Appendix 3

Percolation in directed networks

In Chapter 6, the case of percolation and resilience of undirected graphs has been considered. Various authors have tackled the problem of its generalization to directed networks (Newman *et al.*, 2001; Schwartz *et al.*, 2002; Dorogovtsev *et al.*, 2001a; Boguñá and Serrano 2005).

As seen in Chapter 1, directed networks are topologically richer and more complex than their undirected counterparts. The links attached to a node can be either incoming or outgoing. The number of in-links is the in-degree $k_{\rm in}$, and the number of out-links is the out-degree $k_{\rm out}$. Since some pairs of nodes can be linked by two directed links pointing in the opposite directions, it is also possible to distinguish between strictly incoming, strictly outgoing or bidirectional links, the total degree being the sum of their respective numbers $k_{\rm i}$, $k_{\rm o}$, $k_{\rm b}$, with $k_{\rm in} = k_{\rm i} + k_{\rm b}$ and $k_{\rm out} = k_{\rm o} + k_{\rm b}$. The degree distribution is thus the joint distribution of these degrees. Most often, strictly directed networks are considered, with no bidirectional links: $k_{\rm b} = 0$. Note that by a simple conservation rule, the averages of $k_{\rm i}$ and $k_{\rm o}$ are then equal: $\langle k_{\rm i}\rangle = \langle k_{\rm o}\rangle$. Each (weakly) connected component (WCC) of a directed graph has also an internal structure as described in Chapter 1, and is composed of a strongly connected component (SCC), an in and an out components, tendrils and tubes. The giant WCC (GWCC) appears at the percolation threshold for the undirected version of the graph. On the other hand, giant SCC, in and out components as sketched in Figure 1.2 appear at another phase transition, as explained in the following.

A3.1 Purely directed networks

Let us first consider the case of purely directed networks: all links are directed, and no bidirectional links exist. The formalism of generating functions (see Appendix 2) can be adapted (Newman *et al.*, 2001) to this case by defining $p_{k_{\rm i}k_{\rm o}}$ as the probability for a node to have $k_{\rm i}$ incoming links and $k_{\rm o}$ outgoing links, and its generating function of two variables $\sum_{k_{\rm i}k_{\rm o}} p_{k_{\rm i}k_{\rm o}} x^{k_{\rm i}} y^{k_{\rm o}}$. We follow here instead the simple argument given by Schwartz *et al.* (2002) for the existence of a giant connected component. Let us consider a site b, reached from another site a. In order for it to be part of the giant component, such a site must have on average at least one outgoing link:

$$\langle k_{\rm o}(b)|a \to b\rangle \geq 1, \tag{A3.1}$$

https://doi.org/10.1017/CBO9780511791383.017 Published online by Cambridge University Press

and, if $P(k_\mathrm{i}, k_\mathrm{o}|a \to b)$ is the probability for b to have in-degree k_i and out-degree k_o *knowing that* there is a link from a to b, we obtain

$$\sum_{k_\mathrm{i},k_\mathrm{o}} k_\mathrm{o} P(k_\mathrm{i}, k_\mathrm{o}|a \to b) \geq 1. \tag{A3.2}$$

We can now use Bayes rule to write, with obvious notations:

$$\begin{aligned} P(k_\mathrm{i}, k_\mathrm{o}|a \to b) &= P(k_\mathrm{i}, k_\mathrm{o}, a \to b)/P(a \to b) \\ &= P(a \to b|k_\mathrm{i}, k_\mathrm{o})P(k_\mathrm{i}, k_\mathrm{o})/P(a \to b). \end{aligned} \tag{A3.3}$$

In a random network of N nodes, $P(a \to b) = \langle k_\mathrm{i}\rangle/(N-1)$ and $P(a \to b|k_\mathrm{i}, k_\mathrm{o}) = k_\mathrm{i}/(N-1)$, so that the condition which generalizes the Molloy–Reed criterion finally reads

$$\sum_{k_\mathrm{i},k_\mathrm{o}} \frac{k_\mathrm{i} k_\mathrm{o}}{\langle k_\mathrm{i}\rangle} P(k_\mathrm{i}, k_\mathrm{o}) \geq 1 \tag{A3.4}$$

i.e.

