T]$, one can introduce the ratio of the total traffic through i during the observation time and the total overall traffic in the network

$$A_i = \frac{\sum_{t'=1}^{T} f_i(t')}{\sum_{t'=1}^{T} \sum_{j=1}^{N} f_j(t')}, \tag{11.1}$$

so that the expected amount of traffic through i can be written as

$$f_i^{\text{ext}}(t) = A_i \sum_{j=1}^{N} f_j(t). \tag{11.2}$$

Indeed, $\sum_{j=1}^{N} f_j(t)$ is the total traffic on the network at time t, and is assumed to be dispatched onto the various nodes in proportion to the values of A_i.

This expected traffic f_i^{ext} is thus a quantity which fluctuates because of externally induced time changes in the global network traffic. On the other hand, $f_i^{\text{int}} = f_i - f_i^{\text{ext}}$ captures the deviations from the traffic average through node i.

The systematic measure of (f_i^{ext},f_i^{int}) and of the corresponding variances (σ_i^{ext},σ_i^{int}) shows that $\sigma_i^{\text{ext}} \ll \sigma_i^{\text{int}}$ in the cases $\alpha = 1/2$ while $\sigma_i^{\text{ext}} \sim \sigma_i^{\text{int}}$ when $\alpha = 1$.[2] The Internet and the microchip are thus networks with internally robust dynamics, while the activity of the other three studied networks (WWW, the highway network, and the river network) is driven by the external demand.

Notably, such studies allow us to gain some understanding of global features of the traffic without requiring precise knowledge of the network structure. On the other hand, a theoretical modeling of traffic in large-scale networks cannot neglect the overall topology of the system. In this area most of the modeling activity is concerned with the self-similarity of traffic and the emergence of congestion in technological networks such as the Internet. These approaches propose that the presence of phase transition and emergent phenomena in simple models mimicking the routing of information packets is what lies behind the appearance of congestions and traffic self-similarity (Ohira and Sawatari, 1998; Takayasu, Fukuda and Takayasu, 1999; Fukuda, Takayasu and Takayasu, 2000; Solé and Valverde, 2001; Valverde and Solé, 2002). In general, the phase transition mechanism appears as an elegant explanation that arises as a complex emergent phenomenon due to the global dynamics of the network. Many other mechanisms which depend, however, on the specific nature of the system can be put forward and validated by specific experimental work. In contrast, the very abstract nature of statistical mechanics models does not often find clear-cut support from experimental data.[3] In the following we focus on a series of works that consider the effect of a network's complexity on routing algorithms and congestion phenomena clearly inspired by problems in the information technology world. These models are naturally fitting in the present book, but we warn the reader about the existence of a large body of literature that approaches these problems in the orthogonal perspective of a very detailed account of the microscopic processes as dictated by the specific technology under study.

11.2 Traffic and congestion in distributed routing

Most studies about traffic and congestion phenomena on complex networks focus on technological networks, particularly on packet switched dynamics such as the one used in the Internet to route data packets. In these attempts, the Internet is modeled as a network whose vertices are assumed to be hosts that generate a certain number of packets (the traffic to be routed) or routers that forward packets.

[2] Duch and Arenas (2006) show, however, that for models in which packets perform random walks on the network, varying the number of steps each packet remains in the network and taking into account the finite capacity of the nodes can lead to continuously varying exponents α between 1/2 and 1.

[3] An interesting discussion of the models' validation issues is provided by Willinger *et al.* (2002).

These packets travel to their destination following routing dynamics that, in general, involve a nearest neighbor passing of packets. Most studies consider for the sake of simplicity that all nodes can play both roles of hosts and routers, and only a few distinguish, in a more realistic way, between hosts which can generate and receive packets, and routers which can only forward them to another node (Ohira and Sawatari, 1998; Solé and Valverde, 2001). In this configuration the network's behavior strongly depends upon the injection rate of new packets. This is generally modeled by the creation of packets at the rate R on randomly chosen nodes of the network. This simply implies that an average of NR packets are created at each unitary time step. Each packet is also assigned a random destination at which it is annihilated upon arrival. The packets then travel from one node to another: at each time step a node can treat and forward a certain number of packets to its neighbors. Such a capacity is usually taken as uniform (Ohira and Sawatari, 1998; Arenas, Díaz-Guilera and Guimerà, 2001), but can also depend on the node's characteristics (Zhao *et al.*, 2005; Lin and Wu, 2006). Each vertex is also supposed to have a buffer (a queue) in which packets can be stored if they arrive in a quantity larger than the handling limit. Packets may accumulate indefinitely or can be discarded if they exceed the buffer limit. The forwarding rules, from a node to its neighbor, define the *routing policy*: they can either follow the shortest path on the network (Ohira and Sawatari, 1998; Solé and Valverde, 2001; Arenas *et al.*, 2001; Guimerà *et al.*, 2002a), or perform a random walk until they reach their destination or its neighborhood (Tadić, Thurner and Rodgers, 2004; Tadić and Thurner, 2004; Valverde and Solé, 2004), as shown schematically in Figure 11.1.

In the context of packet routing models, the total network load or traffic is given by the number of packets traveling simultaneously on the network. As shown in

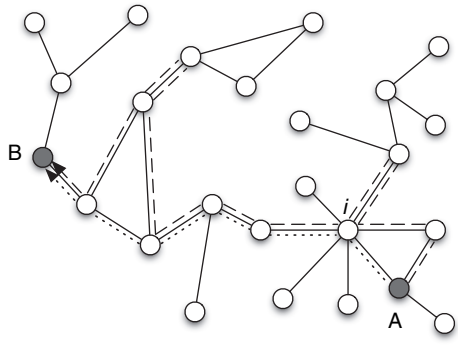

Fig. 11.1. Schematic comparison of shortest-path (dotted line) routing and random walk (dashed line) routing for a packet generated in A and sent to B. Node i has a large degree and betweenness centrality and will therefore typically receive high traffic.

https://doi.org/10.1017/CBO9780511791383.012 Published online by Cambridge University Press

Figure 11.2, this load fluctuates around a steady-state value if the creation rate is not too large. If, on the other hand, the routing strategy does not allow the delivering of packets quickly enough, they start accumulating in the network and the load steadily increases with time, defining a congested (sometimes called jammed) phase.

