Fast Low Degree Connectivity of Ad-Hoc Networks via Percolation

Fabrizio Grandoni

Joint work with E. De Santis and A. Panconesi

Università di Roma "La Sapienza"

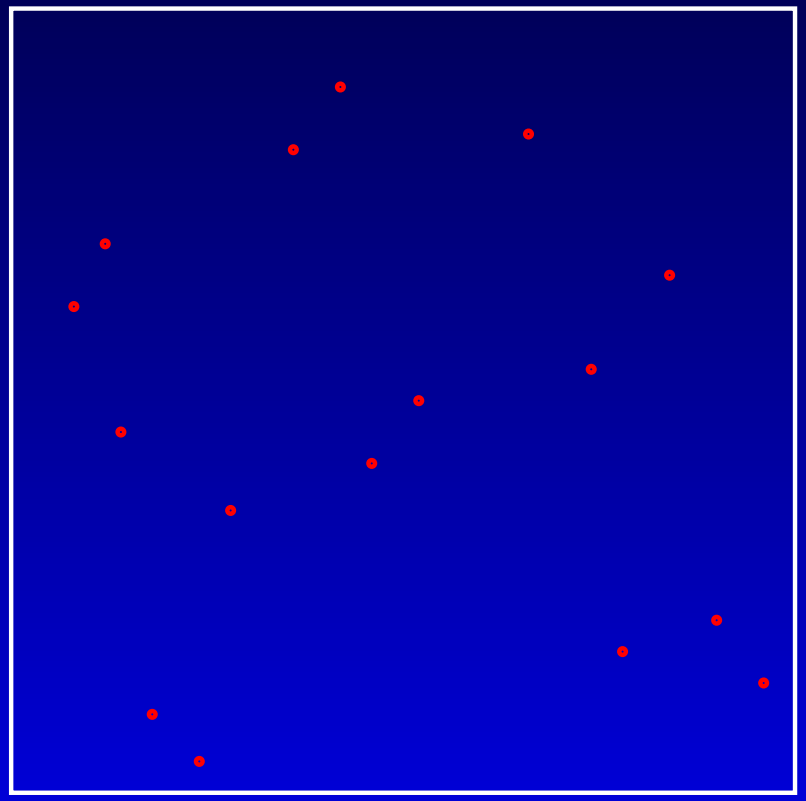
grandoni@di.uniroma1.it

The problem

- n points (devices) are distributed uniformly at random in the unit square.
- Each point v can set its own transmission radius in order to reach the k_v closest devices
  - $O(1)$ rounds to find the next closest device
- There is a communication channel between two devices iff they are within the transmission radius of each other.
  - ack's in wireless standards.

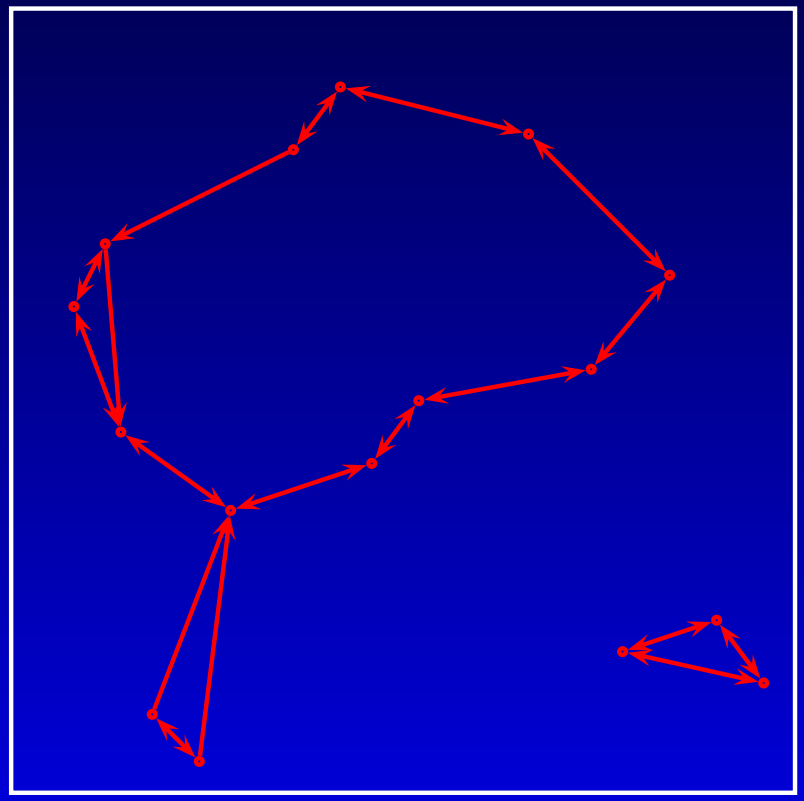
The problem

If $k_v = 2$ for each v:



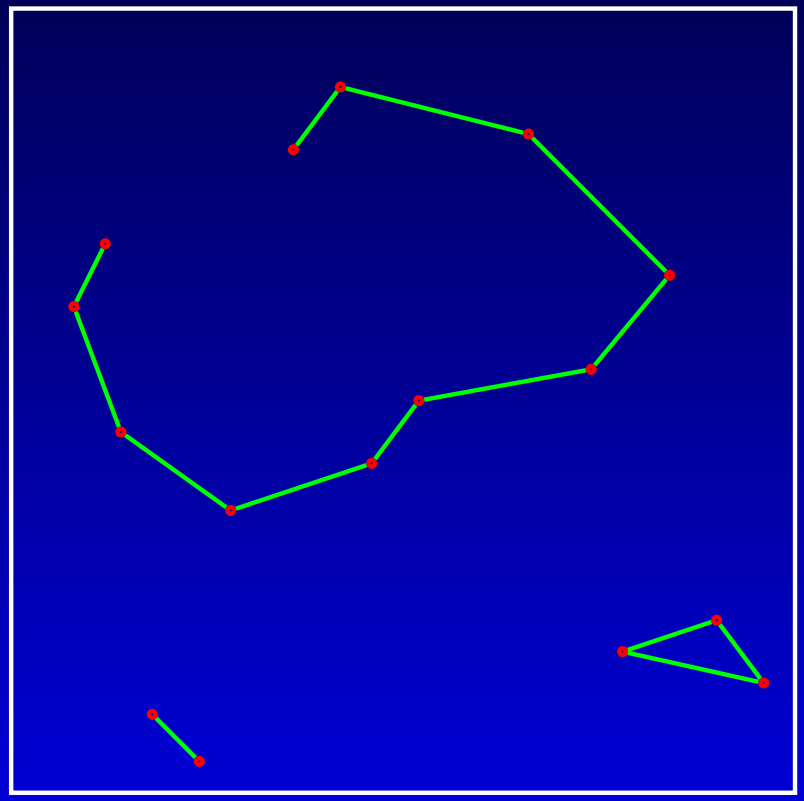
The problem

If $k_v = 2$ for each v:



The problem

If $k_v = 2$ for each v:



The problem

Prb: Design a distributed algorithm to choose the k_v's which

- is fast (in terms of communication rounds),
- sets up a connected network, and
- minimizes the average/maximum out-degree.
  - ensures low traffic/congestion and low power consumption.

- The three goals are partially conflicting.

Known results

[Xue & Kumar'04]

- Set $k_v = K \log n$.

:-) The network is connected whp.

:-) Very fast: $O(\log n)$.

:-(avgdeg = maxdeg = $O(\log n)$.

- best possible for uniform k_v's.

Known results

[Kucera'05]

- Set $k_v = K$.
- If $|component(v)| \leq n/2$ increase k_v by one and repeat the process.
- Otherwise, respond to the "requests" of the nodes enlarging their out-degree.

:-) The network is connected whp.

:-) avgdeg = $O(1)$, maxdeg = $O(\log n)$.

- best possible for arbitrary k_v's.
- $O(opt)$ power consumption.

:-(Very slow: $\Omega(n)$ (?).

Our results

Thr: By setting $k_v = K$ for each v, for a sufficiently large universal constant K, whp

- there is a giant component of $\Theta(n)$ nodes,
- all the other components contain at most $C \log^2 n$ nodes, and
- small components cannot merge into a component larger than $C \log^2 n$ without reaching the giant component before.

$\Rightarrow$ Idea: replace the (slow) test in Kucera's algorithm with a faster probabilistic test.

Our results

- ◇ Set $k_v = K$.
- ◇ If $|component(v)| < C \log^2 n$ increase k_v by one and repeat the process.
- ◇ Otherwise, respond to the "requests" of the nodes enlarging their out-degree.

:-) The network is connected whp.

:-) avgdeg = $O(1)$, maxdeg = $O(\log n)$.