$$\langle k_\mathrm{i} k_\mathrm{o}\rangle \geq \langle k_\mathrm{i}\rangle. \tag{A3.5}$$

Let us now consider a directed network from which a fraction p of nodes has been randomly deleted. A node has then in-degree k_i and out-degree k_o if it has in-degree k_i^0 and out-degree k_o^0 in the undamaged network, and if $k_\mathrm{i}^0 - k_\mathrm{i}$ of its in-neighbours have been removed, together with $k_\mathrm{o}^0 - k_\mathrm{o}$ of its out-neighbours. The joint probability distribution of k_i and k_o can thus be written as

$$P_p(k_\mathrm{i}, k_\mathrm{o}) = \sum_{k_\mathrm{i}^0,k_\mathrm{o}^0} P_0(k_\mathrm{i}^0, k_\mathrm{o}^0) \binom{k_\mathrm{i}^0}{k_\mathrm{i}} (1-p)^{k_\mathrm{i}} p^{k_\mathrm{i}^0 - k_\mathrm{i}} \binom{k_\mathrm{o}^0}{k_\mathrm{o}} (1-p)^{k_\mathrm{o}} p^{k_\mathrm{o}^0 - k_\mathrm{o}}, \tag{A3.6}$$

where P_0 is the degree distribution of the undamaged network. According to the condition (A3.5), a giant component is present if and only if

$$\sum_{k_\mathrm{i},k_\mathrm{o}} k_\mathrm{i} k_\mathrm{o} P_p(k_\mathrm{i}, k_\mathrm{o}) \geq \sum_{k_\mathrm{i},k_\mathrm{o}} k_\mathrm{o} P_p(k_\mathrm{i}, k_\mathrm{o}). \tag{A3.7}$$

We use on each side of the inequality the relation

$$\sum_{k=1}^{k^0} k \binom{k^0}{k} x^k y^{k^0-k} = x \frac{\mathrm{d}}{\mathrm{d}x}(x+y)^{k^0} = x k^0 (x+y)^{k^0-1} \tag{A3.8}$$

with $x = 1 - p$ and $y = p$, to obtain

$$(1-p)^2 \langle k_\mathrm{i}^0 k_\mathrm{o}^0\rangle \geq (1-p)\langle k_\mathrm{o}^0\rangle. \tag{A3.9}$$

The critical value for the existence of a giant component is therefore given by

$$p_c = 1 - \frac{\langle k_\mathrm{o}^0\rangle}{\langle k_\mathrm{i}^0 k_\mathrm{o}^0\rangle}. \tag{A3.10}$$

Such a result shows that directed networks are resilient if and only if $\langle k_\mathrm{i}^0 k_\mathrm{o}^0\rangle$ diverges in the infinite-size limit.

https://doi.org/10.1017/CBO9780511791383.017 Published online by Cambridge University Press

The case of uncorrelated in- and out-degree corresponds to $\langle k_i^0 k_o^0 \rangle = \langle k_i^0 \rangle \langle k_o^0 \rangle$, i.e. to a critical value

$$p_c = 1 - \frac{1}{\langle k_o^0 \rangle} \tag{A3.11}$$

strictly lower than 1: contrarily to undirected networks, uncorrelated directed scale-free networks are not particularly resilient to random failures. On the other hand, correlations between in- and out-degrees are often observed and can modify this property. For example, Schwartz *et al*. (2002) study the case of scale-free distributions of in- and out-degrees, given by

$$P_i(k_i) = (1 - B)\delta_{k_i,0} + Bc_i k_i^{-\lambda_i}(1 - \delta_{k_i,0}); \;\; P_o(k_o) = c_o k_o^{-\lambda_o}. \tag{A3.12}$$