In order to distinguish between the free flow and congested phases, Arenas *et al.* (2001) define a specific global parameter (see also Guimerà *et al.* [2002a]) given by

$$\eta = \lim_{t\to\infty} \frac{n(t+\Delta t) - n(t)}{N R \Delta t}, \tag{11.3}$$

where $n(t)$ is the number of packets in the network of size N at time t and the $t \to \infty$ limit is taken in order to ensure that the transient initial state does not influence the result. The quantity $NR\Delta t$ is the number of packets injected into the system in the interval time Δt, and η is equal to or smaller than zero if the system is able to route to destination a number of packets larger than the generated traffic. Otherwise, if $\eta > 0$ the system enters a congested phase with packets accumulating in the network. Therefore, if the system is in a free flow phase, $\eta = 0$, while the congestion phase corresponds to $\eta > 0$. In practice, an average over a large enough Δt is performed, as fluctuations may create a negative η as well as transient positive values.

Analogous to phase transitions, the study of the congestion phenomena consists of identifying the critical R_c separating the free flow from the jammed phase, finding the mechanisms leading to the congestion, and understanding how the various ingredients of the model (network structure, routing policy) modify the congestion transition. Implicitly, the congestion transition also depends on the topology of the

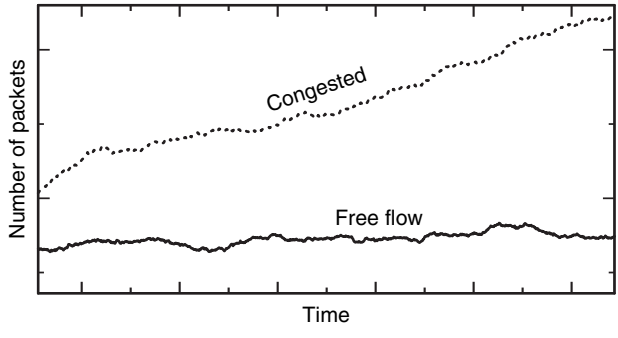

Fig. 11.2. Evolution in time of the number of data packets in the network: if the creation rate is small enough, this number fluctuates around an average value. If the creation rate is too large, the system becomes congested and the number of packets increases (roughly) linearly with time.

https://doi.org/10.1017/CBO9780511791383.012 Published online by Cambridge University Press

network as the routing dynamics is clearly affected by the underlying connectivity pattern. Indeed, most routing policies can be approximated by a shortest path routing from the origin to the destination node. This assumes that a routing table is associated to each node which indicates, according to all possible final destinations, to which neighbor the packet should be routed. Since each packet follows the shortest path to its destination, it is evident that the concept of betweenness centrality will play a central role in the study of traffic properties. We have seen in Chapter 1 that the betweenness of the vertex i is defined as

$$b_i = \sum_{h \neq j \neq i} \frac{\sigma_{hj}(i)}{\sigma_{hj}}, \tag{11.4}$$

where σ_{hj} is the total number of shortest paths from h to j and $\sigma_{hj}(i)$ is the number of these shortest paths that pass through the vertex i. Equation (11.4) is in fact very general, and can be used for any *static routing protocol* (SRP), i.e. assignment of routes between all $N(N-1)/2$ pairs of nodes.[4] In this case σ_{hj} and and $\sigma_{hj}(i)$ are the numbers of paths corresponding to the routing dynamics implemented in the network. In other words, the betweenness centrality as defined by the routing protocol is a first indicator of the relative traffic that a node may eventually handle.

Let us now consider a given SRP. Packets are generated at random, with random destinations (chosen among the $N-1$ other nodes) so that for each source h, the traffic generated toward each destination j is $R/(N-1)$. Therefore, the average number of packets handled at a node i at each time step is on average $Rb_i/(N-1)$, where b_i is the betweenness centrality due to the routing dynamics. Moreover, if each node i can handle c_i packets at each time step, the condition for the system not to be congested reads as

$$R\frac{b_i}{N-1} \leq c_i \quad \forall i, \tag{11.5}$$

which, in the simple case of a uniform capacity $c_i = 1$, gives the congestion threshold (Arenas *et al.*, 2001)

$$R_c = \frac{N-1}{\max_i b_i} = \frac{N-1}{b^*}, \tag{11.6}$$

where b^* is the largest value of centrality in the network; i.e. the value belonging to the most central node in the network which is also the first one to become congested. The above equation clearly shows the role of network topology in the onset of congestions as the maximum betweenness is determined by both the routing

[4] Newman (2005a) also defines the random walk betweenness centrality as the number of times a random walk between h and j passes through i, averaged over all pairs (h, j).

https://doi.org/10.1017/CBO9780511791383.012 Published online by Cambridge University Press

dynamics and the network connectivity pattern. The larger the maximal betweenness of the system, the smaller the traffic injection rate R_c at which the network results in congestion. In addition, the evolution of the congestion threshold with the system size can also be evaluated in terms of the scaling of the maximal betweenness with the number of nodes N. The simple expression (11.6) explains why homogeneous networks have typically much larger congestion thresholds than heterogeneous graphs which display broad distributions of both degrees and betweennesses. In most heterogeneous networks, the betweenness typically grows with the degree, so that the hubs become easily congested, leading to a small R_c even if most nodes are far from being overloaded (Echenique, Gómez-Gardeñes and Moreno, 2005). On the other hand, at small load, the presence of hubs can lead to particularly short paths, and therefore small transit times from one point of the network to another. The combination of these observations shows that the optimal structure of a traffic carrying network depends on the imposed global load (Guimerà *et al.*, 2002b). A star-like network is very efficient (with paths of length at most 2) if the number of packets is small. In the case of large traffic, on the other hand, the minimization of b^* corresponds to the most homogeneous possible network, in which the load is balanced between all nodes.

Once we understand the effect of the network topology on traffic, we have to cope with the fact that, in most cases, the structure of networks on which traffic takes place is given and cannot be changed at will,[5] and obtaining optimized topologies in a technologically efficient manner turns out to be quite challenging (Krause *et al.*, 2006). Given a fixed network topology, we will see in the following different routes to improve the traffic properties therefore need to be followed: refined static routing protocols may be defined in order to modify the betweenness properties; alternatively, adaptive protocols have been proposed, which dynamically change routing such that the most loaded nodes are avoided.

