Probability for Data Science – Master's Degree in Data Science – A.Y. 2023/2024
Academic Staff: Francesca Collet, Paolo Dai Pra

PROJECT 4: PERCOLATION ON BOOLEAN NETWORKS

Abstract. Percolation theory describes how the size of clusters of connected set of edges of large random networks varies as the connectivity - i.e. the probability that two vertices are connected by an edge - changes. As the connectivity increases the system undergoes a transition from a situation where the clusters are small in size to a picture with significantly larger clusters. This transition is not smooth, but rather it happens sharply as the connectivity crosses a critical value, called percolation threshold*. The scope of the present project is to study numerically this phenomenon in Boolean networks.

Boolean model network on $\mathbb{R}^2$

To define Boolean model networks we need to introduce first Poisson processes on the plane. A random set of points $X \subset \mathbb{R}^2$ is said to be a **Poisson process of density** $\lambda > 0$ on the plane if it satisfies the following conditions:

- Let $X(D)$ denote the random number of points of X inside a domain D. For mutually disjoint sets $D_1, D_2, \ldots, D_k \subset \mathbb{R}^2$, the random variables $X(D_1), X(D_2), \ldots, X(D_k)$ are independent.
- Let $|D|$ denote the area of a domain D. For any bounded $D \subset \mathbb{R}^2$ and any $k \geq 0$, it holds

$$P\left(X(D) = k\right) = e^{-\lambda|D|} \frac{\left(\lambda|D|\right)^k}{k!},$$

  that is $X(D) \sim \text{Poi}(\lambda|D|)$. In particular, as a consequence, the mean number of points of X in D is $E[X(D)] = \lambda|D|$.

Consider X a Poisson process of density $\lambda > 0$ on the plane and take $r > 0$. A **Boolean random network** on the plane, denoted by (X, λ, r), is constructed as follows: given a realization of the Poisson process X, two elements $x, y \in X$ are connected if their Euclidean distance is smaller than or equal to $2r$. This geometrically corresponds to placing discs of radius r at the points of the Poisson process and considering connected components formed by clusters of overlapping discs. See the figure below, taken from [1].

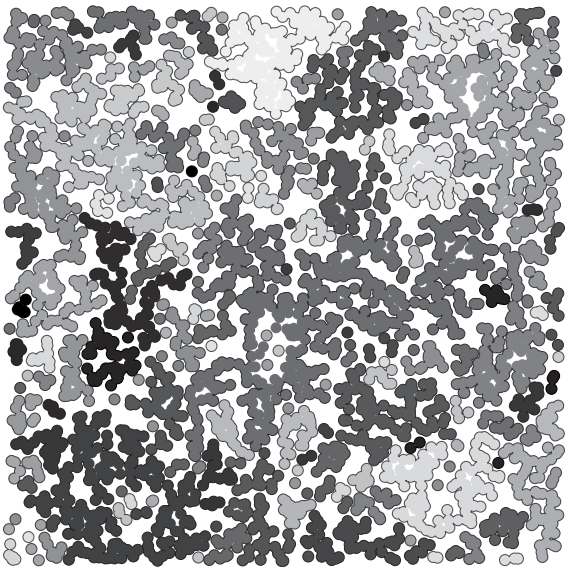

Figure 1: Boolean model. Connected components of overlapping discs are drawn with the same grey level.

*More about percolation theory and its applications may be found in [1] or at `https://en.wikipedia.org/wiki/Percolation_theory`.

mostrare la differenza tra algorith e bool, quindi creare N all interno di creation_boolen però non rispetta la formulazione nel progetto che è quella di passare N a creation_boolean (quindi avere N sempre uguale per ogni M oppure ad ogni M creare un nuovo N?? in ogni caso più è grande M più N ai avvicina alla media generale quindi non devrebbe cambiare nulla??

Algorithm to simulate a Poisson process on a square

We describe a simple algorithm to simulate a Poisson process X of density $\lambda > 0$ on the squared domain $\Lambda = [0, T] \times [0, T]$, with $T > 0$. The algorithm is based on two key facts:

- the number of points in Λ has distribution $\text{Poi}(\lambda T^2)$;
- conditional on $X(\Lambda)$ (number of points in Λ), locations of points $x \in X$ are independent and uniformly distributed on the domain Λ (see [3, p. 359]).

Input: size of the squared domain T;
density of the Poisson process λ;

Output: Euclidean coordinates of N points of the Poisson process;

Procedure

Step 1. Generate $N \sim \text{Poi}(\lambda T^2)$.

Step 2. If $N = 0$, then stop; there are no points in Λ. Otherwise, generate $2N$ independent random numbers in $(0, 1)$: $U_1, U_2, \ldots, U_N, V_1, V_2, \ldots, V_N$.

Step 3. The Euclidean coordinates of the N points of the Poisson process are $(U_i T, V_i T)$, for $i = 1, \ldots, N$.

Remark. Step 2 above relies on the fact that we are dealing with a squared domain. In general, different domain shapes require different algorithms. For instance, in [2, p. 279], an algorithm to generate a homogeneous Poisson process on a circle is given.

Project

Given a random network, a cluster is a set of connected points and the cluster size is defined as the number of nodes belonging to the cluster itself. Natural and relevant questions to address concern the number of clusters present in the network and the size of the largest cluster.
The current project consists in what follows. Set $r = 1$ and take a squared domain $\Lambda = [0, T] \times [0, T]$, with $T > 0$ large. Fix a value of λ and make M independent simulations of the Boolean random network $(X, \lambda, 1)$. Then, compute the empirical average of the M sizes of the largest cluster.
Clusters may be identified by using algorithms such as Depth First Search (DFS) or Breadth Search First (BFS) available on the Internet.
By running several simulations and collecting the results in appropriate plots, investigate the following problems:

- How the size of the largest cluster depends on λ. Let $\lambda_c \simeq \frac{4.512}{4\pi}$. You should be able to identify two regimes: one for $\lambda < \lambda_c$ where clusters are small and a different regime for $\lambda > \lambda_c$ where the largest cluster cover (almost all of) the network.
- (Optional) How the number of clusters depends on λ.

References

[1] Franceschetti M. and Meester R.W.J., *Random networks for communication*, Cambridge University Press, 2007

[2] Ross, S.M., *Simulation*, Academic Press, 2006

[3] Resnick, S.I., *Adventures in stochastic processes*, Birkhäuser Boston, 1992