Chapter 5

Processes on complex networks. Percolation

Up till now we discussed the *structure* of the complex networks. The actual reason to study this structure is to understand how this structure influences the behavior of random *processes* on networks. I will talk about two such processes. The first one is the *percolation* process. The second one is the *spread of epidemics*. There are a lot of open problems in this area, the main of which can be innocently formulated as: How the network topology influences the dynamics of random processes on this network. We are still quite far from a definite answer to this question.

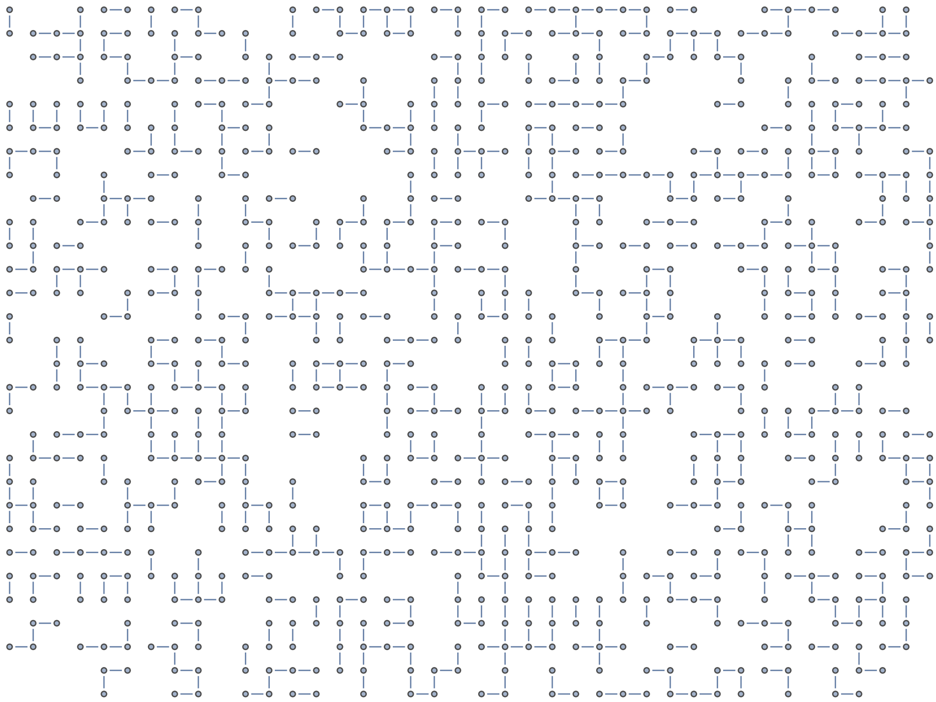
5.1 Percolation

5.1.1 Introduction to percolation

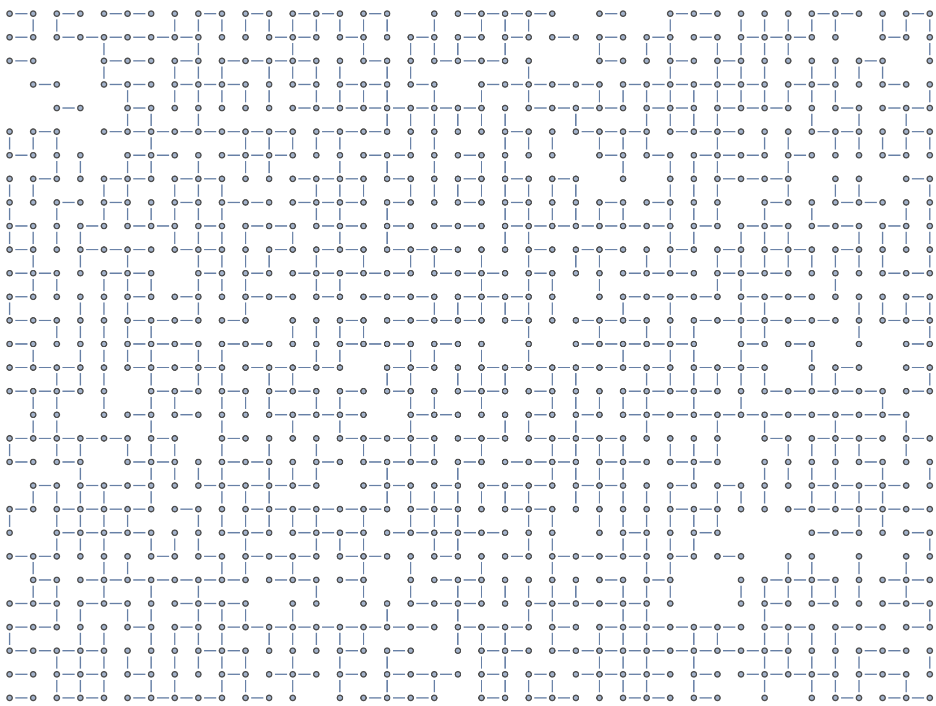
Percolation is one of the simplest processes that exhibit the *critical phenomena* or *phase transition*. This means that there is a parameter in the system, whose small change yields a large change in the system behavior.