11.2.1 Heterogeneity and routing policies

Most networks are characterized by large heterogeneities in the nature of their constituent elements. This implies different capacity c_i of the elements in traffic handling. In the case of heterogeneous capacities of the nodes, Equation (11.6) can be rewritten as

$$R_c = \frac{(N-1)C^*}{b^*}, \tag{11.7}$$

[5] A noteworthy exception concerns the case of wireless ad hoc networks, whose topologies can be changed by modifying the transmission power of the nodes (Krause *et al.*, 2004; Glauche *et al.*, 2004; Krause, Scholz and Greiner, 2006).

where C^* is the capacity of the node with the largest betweenness. Since hubs are typically the first nodes to become overloaded, Zhao *et al.* (2005) propose to delay their congestion by considering a capacity which increases linearly either with the degree (see also Lin and Wu [2006]) or with the betweenness. In the second case in particular, the congestion threshold becomes independent of the topology: if $c_i = \beta b_i / N$ then $R_c = \beta$. The extension of the nodes' capacities may, however, not be easy, especially for large networks, since b^* may diverge with the network size as N^γ with $\gamma > 1$ (Danila *et al.*, 2006a,b; Sreenivasan *et al.*, 2007). Various attempts have thus been directed towards the optimization of the routing policy.

First, and not surprisingly, a comparison between random walks and deterministic shortest paths shows that the latter leads to better performances, although the former is of course easier to implement. In particular, Tadić and Thurner (2004) (see also Tadić *et al.* [2004]) consider that packets perform random walks on the network, but switch to a deterministic shortest path if they reach a node which is next-to-nearest neighbor of their destination. This small modification yields a large improvement, especially in terms of transit times. The crossover between random and deterministic paths has been further investigated by Valverde and Solé (2004) who consider that when a randomly diffusing packet reaches a node which is at distance m from its destination, it follows the shortest path to this destination. The parameter m determines the change from a fully random walk at $m = 0$ to a fully deterministic routing using only shortest paths at values of m larger than the network diameter. Although congestions are not considered by Valverde and Solé (2004) (the packets creation rate is tuned below the congestion threshold) they have shown that large values of m determine a much better performance with smaller load imposed on the network.

Since pure shortest path routing leads to rapid overloading of the hubs, while pure random walks are rather inefficient (and tend to overload hubs, see Chapter 8), routing strategies aimed at obtaining larger congestion thresholds naturally try to avoid paths through hubs, at the possible cost of increasing the path length for the transmitted packets. Such *hub avoidance strategies* can be defined in various ways. Yan *et al.* (2006), for instance, propose to use between any source h and destination j the path $\mathcal{P}_{hj}$ which minimizes $\sum_{i \in \mathcal{P}_{hj}} k_i^\beta$, where β is a parameter of the routing strategy. For $\beta = 0$ the usual shortest path routing is recovered. For $\beta > 0$ the paths, when possible, go around the hubs and R_c is noticeably increased. While computing such optimal paths implies the knowledge of the whole topology of the network and may be computationally very expensive and hardly feasible in real-world applications, local stochastic strategies can be devised. A node can forward a packet to a neighbor i with a probability $\propto k_i^\alpha$; i.e. depending on the degree of the neighbor itself. The value $\alpha = -1$ turns out to be optimal in increasing R_c, as

https://doi.org/10.1017/CBO9780511791383.012 Published online by Cambridge University Press

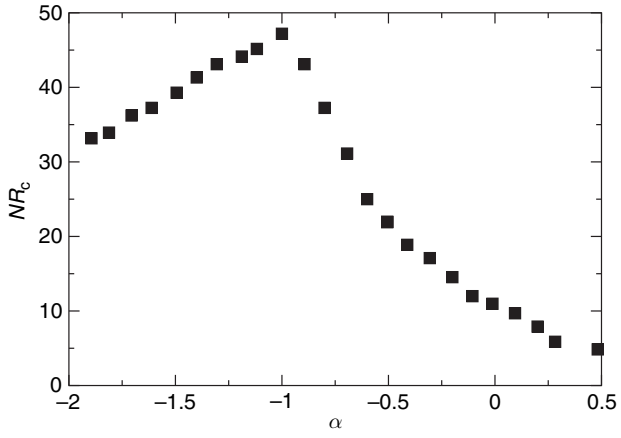

Fig. 11.3. Congestion threshold R_c as a function of the hub avoidance parameter α: each packet sitting at a node is forwarded to one of its neighbors i with probability $\propto k_i^{\alpha}$. An additional rule which forbids the packet to travel back and forth on the same link is used to obtain reasonable values of R_c. Here the network used is a Barabási–Albert model with $m = 5$, $N = 1000$. Each node can treat and forward 10 packets at each time step. Data from Wang *et al*. (2006a).

displayed in Figure 11.3 (Wang *et al.*, 2006a). Indeed, it can be shown (for uncorrelated networks) that the average number of packets located at a node i depends on its degree according to the form $n_i \propto k_i^{1+\alpha}$. The load is thus most evenly balanced between various degree classes when $n_i = \text{const}$, yielding the condition $1 + \alpha = 0$.

The quest for increasing the congestion threshold R_c has led to the proposal of several other routing strategies that consider explicitly the network topology. Sreenivasan *et al*. (2007) put forward another hub avoidance strategy: (i) starting from a given network, a certain number of hubs are "removed", which typically will disconnect the network in various clusters (see Chapter 6); (ii) in each of these clusters, routing between each pair of nodes is assigned through shortest paths; some of these shortest paths may differ from (and be longer than) the shortest paths on the original network, which may pass through the removed hubs; (iii) the hubs are put back into the network and all pairs of nodes which do not yet have a routing path are joined through shortest paths. In this way, some of the paths assigned are not the shortest paths on the full network and avoid hubs. Danila *et al*. (2006b) alternatively, study a heuristic iterative algorithm that, given a network structure, minimizes the difference between the maximum and the average betweenness, i.e. obtains a distribution of betweenness as narrow as possible. Such a goal is obtained by iteratively assigning larger "costs" to the edges connected to

https://doi.org/10.1017/CBO9780511791383.012 Published online by Cambridge University Press

the node with largest betweenness. The growth of b^* with N is substantially lowered, but the algorithm computational time scales as $\mathcal{O}(N^4)$, which is too large for most real networks.

An extremely valuable result has been derived by Sreenivasan *et al.* (2007) who found that the network topology in fact induces a lower bound for b^*, whatever the routing protocol. Different static routing protocols can lead to different nodes having maximal betweenness, and different betweenness distributions, but b^* cannot be decreased indefinitely. For a scale-free network with degree distribution $P(k) \sim k^{-\gamma}$, this bound is estimated as $\mathcal{O}(N^{\gamma/(\gamma-1)})$: as $\gamma \to 2$, the network becomes increasingly star-like and the betweenness of the most central node scales as N^2 while, as $\gamma \to \infty$, one obtains a homogeneous network for which the best possible scaling of b^* with N is obtained. Consequently, the congestion threshold for deterministic routing algorithms has an upper bound due to the topology. Such reasoning does not, however, apply to adaptive strategies, considered in the next section, in which each node decides in real time the forwarding routes according to the local information about the most loaded nodes.