:-) Fast: $O(\log^3 n)$.

- By reaching the giant component directly: time=O$(\log^2 n)$, maxdeg=$O(\log^2 n)$.

Our results

authors	X&K	K	D,G&P
time	$O(\log n)$	$\Omega(n)$	$O(\log^3 n)$
avgdeg	$O(\log n)$	$O(1)$	$O(1)$
maxdeg	$O(\log n)$	$O(\log n)$	$O(\log n)$

Outline

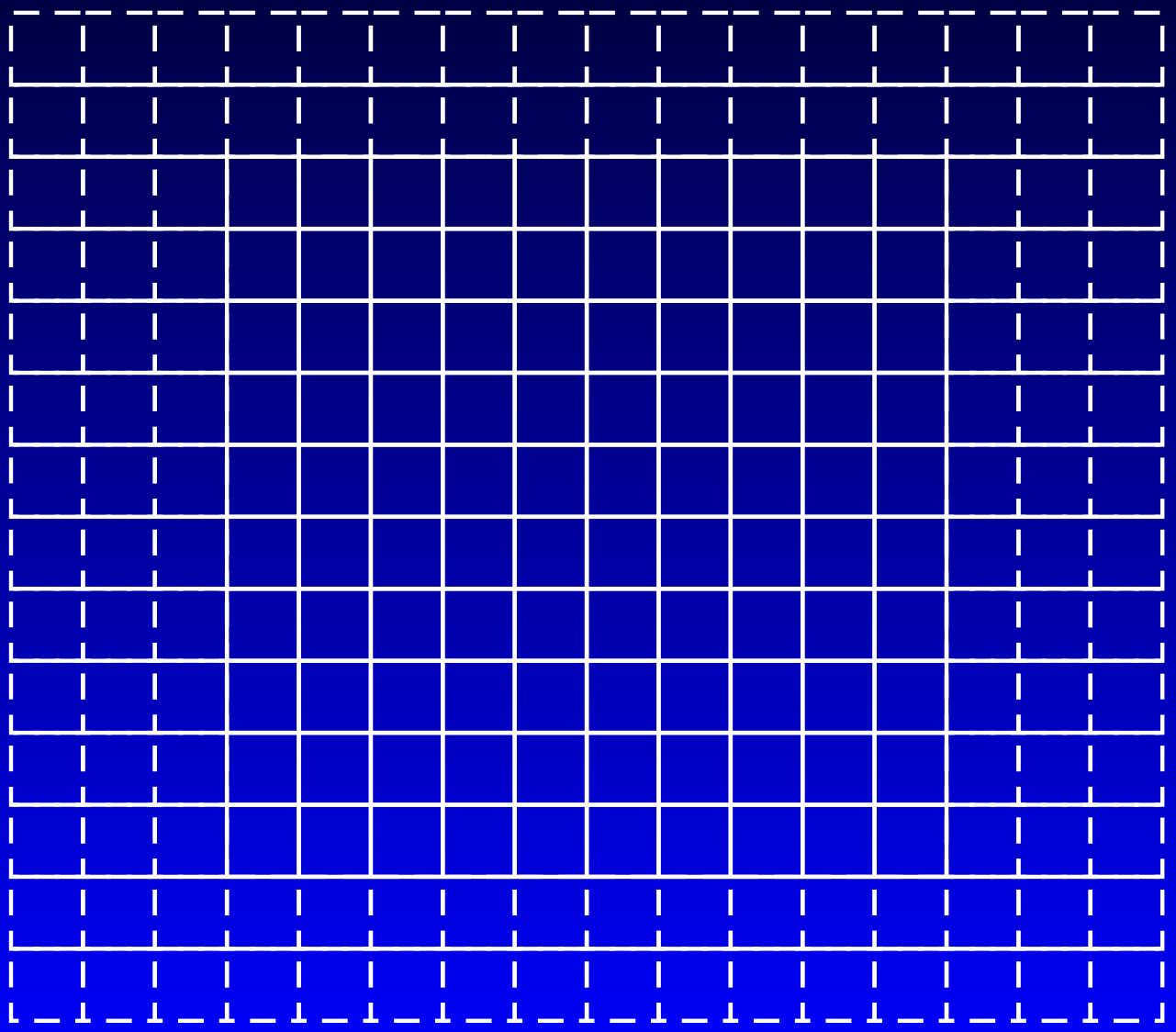
- In the rest of this talk we'll focus on our main theorem, which guarantees connectivity whp.
- The proof of the theorem is based on standard results from percolation theory.
  - Short introduction to percolation.
  - Reduction to percolation.
- The proof of the claims on time and degree follows along the same line (but it is more technical).