To define the percolation process, consider a graph, that has a large connected component. In the classical settings, percolation was actually studied on infinite graphs, whose vertices constitute the set $\mathbf{Z}^d$, and edges connect each vertex with nearest neighbors, but we consider general random graphs. We have parameter ϕ, which is the probability that any edge present in the underlying graph is *open* or *closed* (an event with probability $1-\phi$) independently of the other edges. Actually, if we talk about edges being open or closed, this means that we discuss *bond percolation*. It is also possible to talk about the vertices being open or closed, and this is called *site percolation*. Obviously, if $\phi = 1$ then we get exactly the original graph, and if $\phi = 0$ then no edges are open. If parameter ϕ is increased from 0 to 1 there will be a moment when a giant open component appears; in the language of percolation theory this component is called *giant cluster* or *spanning cluster*. The value at which the giant cluster appear is called the *percolation threshold*, and of the main goals of the theory is to determine this value depending on the structure of the underlying graph. It is remarkable how difficult this problem can be even in simply looking cases.

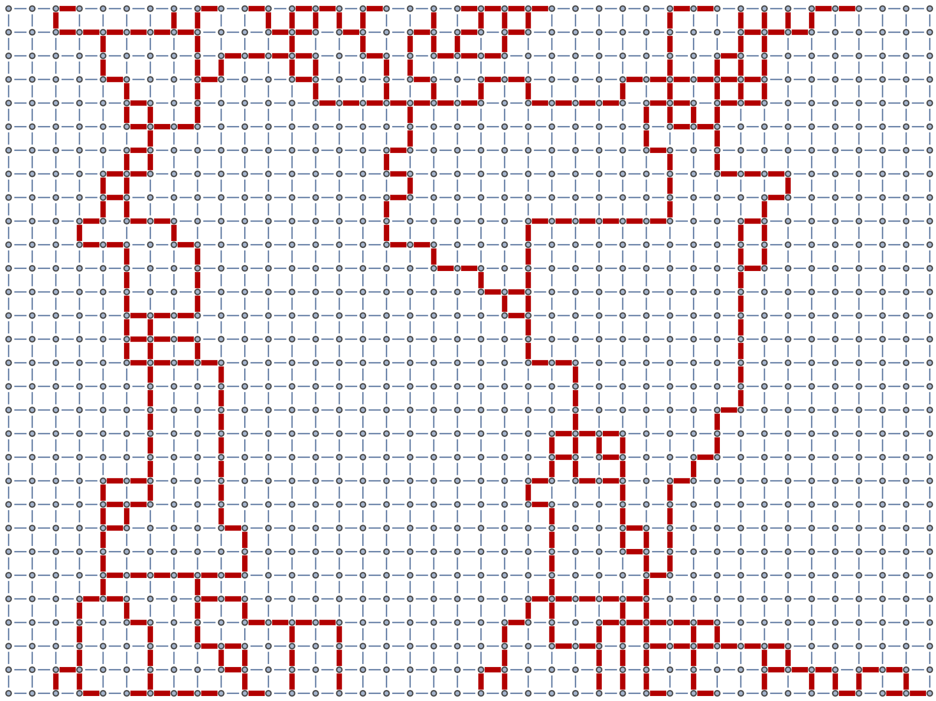
As an example consider a piece of the square lattice $\mathbf{Z}^2$ of the size 30×40. If $\phi = 0.3$ we can observe, e.g., the following picture:



Even perfunctory inspection is enough to confirm that there is no giant component in this graph. However, if we change the value of the parameter to $\phi = 0.55$ the picture changes dramatically:



In this case we definitely have the giant component. The presence of the paths connecting different sides of the lattice explains the name *spanning cluster*. Here you can see some of these paths from the previous figure:



This numerical experiment confirms that the percolation threshold is between 0.3 and 0.55. Indeed, for the bond percolation on the square lattice $\mathbf{Z}^2$ we know that $\phi_c = 1/2$. The proof of this fact is non-elementary[1].