11.2.2 Adaptive (traffic-aware) routing policies

The fundamental feature of the traffic-aware routing policies is the possibility for each node to gather information about the load status of its neighbors. The decision of where to forward a packet can thus be taken in order to avoid overloading nodes with large queues. In the context of complex networks, such a scheme has been proposed by Echenique, Gómez-Gardeñes and Moreno (2004). In this work, the routing dynamics incorporates a tunable amount of information about traffic into the shortest-path-based routing. Let us consider that each node can treat and forward only one packet at each time step (see Chen and Wang [2006] for the case of a capacity depending on the degree) and that the packet held by a certain node l has destination j. For each neighbor i of l, the effective distance to j is then defined as

$$\delta_i = hd_i + (1-h)q_i, \tag{11.8}$$

where d_i is the shortest-path distance between i and j, and q_i is the number of packets held by i (the length of its queue). The parameter h controls the interplay between traffic and shortest paths and, for $h < 1$, packets may be diverted to longer but less congested paths. The forwarding decision of node l can then either be probabilistic, each neighbor i having probability $\exp(-\beta\delta_i)$ of being chosen (Echenique *et al.*, 2004), or deterministic (Echenique *et al.*, 2005). According to the latter rule the packet is then sent to the neighbor i which minimizes the effective distance δ_i. For $h = 1$, the usual transition to congestion is observed as shown in Figure 11.4. The order parameter η increases continuously from 0 at $R < R_c$

https://doi.org/10.1017/CBO9780511791383.012 Published online by Cambridge University Press

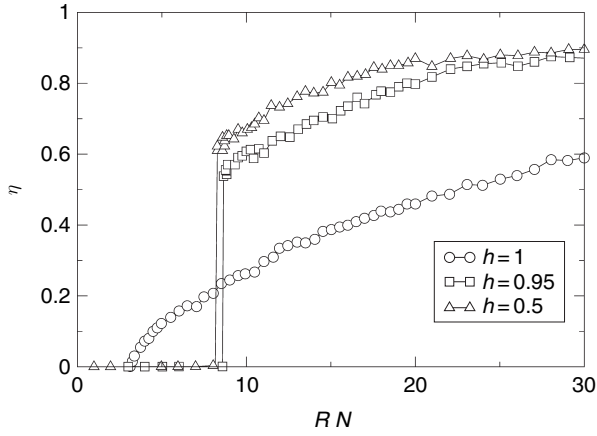

Fig. 11.4. Jamming transitions in the traffic-aware scheme proposed by Echenique *et al.* (2005), for various values of the parameter h (see Equation (11.8)). The order parameter η is given by Equation (11.3). The case $h = 1$ corresponds to the standard shortest path routing with no traffic awareness. Data from Echenique *et al.* (2005).

to positive values at $R > R_c$. On the other hand, as soon as some amount of traffic-awareness is included, $h < 1$, the appearance of congestion is retarded, i.e. R_c increases. It must be stressed that the parameter η presents a discontinuous jump at the transition and takes values larger than in the case $h = 1$. The study of congestion levels as a function of the node's characteristics and as a function of time, displayed in Figure 11.5, allows an understanding of how congestion arises and how the traffic-aware policy improves the situation. For $h = 1$ (usual shortest path routing), the nodes with large betweenness are congested while the nodes with small betweenness are not overloaded. In the case of $h < 1$, the nodes with large betweenness become congested first, but the traffic-awareness then leads the packets to follow paths which avoid these congested nodes, slowly spreading the congestion to the rest of the network. The interest of the traffic-aware scheme with respect to the shortest path routing thus actually depends on a trade-off between the value of R_c and the way in which congestion arises (continuously or abruptly).

Variations on the basic idea of dynamically avoiding the most loaded nodes have proposed routing mechanisms biased by the value of the queue q_i of each network element. In particular, each node l can decide stochastically to forward a packet, each neighbor i being chosen with probability $\propto (1 + q_i)^{-\beta}$ (Danila *et al.*, 2006a), or $\propto k_i(1 + q_i)^{-\beta}$ (Wang *et al.*, 2006b). It turns out that an optimal value for β is obtained (namely $\beta = 1$ for the first case, $\beta = 3$ for the second case), with a strong increase in the observed values of R_c. While $\beta > 0$ allows the most loaded nodes to be avoided, an excessively high β, i.e. too strict a load avoidance policy,

https://doi.org/10.1017/CBO9780511791383.012 Published online by Cambridge University Press

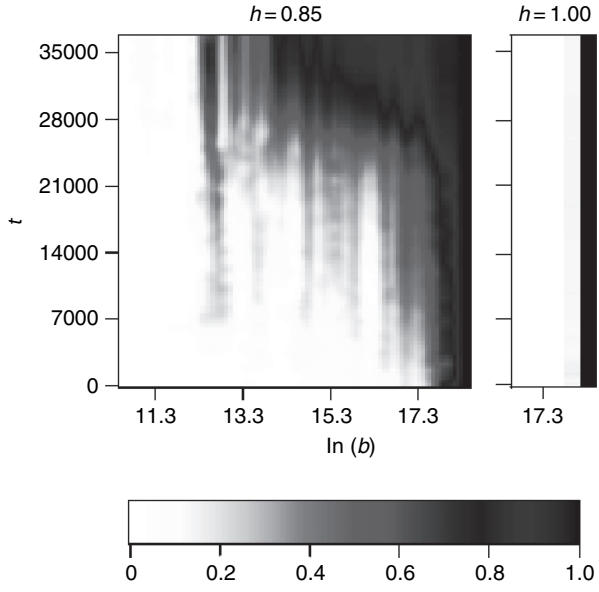

Fig. 11.5. Congestion levels in the congested phase as a function of time t and nodes' betweenness b. At each time step, the gray scale is normalized by the number of packets in the queue of the node with the largest congestion. Two radically distinct behaviors are obtained for the standard ($h = 1$, right panel) and the traffic-aware ($h = 0.85$, left panel) protocols. In both cases $R > R_c$. From Echenique *et al.* (2005).

