# Link metadata percolation - Final report

Alessandro Canevaro
supervisor: Oriol Artime
Fondazione Bruno Kessler - 26/08/2022

## 1 Introduction

The focus of the internship is on the analysis of complex systems, namely percolation processes. During the first part of the internship, the current state-of-the-art results in this topic were studied. For example, a series of experiments from the Newman [New10] book and from recently published papers were reproduced. In the second part, the focus was shifted toward the research of feature-enriched bond percolation processes. All the analytical results are also compared with simulated experiments conducted on the developed network simulator.

## 2 Algorithms

To test the analytical solutions, a network percolation simulator was developed. To fastly simulate large-scale networks with 100K nodes, two strategies were used. Firstly the core percolation algorithms were coded in C++. Secondly, the low time-complexity percolation algorithm proposed by Newman [NZ01] was used. For analysing the results and transforming the raw data into meaningful plots, a Python script was developed. The parameters of each experiment (such as the network size, the degree distributions, the percolation type, etc..) are specified through a YAML configuration file.

## 3 Network generation

All the studied processes happen on networks generated through the well-known and widely-used configurational model [New10]. The main advantage of this model is that it allows to specify a degree distribution $p_k$. In this project, the two most studied degree distributions are the binomial distribution (Erdős–Rényi networks) and the power-law distribution (Scale-free networks). Moreover, another quantity that will recur many times in this project is the excess degree distribution $q_k$, which is the probability that by following an edge a node with degree k is reached. Note that $q_k$ can be expressed in terms of $p_k$:

$$q_k = \frac{(k+1) \cdot p_{k+1}}{\langle k \rangle} \tag{1}$$

Finally, let us introduce the generating functions for the degree and excess degree distributions, that will be useful in the following calculations:

$$g_0(z) = \sum_{k=0}^{\infty} p_k z^k, \quad g_1(z) = \sum_{k=0}^{\infty} q_k z^k \tag{2}$$

# 4 Percolation processes

## 4.1 Uniform random removal of nodes

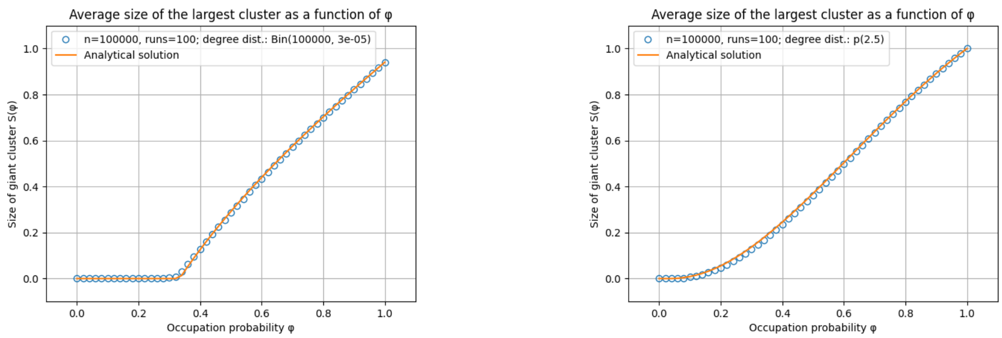
This experiment consists of uniformly removing nodes from the network with a given probability $(1-\phi)$, and analysing the size of the giant component as a function of the occupation probability $\phi$.
The analytical solution is given by:

$$S = \phi(1 - g_0(u)) \tag{3}$$

where $u$ is the solution of the following self-consistent equation:

$$u = 1 - \phi + \phi g_1(u) \tag{4}$$

Figure 1a shows the comparison between the analytical solution and the simulation of a network with binomial degree distribution. Similarly, Figure 1b compares the analytical solution with a simulated scale-free network.




(a) ER network

(b) SF network

Figure 1: Site percolation

## 4.2 Targeted attacks

Another interesting case is to study what happens when the nodes are removed according to a given rule, i.e. all the nodes with degrees greater than a threshold are removed. This is also called a targeted attack. The above-defined occupation probability can be expressed in mathematical terms as:

$$\phi_k = \begin{cases} 0 \text{ if } k \geq k_0 \\ 1 \text{ otherwise} \end{cases} \tag{5}$$

then, by redefining the generating functions of the degree and excess degree distribution as follows:

$$f_0(z) = \sum_{k=0}^{\infty} \phi_k p_k z^k, \quad f_1(z) = \sum_{k=0}^{\infty} \phi_{k+1} q_k z^k \tag{6}$$

the size of the giant component can be expressed as:

$$S = f_0(1) - f_0(u) \tag{7}$$

with $u$ equal to:

$$u = 1 - f_1(1) + f_1(u) \tag{8}$$

Figure 2 shows that the analytical solution once again correctly predicts the results of the simulations.