where the additional parameter B is needed to have the possibility to ensure $\langle k_i \rangle = \langle k_o \rangle$. Moreover, for each site, k_o is either completely uncorrelated from k_i (with probability $1 - A$) or fully determined by k_i (with probability A) through a deterministic function $k_o = f(k_i)$. The condition of scale-free distributions implies that $f(k_i) \sim k_i^{\lambda_i - 1/\lambda_o - 1}$, and the joint distribution of in- and out-degrees thus reads

$$\begin{aligned} P(k_i, k_o) &= (1 - A)Bc_i c_o k_i^{-\lambda_i} k_o^{-\lambda_o} + ABc_i k_i^{-\lambda_i} \delta_{k_o, f(k_i)} \;\text{ if } k_i \neq 0 \\ &= (1 - B)c_o k_o^{-\lambda_o} \;\text{ if } k_i = 0. \end{aligned} \tag{A3.13}$$

It is then easy to see that, for any finite fraction of fully correlated sites, the average $\langle k_i k_o \rangle$ diverges if and only if

$$(\lambda_i - 2)(\lambda_o - 2) \leq 1, \tag{A3.14}$$

leading to a very resilient network with $p_c \to 1$, while full uncorrelation ($A = 0$) gives $p_c = 1 - 1/\langle k_o \rangle < 1$ even for scale-free distributions.

A3.2 General case

Boguñá and Serrano (2005) have treated the more general case in which a network contains both directed and undirected (bidirectional) links. The degree of a node is then a three-component vector $\mathbf{k} = (k_i, k_o, k_b)$ where k_i, k_o, k_b are respectively the number of incoming, outgoing and bidirectional links. The degree distribution is then denoted $P(\mathbf{k})$, and correlations are encoded in the conditional probabilities $P_a(\mathbf{k}'|\mathbf{k})$ with $a =$ i or $a =$ o or $a =$ b for the probability of reaching a vertex of degree $\mathbf{k}'$ when leaving a vertex of degree $\mathbf{k}$ through respectively an incoming, out-going or bidirectional link. It can then be shown through the generating function formalism that giant in- and out-components appear when $\Lambda_m > 1$, where Λ_m is the largest eigenvalue of the correlation matrix (Boguñá and Serrano, 2005).

$$C^{o}_{\mathbf{k}\mathbf{k}'} = \begin{pmatrix} k'_o P_o(\mathbf{k}'|\mathbf{k}) & k'_b P_o(\mathbf{k}'|\mathbf{k}) \\ k'_o P_b(\mathbf{k}'|\mathbf{k}) & (k'_b - 1) P_b(\mathbf{k}'|\mathbf{k}) \end{pmatrix} \tag{A3.15}$$

(the matrix $C^{i}_{\mathbf{k}\mathbf{k}'}$ obtained by replacing the indices o by i in the expression of $C^{o}_{\mathbf{k}\mathbf{k}'}$ has the same eigenspectrum, so that the condition for the presence of giant in- and out-components are the same).

Note that both cases of purely directed and purely undirected networks can be recovered from the condition $\Lambda_m > 1$. Indeed, for purely undirected networks only k_b is defined,

https://doi.org/10.1017/CBO9780511791383.017 Published online by Cambridge University Press

and the correlation matrix becomes $(k' - 1)P(k'|k)$, giving back the results of Moreno and Vázquez (2003). For purely directed networks, on the contrary, $k_\mathrm{b} = 0$ and $C^\mathrm{o}_{\mathbf{k}\mathbf{k}'} = k'_\mathrm{o} P_\mathrm{o}(\mathbf{k}'|\mathbf{k})$; if the degrees of neighbouring nodes are uncorrelated, moreover, $P_\mathrm{o}(\mathbf{k}'|\mathbf{k}) = k'_\mathrm{i} P(\mathbf{k}')/\langle k_\mathrm{i}\rangle$ and one recovers the condition $\langle k_\mathrm{i} k_\mathrm{o}\rangle > \langle k_\mathrm{i}\rangle$ for the existence of a giant component.