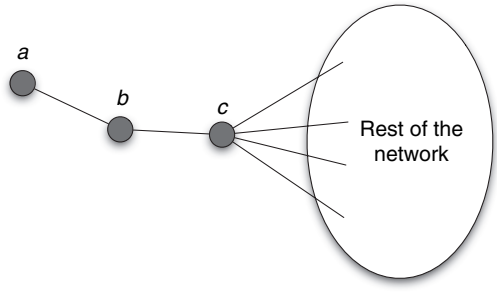

Fig. 11.6. "Traps" in the congestion aware scheme of Danila *et al.* (2006a). At large β, many packets get trapped between nodes b and a.

is counter-productive since it leads to the appearance of "traps" such as the one depicted in Figure 11.6 (Danila *et al.*, 2006a). The reason is that nodes a and b in Figure 11.6, having small betweenness, will typically receive few packets and have small queue lengths. Node c would then tend to send packets to b, and these packets would then get stuck for a long time traveling back and forth between a, b, and c. A value of β that is not too large allows the avoidance of such blocking phenomena.

11.3 Avalanches

In the previous section, we have considered the problem of how congestion arises (and may be delayed thanks to clever routing strategies) when the rate of packet creation increases in an otherwise intact and functioning network. Once congestion arises, the overloading of nodes may lead to their failure and their traffic will have to be redistributed to other elements of the network. This may induce the overload of other elements and eventually trigger other failures, giving rise to a cascade or avalanche of failures which propagates in the network. While in some cases the avalanche may just involve a limited number of elements, in other occasions it may affect a sizeable portion of the network leading to a macroscopic breakdown of the entire system. Avalanche dynamics is not only found in technological networks. A wide variety of phenomena ranging from material science to biological and social systems exhibit avalanche phenomena triggered by a progressive rearrangement of the system under the failure of some of its elements. Two ingredients appear as fundamental in the modeling of avalanche and breakdown phenomena in heterogeneous media (Herrmann and Roux, 1990; Alava, Nukala and Zapperi, 2006). First, the existence of local thresholds ensures that, above a certain level of stress, a local element will fail. The second important point lies in the existence of disorder (due to structural heterogeneities): a model in which all elements fail at the same time would be neither realistic nor very interesting. In the context of complex networks, the determination of the heterogeneity of the thresholds is a non-trivial point, since little is known about the real capacity of each element of real-world networks. In the absence of reliable information, one possibility is to distribute randomly the thresholds of the links, without any a priori correlations with the underlying topology. This corresponds to models akin to the fiber bundle model (FBM). Another choice corresponds instead to the hypothesis that the *actual* capacity of each link or node is strongly linked to its topological centrality as given by the betweenness centrality. The rationale behind this approach comes from the fact that the betweenness centrality gives the amount of all-to-all traffic (i.e. each node sends one data packet to each other node of the network) if shortest paths are used for transmitting the data, as seen in the previous section.

In all cases, the initial failure may be caused either by the progressive increase of an externally imposed force or load (such as the increasing number of data packets in a congested network), which will typically lead to the failure of the weakest or most overloaded part of the network, or by the sudden (accidental) failure of a given node or link; the cascading consequences of this link or node removal will then depend on its topological characteristics, just as in the case of random failures versus targeted attacks explored in Chapter 6.

11.3.1 Breakdown models

Classical models for avalanches and network breakdowns were first put forward by physicists in the context of disordered material science. While these models were formulated with physics problems in mind, they represent simple abstractions of the physical phenomena and as such can be easily generalized to other systems. These models share the presence of an external load or force which is acting globally on the system and can trigger the failure of network elements.

The random fuse network (RFN) is a simple model (de Arcangelis, Redner and Herrmann, 1985) which describes the breakdown of materials and electric grids. In this model, fuses form a regular lattice on which an electrical current flows. The fuses have a threshold which can fluctuate from one link to the other reflecting the heterogeneity in the system. When the local current reaches the fuse threshold, the corresponding fuse breaks and the current is then redistributed among the surviving links according to Kirchhoff's laws. This current redistribution may lead to an overload of some other fuses which fail in turn, triggering a failure cascade with the ability to break the system into disconnected parts. The RFN model thus demonstrates how the failure of a single element may trigger a cascade leading to a macroscopic breakdown. Among a number of general results, the existence of a macroscopic breakdown current I_c is of particular interest. When the applied external current I is increased, more and more fuses break, some causing small avalanches, until a global failure avalanche breaks the system for $I = I_c$. The value of this breakdown current I_c depends on different parameters such as the threshold distribution and the geometry of the lattice.

Some simple analytical arguments for the RFN on regular lattices might be helpful for the understanding of more involved cases. Let us consider a very large rectangular lattice of fuses with the same conductivity, and thresholds distributed according to $\phi(i_c)$ which we will assume to be uniform in the interval $[\langle i_c\rangle - w, \langle i_c\rangle + w]$. When the applied current is increased very slowly, the weakest fuse with threshold $\langle i_c\rangle - w$ fails first. The current is then redistributed on the lattice (Clerc *et al.*, 1990); for example, the current flowing through the neighbors of the broken link is multiplied by some factor $\alpha > 1$, equal to $\alpha = 4/\pi$ for an infinite rectangular lattice (the multiplication factor decreases with the distance of the links from the breaking point). If α is small or if w is large enough, these nearest neighbors will probably be able to cope with the current increase and will not break. On the other hand, a sufficient condition in order to trigger an avalanche leading to a global breakdown of the system is that the redistributed current is larger than the largest threshold in the system:

$$\alpha(\langle i_c\rangle - w) > \langle i_c\rangle + w. \tag{11.9}$$

https://doi.org/10.1017/CBO9780511791383.012 Published online by Cambridge University Press

This equation implies an important qualitative result. If the disorder is weak (small w) then the failure of the first fuses will very likely trigger a very large avalanche ("brittle" behavior). In contrast, if the threshold disorder is large, the increase of the applied current will break more and more fuses leading to microcracks which will grow and lead to global breakdown when they coalesce ("ductile" behavior).

In a similar spirit, the fiber bundle model represents a simplification of the RFN (Herrmann and Roux, 1990): instead of recalculating the currents in all the network when a link (or fiber) breaks, one assumes that the current is redistributed equally either on all the remaining links, or simply on the nearest neighbors of the failure. As in the RFN, each of the initially intact N fibers has a local threshold i_c distributed according to some distribution $\phi(i_c)$; if the applied current (or force) I is equally divided on all the fibers, a fiber breaks when $I/n(I)$ exceeds its threshold i_c, where $n(I)$ is the number of remaining fibers, which thus obeys

$$n(I) = N\text{Prob}(I/n(I) < i_c) = N \int \mathrm{d}i_c \phi(i_c) \theta \left(i_c - \frac{I}{n(I)} \right), \qquad (11.10)$$

where θ is the Heaviside function. Depending on the disorder distribution, the model exhibits different regimes of abrupt rupture or more progressive damage leading to global breakdown.