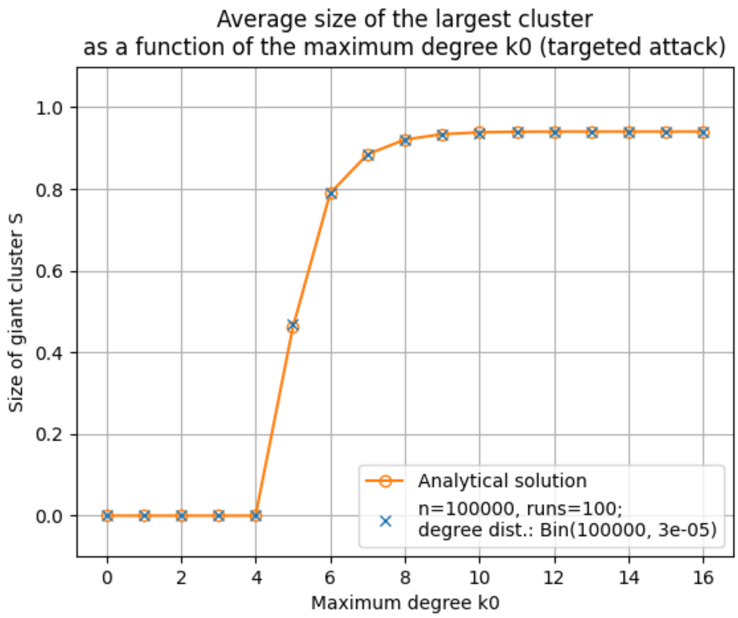



Figure 2: Targeted attack on ER network

### 4.3 Bond percolation

In the case of bond percolation, the links of the networks are removed. Similarly to the site percolation case, the links can be removed uniformly at random or by a targeted attack. In the case of a uniform removal, the analytical solution can be derived with the same equation as for site percolation except for eq. 3 that becomes:

$$S = 1 - g_0(u) \tag{9}$$

namely site and bond percolation differ by the factor $\phi$. However, this is not true in the case of a targeted attack, where the analytical solution for the bond percolation case is much more

complex as shown in the following paragraphs.
Figure 3 shows that the simulations match the analytical solution.

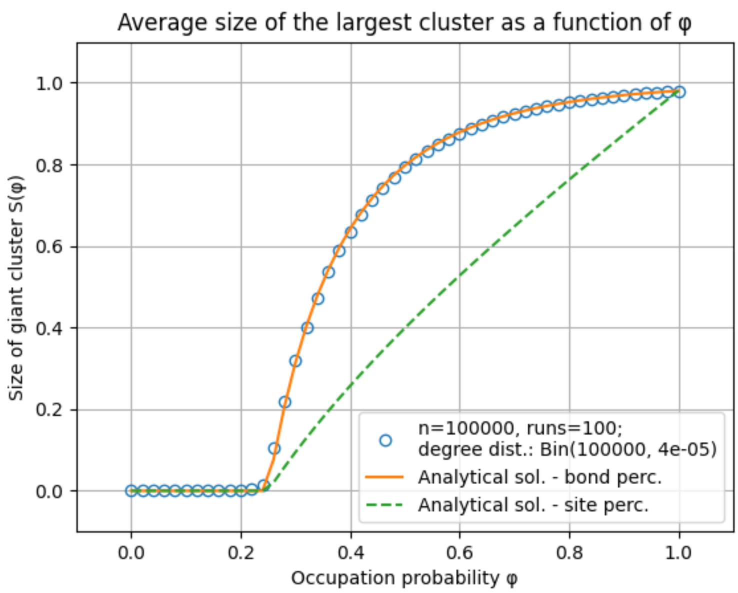



Figure 3: Bond percolation on ER network

# 5 Feature-enriched bond percolation

## 5.1 Independent case

Let us consider a network generated through the configurational model, with degree distribution $p_k$. Consider also a feature value for each network's link, and assume that these values are distributed according to a feature distribution $p_F$. In the following $p_F$ will be assumed, without loss of generality, to be a discrete distribution.
The goal is to derive an analytical solution for the giant cluster size in the case that the edges with the highest feature value are removed first during the percolation process.
In other words, all the edges with feature value greater or equal to the threshold value $F_0$ are occupied with probability 0 (removed from the network), while all the other edges are present in the network.
This is translated mathematically into the following edge occupation probability:

$$\phi_F = \begin{cases} 0 \text{ if } F \geq F_0 \\ 1 \text{ otherwise} \end{cases} \tag{10}$$

Note that the $p_F$ and $\phi_F$ are uncorrelated from every other quantity of the network.