5.1.2 Bond percolation on d-ary trees

Consider a d-ary tree, i.e., a tree, where each node has exactly d offspring. Hence all the nodes, except for the root, have degree $d+1$. Now assume that we are given a percolation probability ϕ that an edge is open. Denote $\theta(\phi)$ the probability that there is an infinite open cluster, containing the root, in our graph. Then the percolation threshold is defined to be $\phi_c = \inf\{\theta(\phi) > 0\} = \sup\{\theta(\phi) = 0\}$. By comparing the percolation process with the Galton–Watson branching process, we see that we are looking exactly for the condition on ϕ when we have supercritical branching process with the binomial probability that any individual has k offspring, with parameters d and ϕ.

Recall that the generating function of the binomial random variable is given by $\varphi(s) = (1 - \phi + \phi s)^d$. Let $d = 2$ for simplicity, then probability for extinction of the corresponding branching process is given by $\pi = 1$ if $\phi \leq 1/2$ and $\pi = ((1 - \phi)/\phi)^2$ if $\phi > 1/2$. Since we have $\theta(\phi) = 1 - \phi$, we obtain the final result

$$\theta(\phi) = \begin{cases} 0, & \phi \leq 0.5\,, \\ 1 - \left(\frac{1-\phi}{\phi}\right)^2, & \phi > 0.5\,. \end{cases}$$

And therefore $\phi_c = 0.5$.

Problem 5.1. If $\phi \leq \frac{1}{2}$ we have that the open cluster size is finite. Find its average size and distribution for the case of binary tree.

[1]see Jeffrey E. Steif, *A mini course on percolation theory* for an introduction to the classical percolation theory

5.1.3 Site percolation on a configuration model

Recall that the configuration model $\mathcal{G}(n, \boldsymbol{d})$ is a probability space where each graph with exactly degree distribution $\boldsymbol{d} = (d_1, \ldots, d_n)$ has equal probability. With configuration model two probability distributions are associated: the degree distribution p_k and the excess degree distribution

$$q_k = \frac{(k+1)p_{k+1}}{\mathsf{E}D},$$

where $\mathsf{E}D$ is the expected degree distribution:

$$\mathsf{E}D = \sum_{k=1}^{\infty} kp_k = \frac{\sum_{i=1}^{n} d_i}{n}.$$

We denote $\varphi_0(s)$ and $\varphi_1(s)$ the probability generating functions for p_k and q_k respectively. Recall that the condition for the existence of the giant component in the configuration model takes the form

$$\nu = \frac{\mathsf{E}D^2 - \mathsf{E}D}{\mathsf{E}D} > 1.$$

Note that for the power law distribution with the exponent $2 \leq \gamma \leq 3$ the second moment is infinite, and hence the configuration model always has the giant component. Now we are interested in site percolation on the configuration model. I.e., each vertex in the configuration model with $\nu > 1$ can be "open" with probability ϕ and "closed" with probability $1 - \phi$. We are looking for the condition on ϕ such that a *giant cluster* appears in our graph. And we also interested in the size of the giant cluster.

Two situating are useful to have in mind while discussing what exactly we model by this site percolation. First, you can think of the Internet graph, where vertices are the routers, which are working with probability ϕ and are out of order with probability $1 - \phi$. We are interested in the proportion of routers that has to be kept in a working condition such that the Internet is still connected (i.e., the giant cluster is present). The second situation is the vaccination process. Vaccination not only defends the person who was vaccinated from the disease. It also protects others by preventing the disease spread through this particular person. This actually leads to the concept of *herd immunity*, i.e., to the proportion of people that has to be vaccinated to guarantee that the disease will not spread. The vertex is open, or not vaccinated with probability ϕ, and hence the presence of the giant cluster means that there is a way for the disease to spread throughout the population. Note that for the example with the Internet we would like to construct the network such that the percolation threshold is low (ϕ_c is small) whereas for the vaccination, we would prefer to increase the percolation threshold as much as possible, to reduce the number of vaccinated people necessary for the herd immunity of the population.

Now consider heuristic arguments that lead to the condition for the percolation threshold.