Although these models have been extensively studied on regular lattices, they can also describe avalanches in complex networks. In particular, Moreno, Gómez and Pacheco (2002a) consider a version of the fiber bundle model in which each node (rather than each link) of a Barabási–Albert network represents a fiber, with a random threshold distributed according to a given probability distribution. For definiteness, a Weibull distribution is taken, with a variance depending on a parameter ρ (the larger ρ, the smaller the variance and the narrower the range of threshold values). A global force or load F is applied to the network and initially equally divided among the nodes. When the load on a node is larger than its threshold, the node breaks and the load is equally redistributed among its neighbors, possibly causing an avalanche. As the externally applied force slowly increases, more and more nodes break, leading finally to a global destruction of the network at a critical load σ_c which depends on ρ: for larger ρ, i.e., for narrower distribution of thresholds, the rupture of the network becomes more abrupt. In this respect, the behavior of the scale-free network is no different from a regular lattice. A more detailed inspection shows that when the critical load is approached, the fraction of overloaded nodes is an increasing function of the degree; the macroscopic breakdown is reached when the hubs start to fail. When the last hubs are eventually overloaded, the network fails. This result is of course linked to the vulnerability of heterogeneous networks with respect to the removal of the most connected nodes (see Chapter 6), and appears here as a consequence of the local dynamics: the hubs

have more neighbors and thus have a higher probability of receiving additional load from the breaking of a neighbor (see also Kim, Kim and Jeong [2005]).

11.3.2 Avalanches by local failures

As previously mentioned, the failures in a network can also occur because of the random breakdown of one element, whose load has to be redistributed. Other elements may become overloaded and fail, leading to a new redistribution, and so on. This cascade of events stops when all the elements of the network have a load below their respective capacity. The size of the avalanche is given by the number of nodes which have been broken in the process, and the damage is quantified by the size of the largest connected component remaining in the network, in a way similar to Chapter 6. In this case we are not driving the system with an external quantity (force, generated traffic etc.), and the dynamics of the avalanche propagation and the extent of the damage, depend on the specific system properties and dynamics such as the capacities of the nodes or the redistribution rules.

In the absence of precise knowledge about the traffic flow on real networks, the simplest hypothesis considers that in the initial state the load or traffic on each link (i, j) is a random variable g_{ij} drawn from a probability distribution $U(g)$ (Moreno *et al.*, 2003b), and that the local threshold or capacity is uniform ($c = 1$ for each link). A failure can be simulated by selecting a link at random and overloading it, by raising its traffic to $g_{ij} = c$. At this point the traffic of the link is redistributed among the (non-overloaded) links which depart from the end nodes of the congested link. This model is once again thought for networks carrying physical quantities or information packets such as power grids and the Internet. Two types of redistribution can be considered. The first process attempts to deterministically redistribute the load equally among the links. In contrast, "random redistribution" consists of randomly distributing the corresponding load across the neighbors. When a failing link has no active neighbors, its load is equally shared among the remaining functioning links ("conserved dynamics") or can be considered as lost ("dissipative" case). Moreno *et al.* (2003b) compute the phase diagram of such a model when the initial system is a scale-free Barabási–Albert network: the order parameter is given by the probability P_{G} of having a giant component of connected nodes (i.e. with $g_{ij} < c$) of extensive size *after the avalanche*. If this quantity is equal to one, the communication or transport capabilities of the network still function with probability 1. In contrast, when P_{G} is zero, then no information or quantity can propagate from one node of the network to another. The numerical results distinguish three different regimes depending on the value of the average initial load $\langle g \rangle$, as shown in Figure 11.7. For $\langle g \rangle < g_{\mathrm{C}}^{\mathrm{I}}$ where $g_{\mathrm{C}}^{\mathrm{I}}$ depends on the system details and on the redistribution process, one obtains $P_{\mathrm{G}} = 1$. For

https://doi.org/10.1017/CBO9780511791383.012 Published online by Cambridge University Press

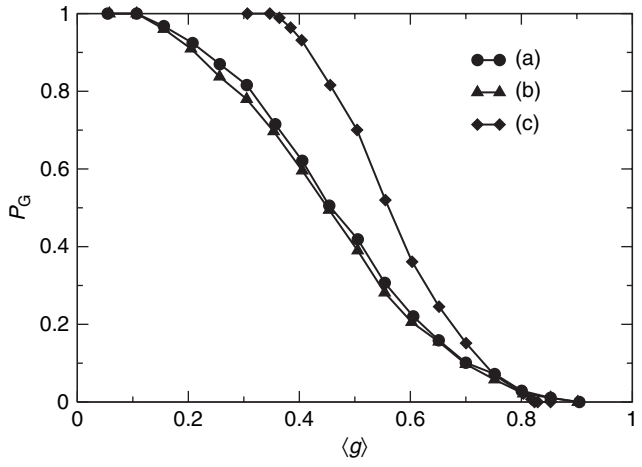

Fig. 11.7. Probability P_G of having a giant component of connected nodes of extensive size after the avalanche caused by the overloading of one link, as a function of the average initial load $\langle g \rangle$, in the model of Moreno *et al.* (2003b). Three redistribution rules are shown: (a) random and dissipative, (b) random and conserved, (c) deterministic and conserved. In all cases the critical values g_C^I and g_C^{II} are easily identified. The initial network has size $N = 10^4$. Data from Moreno *et al.* (2003b).

$g_C^I < \langle g \rangle < g_C^{II}$, P_G becomes smaller than 1 which signals that with a non-zero probability the network can develop a macroscopic breakdown. This region can be considered as fragile with respect to a local congestion. Finally, for $\langle g \rangle > g_C^{II}$, any small instability leads to a completely congested network i.e. $P_G = 0$. A small value is obtained for g_C^I, and quite a large one for g_C^{II}, defining a wide region of load values in which a small instability can propagate and lead to a global congestion with a finite probability. This is indeed what is observed in real congestion or failure phenomena in which apparently similar local disturbances lead either to a global failure or only to a very localized perturbation.