Let $u$ be the average probability that a node is not connected to the giant cluster through a specific edge. $u$ is given by the sum of two contributions, i.e. the probability that an edge is unoccupied, and the probability that it is occupied but the node at the other end of the edge is itself not a member of the giant cluster. Hence, the total probability is: $1-\phi_F+\phi_F u^k$. Averaging over the value of $k$, distributed according to the excess degree distribution $q_k$, and the value of $F$, distributed according to $p_F$, leads to the following self-consistent equation:

$$u=\sum_{F=0}^{\infty}\sum_{k=0}^{\infty} p_F q_k\left(1-\phi_F+\phi_F u^k\right)=1-\left(\sum_{F=0}^{F_0-1} p_F\right)+\left(\sum_{F=0}^{F_0-1} p_F\right) g_1(u) \tag{11}$$

where $g_1(z)$ is the generating function of the excess degree distribution, namely:

$$g_1(z)=\sum_{k=0}^{\infty} q_k z^k=\sum_{k=1}^{\infty}\frac{p_k\cdot k}{\langle k\rangle}\cdot z^{k-1} \tag{12}$$

The probability that a node does not belong to the giant cluster is $u^k$. Thus, the average over the degree distribution $p_k$ is equal to $1-S$, and the sought result is given:

$$S=1-\sum_{k=0}^{\infty} p_k u^k=1-g_0(u) \tag{13}$$

where $g_0(z)$ is the generating function of the degree distribution.

## 5.2 Simulations

### 5.2.1 Erdős–Rényi network

Let us consider a network generated via the configurational model with binomial degree distribution, i.e. $p_k\sim Bin(n,p)$ with $\langle k\rangle=3$. Moreover, let $p_F$ be Poisson distributed with $\mu=8$. Figure 4 compare the analytical solution with the simulation results. The simulation is performed on a network with 100K nodes and averaged over 100 runs.

### 5.2.2 Scale-Free network

Similarly to the Erdős–Rényi case, let us consider a network generated via the configurational model but with power-law degree distribution with $\alpha=3$. Note that the degree is constrained in the range $(2,\sqrt{n})$. Moreover, let $p_F$ be Poisson distributed with $\mu=8$. Figure 5 compare the analytical solution with the simulation results. The simulation is performed on a network with 100K nodes and averaged over 100 runs.

## 5.3 Correlated case

The main idea is that now the feature value of a link is correlated to the degree of the two adjacent nodes. Therefore let us define $P(f,k,k')$ as the probability that a link has feature $f$ and it is connected to a node with degree $k$ and a node with degree $k'$. Note that this joint probability can be rewritten as $P(f,k,k')=P(f|k,k')P(k,k')$. Moreover, note that in the case

