Exercise 1. Cluster size distribution

Goal: In this exercise we study the cluster size distribution of a system (e.g. forest fire of last week's exercise sheet) with the help of the Hoshen-Kopelman algorithm.

A very efficient method to determine the cluster size distribution n_s is the Hoshen-Kopelman algorithm (J. Hoshen and R. Kopelman, Phys. Rev. B **14**, 3428, 1976) which determines which sites belong to each cluster. One advantage of the method is that one has to go through the system only once. (When the algorithm was developed saving memory and computation time was a crucial requirement to get results in reasonable computing time. Nowadays such intelligent algorithms are still very helpful e.g. to investigate very large systems.)

Task 1: Implement the Hoshen-Kopelman algorithm for a square lattice of size $L \times L$ and occupation probablilty p.

The Hoshen-Kopelman algorithm is described in the following:

- Starting point is a 2D lattice consisting of 0's (unoccupied) and 1's (occupied). Each cluster gets a number k (starting from $k = 2$ to distinguish from 'untouched' sites) and the size (mass) $M(k)$ of each cluster is stored.
- Here, we will sweep through the system from left to right and from up to down (starting point: top left corner).
- For every occupied site there are three possible situations for the top and left neighboring sites:
  - Both are unoccupied. In this case create a new cluster $k \to k+1$ and $M(k) = 1$. Set the value of the site to k.
  - One is occupied or both are occupied with the same number k_0. The current site belongs to the cluster. Set the value of the current site to k_0 and set $M(k_0) \to M(k_0) + 1$.
  - Both are occupied with two different numbers k_1 and k_2 (WLOG assume $k_1 < k_2$). Both clusters have to be united. Set the value of the current site to k_1 and set $M(k_1) \to M(k_1) + M(k_2) + 1$ while $M(k_2) \to 0$. Set all sites with value k_2 to k_1.

Task 2: Study the cluster size distribution for different occupation probabilities. What do you observe for $p < p_c$ compared to $p > p_c$?

Hint: The cluster size distribution n_s is normalized per site i.e. the total number of clusters of each size is divided by the total number of sites. (For this reason n_s is always below 1.)