11.3.3 Avalanche and routing dynamics

In Section 11.2, the congestion phenomena have been studied using various routing protocols that are often based on shortest paths. A natural hypothesis considers therefore that the betweenness centrality represents a first order approximation of the real traffic in a technological network. If the capacities of the various elements are uniform, the large betweennesses of the hubs lead to a small congestion threshold. Moreover, the mere fact that a network evolves and grows leads to the creation of more and more shortest paths, i.e. to a traffic increase, which tends to accumulate on certain nodes, leading to a high fragility, as shown by Holme and Kim (2002b)

https://doi.org/10.1017/CBO9780511791383.012 Published online by Cambridge University Press

in the case of a growing Barabási–Albert network. In this case, the nodes with large degree (and therefore large betweenness centrality) easily become overloaded and their removal yields both an important damage and an important redistribution phenomenon, leading eventually to a fragmented network. Another possibility is to assign, together with an initial traffic given by the betweenness centrality, a maximal capacity c_i for each node i *proportional* to its betweenness centrality. The parameter of the model is given by the ratio between capacity and initial traffic g_i: $c_i = (1 + \alpha)g_i$, where the constant $\alpha \geq 0$ is called *tolerance parameter* of the network (Motter and Lai, 2002; Crucitti, Latora and Marchiori, 2004; Kinney *et al.*, 2005). If a node fails, the betweenness centralities of all other nodes change, since shortest paths have to be redirected. The load of some nodes may then be increased above their capacity, triggering an avalanching phenomenon. The extent of damage depends on the tolerance parameter: for large α the cascade remains local with few nodes being overloaded, while as α goes to zero the removal of a single node may lead to the entire breakdown of the system. Moreover, the topological characteristics of the initially broken node determine the extent of damage (Motter and Lai, 2002). As can be intuitively understood from Chapter 6, the removal of a node with large centrality is much more likely to produce significant damage than the removal of a low-centrality node. In the framework considered here, this is because the amount of load to redistribute is then large. In fact, the study of various types of networks shows that global cascades are triggered if the centrality is broadly distributed and if nodes with large centrality fail (Motter and Lai, 2002). In the case of networks with highly correlated degrees and centralities, the failure of a highly connected node easily leads to large damage. For homogeneous networks with narrow betweenness distributions, on the other hand, all nodes are essentially equivalent and large damages are obtained only if the tolerance α is very small.

11.3.4 Partial failures and recovery

The reaction of a network's properties to congestion or to a breakdown of one element may be less drastic than the simple failure of nodes. For instance, Crucitti *et al.* (2004) consider that each link (i, j) has a certain efficiency e_{ij}, which is simply decreased if the link's extremities are congested. The capacity of each node is supposed to be proportional to the initial betweenness centrality ($c_i = (1+\alpha)b_i$) and the links' efficiencies are taken initially uniform ($e_{ij}(0) = 1$ for all links). The efficiency of a *path* between any two nodes of the network is then defined as the inverse of the weighted distance along this path, where the weight of each link is the inverse of the efficiency: the more efficient a link (i, j), the smaller the "distance" between i and j. If ϵ_{lm} denotes the efficiency of the most efficient path between nodes l and m, then

https://doi.org/10.1017/CBO9780511791383.012 Published online by Cambridge University Press

$$\epsilon_{lm} = \max_{\mathcal{P}_{lm}} \left(\sum_{(i,j)\in\mathcal{P}_{lm}} \frac{1}{e_{ij}} \right)^{-1}, \tag{11.11}$$

where $\mathcal{P}_{lm}$ are paths between l and m, and the global efficiency of the network is simply

$$E = \frac{1}{N(N-1)} \sum_{l\neq m} \epsilon_{lm}. \tag{11.12}$$

The load of node i at any time is assumed to be equal to the number of the most efficient paths going through i. The failure of one node leads to a redistribution of the most efficient paths and possibly to the overloading of some nodes. In contrast with the cascades considered in the previous subsection, overloaded nodes are not removed from the network but become less efficient: if a node i has at time t a load $g_i(t)$ larger than its capacity c_i, the efficiency of all links between i and its neighbors j is reduced according to $e_{ij}(t) = e_{ij}(0)c_i/g_i(t)$. The global efficiency of the network drops in a way depending both on the tolerance parameter α and on the initial node removed. Interestingly, the system can either reach a steady value of efficiency or display oscillations due to the fact that overloaded nodes can become efficient again at later times. As α is decreased, the efficiency of the network after the cascading event undergoes a sharp decrease at a certain value α_c which depends on the type of removed nodes (see Figure 11.8): α_c is smaller for random removals than in the case where the node with the largest initial load is removed. The two

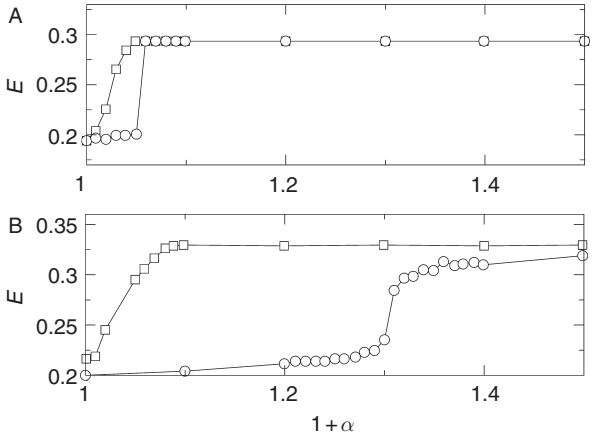

Fig. 11.8. Effect of a cascading failure in (A) an homogeneous ER graph and (B) a BA network, triggered by the removal of either a randomly chosen node (squares) or the node with the largest load (circles). The figure shows the final efficiency of the network vs. the tolerance parameter α. Both networks have size $N = 2000$. Data from Crucitti *et al.* (2004).

https://doi.org/10.1017/CBO9780511791383.012 Published online by Cambridge University Press

critical values are moreover very close for homogeneous graphs, but quite different for BA networks: a range of tolerance parameters in which the damage on the BA network is small for random removal but strong for targeted removal is obtained. Similar results are obtained for internet maps or the US power grid, and in fact for any network displaying a broad distribution of betweennesses. A similar but more detailed study in the case of the US power grid shows the existence of a large number of nodes whose removal causes almost no damage to the network, while there exists a small number of nodes whose removal has a very strong impact on the efficiency of the network (Kinney *et al.*, 2005).