Let u be the probability that a randomly chosen vertex is not in the giant cluster. Any vertex is not in the giant cluster if and only if it is not connected to the giant cluster through any of its neighbors, if this vertex has degree k than this probability is given by u^k. We find that a randomly chosen vertex belongs to the giant cluster with probability

$$1 - \varphi_0(u) = 1 - \sum_k p_k u^k,$$

but this only true for any *open* vertex, i.e., the one that was kept in the percolation process, hence the total fraction of vertices that belong to the giant cluster if given by

$$S = \phi(1 - \varphi_0(u)).$$

Now we calculate u. For a vertex not to belong to the giant cluster, its neighbor has be either closed (an event of probability $1 - \phi$), or it is open (probability ϕ but does not belong to the giant cluster itself). Hence,

$$u = \sum_k q_k(1 - \phi + \phi u^k) = 1 - \phi \sum_k q_k u^k = 1 - \phi + \phi\varphi_1(u).$$

Putting everything together, we obtain

Proposition 5.1. *In the configuration model $\varphi(n, \boldsymbol{d})$, the size of the giant cluster can be found as*

$$S = \phi(1 - \varphi_0(u)),$$

where u is the smallest solution to the equation

$$u = 1 - \phi + \phi\varphi_1(u).$$

Strictly speaking this results holds for the two staged branching process, where the first generation of the descendants has the distribution p_k, and all subsequent generations has distribution q_k. Note that we always have solution $u = 1$, which corresponds to the absence of the giant cluster. It is possible to have another smaller solution to this problem if and only if the slope of the function $1 - \phi + \phi\varphi_1(s)$ at the point $s = 1$ is strictly bigger than one, here the reasoning almost literally repeats the discussion in the branching process section. Therefore, we find that

$$\phi\varphi_1'(1) > 1 \implies \phi > \frac{1}{\nu},$$

since $\nu = \varphi_1'(1)$. Therefore, the critical percolation probability is given by

$$\phi_c = \frac{1}{\nu},$$

and of course this expression makes sense only if $\nu > 1$ as expected.

Problem 5.2. Show that for a Poisson random graph the percolation threshold is given by

$$\phi_c = \frac{1}{\lambda}.$$

Now we are ready to state one of the most famous results in network science: "The Internet is robust." Consider the degree distribution for the configuration model to obey a power law with the exponent $2 < \gamma \leq 3$. In this case, since $\mathsf{E}D^2 = \infty$, then the percolation threshold is simply

$$\phi_c = 0,$$

which implies that independently of with which probability we keep the vertices in the model, there is always a giant cluster in the model. This conclusion was used as a proof that the Internet will say connected even if most of the routers will stop working. On the other hand, if we consider the vaccination of people, it implies that if the contact network of the population has the power law degree distribution with the exponent $2 < \gamma \leq 3$, then it is impossible to get the herd immunity phenomenon in this population.

Let us critically access these far reaching conclusions.

- This conclusion is based on a branching process approximation to the percolation process on a configuration model. The Internet, while probably having power law degree distribution, is not a configuration model, it has some additional structure.
- The size of the giant cluster is somewhat missing in this discussion.

Example 5.2. Consider a network with degree distribution given by

$$p_k = (1 - e^{-\lambda})e^{-\lambda k}.$$

This is so called discrete exponential degree distribution. We find that

$$\varphi_0(s) = \frac{e^\lambda - 1}{e^\lambda - s}, \quad \varphi_1(s) = \left(\frac{e^\lambda - 1}{e^\lambda - s}\right)^2.$$

Hence we find for u:

$$u(e^\lambda - u)^2 - (1-\phi)(e^\lambda - u)^2 - \phi(e^\lambda - 1)^2 = 0,$$

which is a cubic equation. We know that $u = 1$ is always a root, therefore, we can factorize it as

$$(u-1)(u^2 + (\phi - 2e^\lambda)u + \phi - 2\phi e^\lambda + e^{2\lambda}) = 0,$$

we are looking for the smallest root, which is

$$u = e^\lambda - \frac{\phi}{2} - \sqrt{\frac{\phi^2}{4} + \phi(e^\lambda - 1)}\,.$$

It means that

$$S = \frac{3}{2}\phi - \sqrt{\frac{1}{4}\phi^2 + \phi(e^\lambda - 1)},$$

and the percolation threshold here is

$$\phi_c = \frac{1}{2}(e^\lambda - 1).$$

Now we can plot the size of the giant cluster as the function of the percolation probability (make figure).