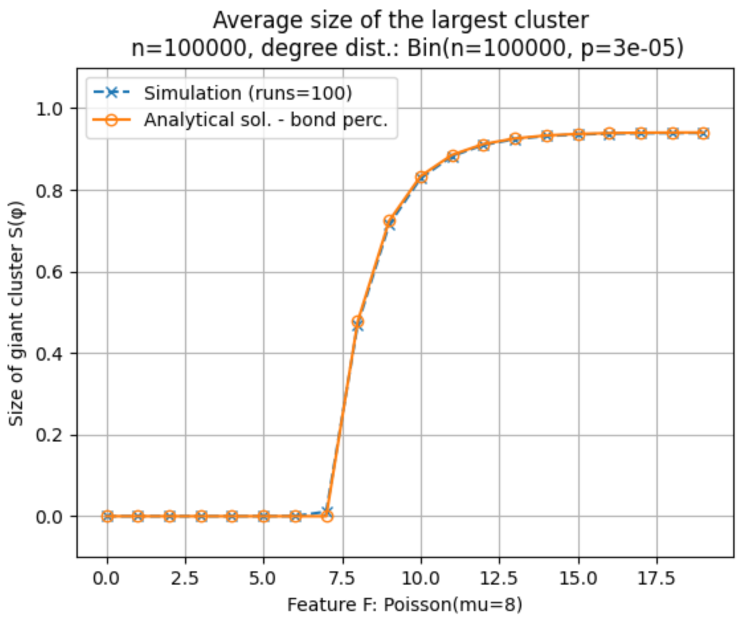



Figure 4: ER network

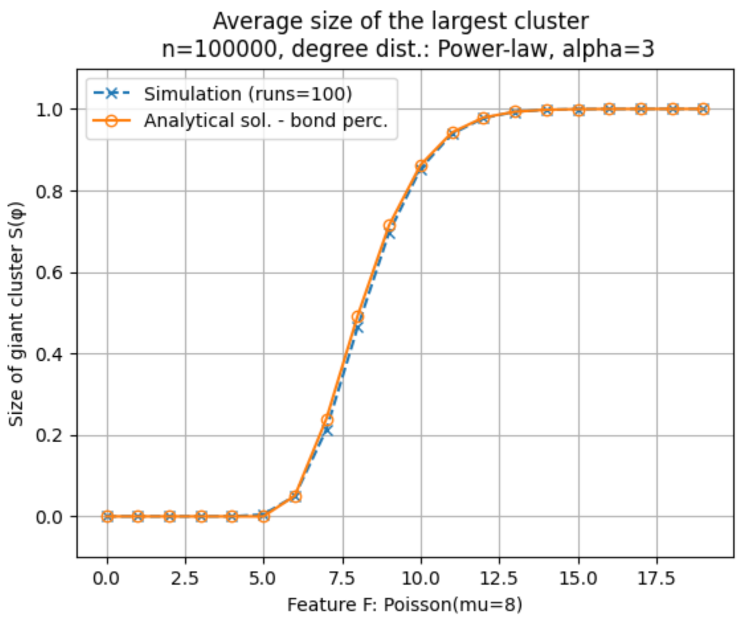



Figure 5: SF network

of degree-uncorrelated networks $P(k,k') = P(k)P(k')$. Here $P(k) = \frac{kp_k}{\langle k \rangle}$ is the probability that a randomly chosen link is connected to a node with degree $k$.

Let us introduce the results proposed by [SK10] based on the cavity method that gives an analytical solution for percolation processes where both a site and a node occupation probability are specified. The main equations are:

$$u_k = \left( \sum_m P(m|k)(1-s_m)(1-b_{mk})(1-u_m) \right)^{k-1} \quad (14)$$

$$w_k = \left( \sum_m P(m|k)(1-s_m)(1-b_{mk})(1-u_m) \right)^{k} \quad (15)$$

$$S = \sum_k p_k(1-s_k)(1-w_k) \quad (16)$$

Here $s_m$ and $b_{mk}$ are the removal probabilities for nodes of degree $k$ and links connected to nodes with degree $k$ and $m$ respectively. $P(m|k)$ as discussed before is equal to $P(m) = \frac{mp_m}{\langle k \rangle}$. For our case $s_m = 0$ as we are dealing only with link removal. $1 - b_{mk} = \phi_f$ with $\phi_f$ defined as in equation 10. Since now equations 14 and 15 depend on $f$, we need to average over the values of the features. The modified equations become:

$$u_k = \left( \sum_f \sum_m P(f|m,k) \cdot \frac{mp_m}{\langle k \rangle} \cdot \phi_f \cdot (1-u_m) \right)^{k-1} \quad (17)$$

$$w_k = \left( \sum_f \sum_m P(f|m,k) \cdot \frac{mp_m}{\langle k \rangle} \cdot \phi_f \cdot (1-u_m) \right)^{k} \quad (18)$$

$$S = \sum_k p_k(1-w_k) \quad (19)$$

## 5.4 Simulations

Let us define $P(f|k,k')$ as a Poisson distribution where the $\mu$ parameter depends on $k$ and $k'$. Namely let:

$$P(f|k,k') \sim P(k+k') \quad (20)$$

this is an example of positively correlated features. Figure 6 reports the results of a simulated ER networks and the analytical solution derived from eqs. 17-19.
Conversely, $p(f|k,k')$ can be redefined as:

$$P(f|k,k') \sim P(50//(k+k')) \quad (21)$$

in order to reproduce negatively correlated features. Figure 7 shows the results of the negatively correlated feature-degree distribution.