11.3.5 Reinforcement mechanisms

The study of congestion phenomena and possible subsequent avalanches helps to rationalize the behavior of infrastructure networks: such networks typically can daily support a large number of random failures without any global damage, while dramatic consequences such as electrical blackouts may be triggered by apparently small initial events. In heterogeneous networks with broad distribution of betweenness centralities, random failures will most often concern nodes with small centrality, whose removal has little impact. On the other hand, a random failure may sometimes (even if with small probability) occur on a large-betweenness node, having the same effect as a targeted attack. In such a case, a large cascading event and a global failure of the network would follow.

A crucial issue is how to improve the robustness of networks or avoid a cascade propagation in a network. It is intuitive that the addition of redundant edges on top of an existing structure may strongly decrease its vulnerability (da Fontoura Costa, 2004). For a given number of nodes and edges, optimizing the sum of thresholds for random and targeted attacks leads to a design in which all nodes have the same degree, except one which has a degree of order $N^{2/3}$ where N is the size of the network: such a design is, however, not very realistic (Paul *et al.*, 2004).

Motter (2004) has proposed an adaptive defense mechanism against cascading failures. It applies to models in which the load of a node is given by the number of shortest paths passing through it, with a capacity proportional to the initial load. It moreover relies on the idea that the initial failure can be detected *before it propagates* throughout the network. The action of immediately removing (or switching off) a certain number of nodes with the smallest loads then allows the cascade to be stopped (see Figure 11.9): these nodes contribute to the global traffic but since they are quite peripheral they do little for its handling. Cutting them out thus reduces the global traffic without causing too much damage in terms of efficiency. The

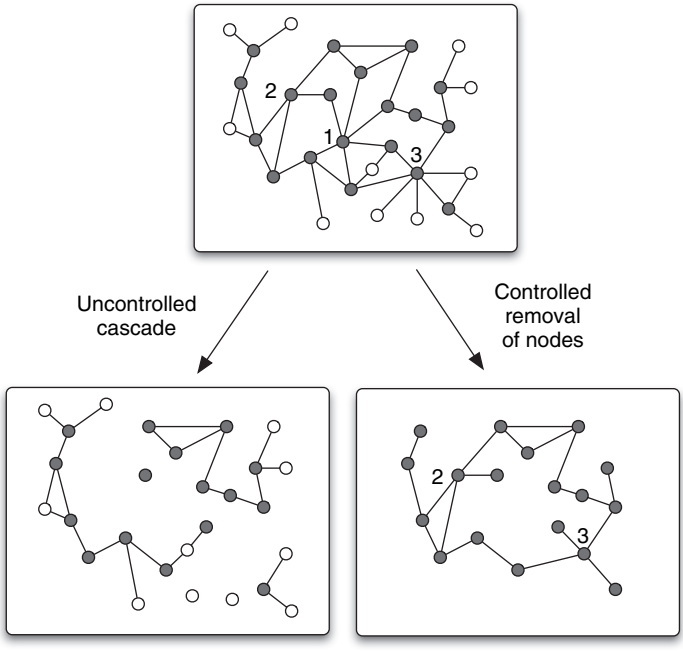

Fig. 11.9. Illustration of the defense mechanism proposed by Motter (2004). The cascade is provoked by the failure of node 1, which is removed from the network. Subsequently, nodes 2 and 3 become overloaded and fail as well, leading to a fragmentation of the network. If, on the other hand, a certain number of the most peripheral nodes, shown in white in the original network, are purposely removed (switched off) right after the failure of node 1, the traffic is alleviated, avoiding the failure of nodes 2 and 3, so that a large functioning component remains connected.

obtained size of the remaining giant component when using this strategy may then be much larger than for a uncontrolled cascade. The assumptions made here – such as the ability to detect the beginning of an important cascade quickly, to distinguish it from one of the many harmless random failures, and to remove nodes easily from the network *before* the cascade propagates – show the need for more investigations into the reinforcement issue.

11.4 Stylized models and real-world infrastructures

At the end of this chapter it is worth stressing again that while the approaches we have presented here are quite general and elegant, they do not account for many features of real-world networks. This has led to criticisms and a general skepticism from the computer science and engineering communities toward these approaches. Let us for instance consider the physical Internet. The routing policies of the

Internet are heavily affected by peering agreements between internet providers, commercial partnerships and other factors that make this system quite different from the stylized models used to study congestion phase transition. Another important issue is that even if the Internet and other information technology networks are self-organized complex systems, they possess, at the smaller scale, a high level of engineering design. In this sense the complex network topologies used in the models are not very good candidates for the modeling of small-scale properties of the network which on the other hand can also heavily affect the onset of congestions and traffic jams. This is also true for transportation systems. Analogously, avalanche processes are strongly dependent on the local nodes' properties and their reaction in case of failure and overload. In simple terms, it is hard to believe that a simple model could account for both a computer network and the railway system.

The value of the simplified models presented in this chapter stems from the fact that while the elements and technologies from which these networks are made are constantly changing, we can reasonably think that similar large-scale patterns and behaviors can be considered at the core of very different systems. This belief is not unmotivated but actually derives from the lesson we learned from statistical physics. The study of matter has taught us that although many materials or physical systems differ at the microscopic level, their large-scale behavior at critical points is essentially identical. For instance, a liquid–gas transition might appear very different from a change in the magnetic properties of ferromagnetic materials. However, both transitions are described by the same statistical laws and show the same phase diagram. More surprisingly, when emergent phenomena take over at the critical point, both systems exhibit the same scaling behavior even at a quantitative level. This feature, dubbed "universality," originates from the fact that when emergent cooperative phenomena set in, the large-scale behavior is essentially determined by the basic symmetry of the interactions and dynamics of the system's components, while other microscopic details become irrelevant (see also Chapter 5 and Chapter 13). In general, this is a common feature of complex systems with emergent properties. Naturally, the strict notion of universality must be relaxed when we leave the domain of phase transitions. However, the qualitative conservation with respect to changes of the very local details, of large-scale emerging properties such as heavy tailed-distributions or the absence of characteristic lengths, is a general property of cooperative phenomena. We can hope, therefore, that a large-scale view of traffic on networks is no exception to this general scenario. On the contrary, the statistical physics approach might acquire a particular conceptual value, being focused on the properties that are likely to be preserved despite continuous technological changes.

It is important, however, that researchers interested in the large-scale view do not neglect the engineering and technology details. In particular, if the statistical approach is extremely useful in the quest to understand the basic mechanisms of a phenomenon, its predictive power can be limited for each particular case. Any real-world application and understanding is usually also based on these detailed features, and a complete view of networks can therefore be obtained only by combining the different perspectives in a unified approach.