For the power law distribution $p_k \sim ck^{-\gamma}$ it can be shown that the size of the giant cluster is asymptotically given by

$$S = c\phi^{\frac{\gamma-2}{3-\gamma}}$$

for $2 < \gamma < 3$, and

$$S = e^{\frac{-(1+o(1))}{c\phi}}.$$

as $\phi \to 0$, i.e., it is vanishingly small. So assume that $\phi = 0.05$, i.e., 95% of the cites were destroyed and assume also that $p_k \sim 0.832k^{-3}$. Our main result says that we still have the giant cluster, but the calculation would tell us that $S \sim 10^{-11}$, so there is no point in talking about actual Internet to be connected.

5.1.4 Intentional damage of sites

Up till now all the vertices were "open" or "closed" independently with the same probability ϕ. Now consider a generalization, in which ϕ_k denotes the probability that vertex with degree k is *open*. For example, we can have that $\phi_k = 1$ for $k < k_0$ and $\phi_k = 1$ for $k \geq k_0$, in this case we remove all vertices of degree k_0 and larger from the network. You should now that we can choose a strategy for damaging a network.

Let u as before be the probability that a vertex is not connected to the giant cluster through its neighbors. If there are k neighbors, then this probability is u^k, and hence $1 - u^k$ is the probability that this vertex is connected to the giant cluster. To belong to the giant cluster this vertex has to be open, an even of probability ϕ_k, hence the probability of being a member of the giant cluster is $\phi_k(1-u^k)$. We find then that the size (the proportion of nodes in the giant cluster) is

$$S = \sum_{k=0}^{\infty} p_k\phi_k(1-u^k) = \sum_{k=0}^{\infty} p_k\phi_k - \sum_{k=0}^{\infty} p_k\phi_k u^k = \psi_0(1) - \psi_0(u),$$

where

$$\psi_0(s) = \sum_{k=0}^{\infty} p_k \phi_k s^k.$$

Note that $\psi_0(1)$ gives the expectation of the number of open vertices.

Again, very similarly to the previous section, we have that the probability that a vertex is not connected to the giant cluster through its neighbors can be calculated as follows: Either a neighbors is closed (probability $1-\phi_{k+1}$) or open, but then it does not belong to the giant cluster ($\phi_{k+1}u^k$). Adding everything together

$$u = \sum_{k=0}^{\infty} q_k(1-\phi_{k+1}+\phi_k u^k) = 1-\psi_1(1)+\psi_1(u),$$

where

$$\psi_1(s) := \sum_{k=0}^{\infty} q_k \phi_k s^k.$$

Note that

$$\psi_1(s) = \frac{\psi_0'(s)}{\varphi_0(1)}\,.$$

We obtain that

Proposition 5.3. *If the percolation probability given by* $(\phi_k)_{k=0}^{\infty}$*, then the size of the giant cluster can be found as*

$$S = \psi_0(1) - \psi_0(u),$$

where u *is the smallest root of*

$$u = 1-\psi_1(1)+\psi_1(u).$$

Example 5.4. Consider again the example with

$$p_k = (1-e^{-\lambda})e^{-\lambda k}.$$

and assume that

$$\phi_k = \begin{cases} 1, & k < k_0, \\ 0, & k \geq k_0. \end{cases}$$

Then

$$\psi_0(s) = (1-e^{-\lambda})\sum_{k=0}^{k_0-1} e^{-\lambda k}s^k = (1-e^{-\lambda k_0}s^{k_0})\frac{e^{\lambda}-1}{e^{\lambda}-s}\,,$$

and

$$\psi_1(s) = \Big((1-e^{-\lambda k_0}s^{k_0}) - k_0 e^{-\lambda(k_0-1)}s^{k_0-1}(1-e^{-\lambda}s)\Big)\left(\frac{e^{\lambda}-1}{e^{\lambda}-s}\right)^2.$$

Using these equations we can numerically find the size of the giant cluster.

Problem 5.3. Use the last example to show numerically that it is enough to remove approximately 8% of vertices with highest degree to destroy the giant cluster. This fact is in large contrast with the random removal of vertices, in which almost 70% of vertices has to be made closed, for the giant cluster to disappear.

For the case of the power law distribution, numerical computations show that it is enough to remove a very tiny fraction of vertices with the highest degree to destroy the giant cluster (the estimates are 3% for γ close to 2, and less than 1% for γ close to 3). This is the explanation of the “fragile” part of the Internet.

Chapter 6

Epidemics on networks