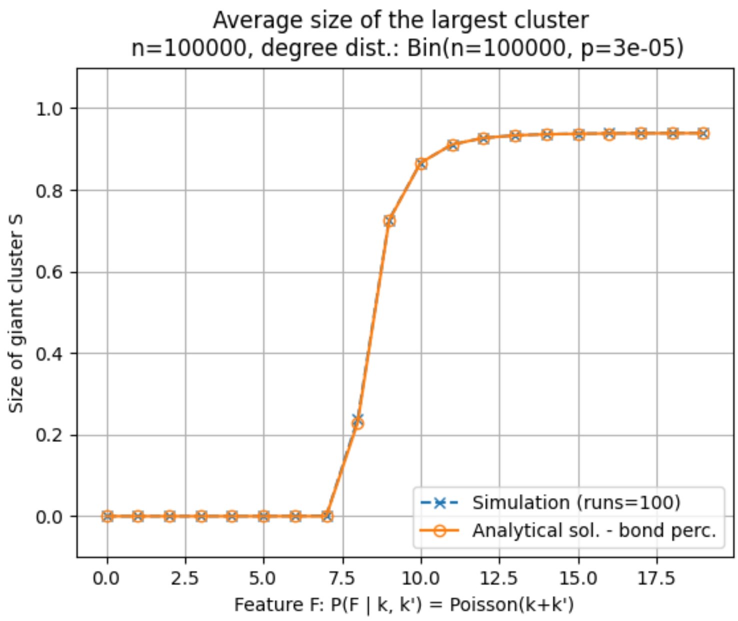



Figure 6: Positively-correlated feature-enriched targeted attack on ER network

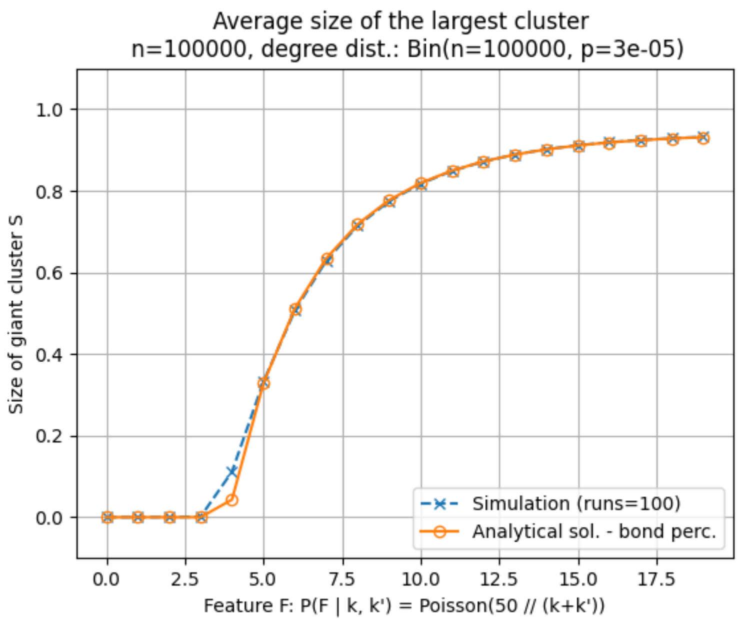



Figure 7: Negatively-correlated feature-enriched targeted attack on ER network

## References


[NZ01] M. E. J. Newman and R. M. Ziff. "Fast Monte Carlo algorithm for site or bond percolation". In: *Physical Review E* 64.1 (June 2001). DOI: `10.1103/physreve.64.016706`. URL: `https://doi.org/10.1103%2Fphysreve.64.016706`.

[New10] M. E. J. Newman. *Networks: an introduction*. Oxford; New York: Oxford University Press, 2010. ISBN: 9780199206650 0199206651. URL: http://www.amazon.com/Networks-An-Introduction-Mark-Newman/dp/0199206651/ref=sr_1_5?ie=UTF8&qid=1352896678&sr=8-5&keywords=complex+networks.

[SK10] Yoshifumi Shiraki and Yoshiyuki Kabashima. "Cavity analysis on the robustness of random networks against targeted attacks: Influences of degree-degree correlations". In: *Physical Review E* 82.3 (Sept. 2010). DOI: 10.1103/physreve.82.036101. URL: https://doi.org/10.1103%2Fphysreve.82.036101.