Percolation

Consider an infinite grid of cells, where each cell is (independently) on with probability p_{on}, and off otherwise

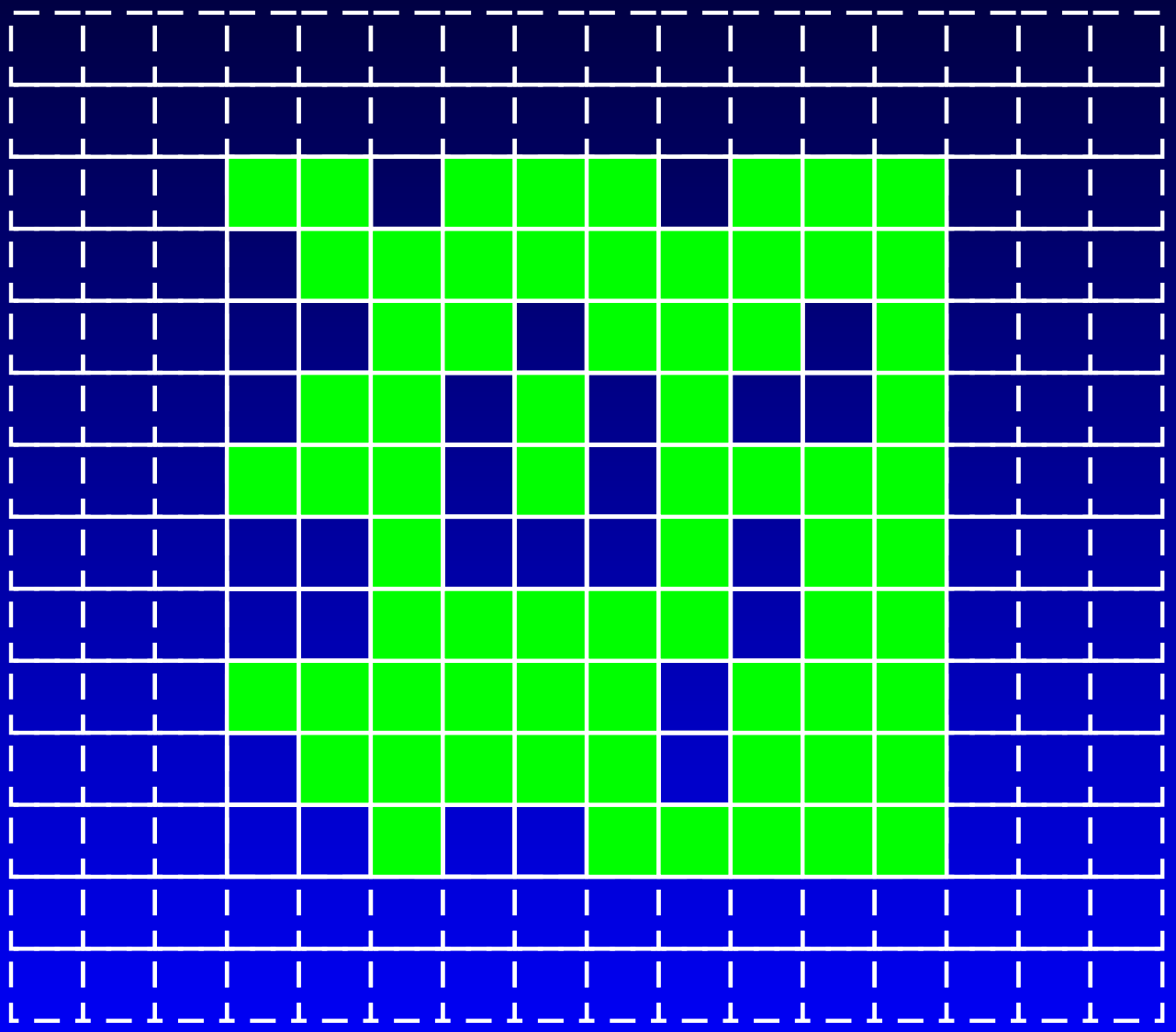
Percolation

Consider an infinite grid of cells, where each cell is (independently) on with probability p_{on}, and off otherwise



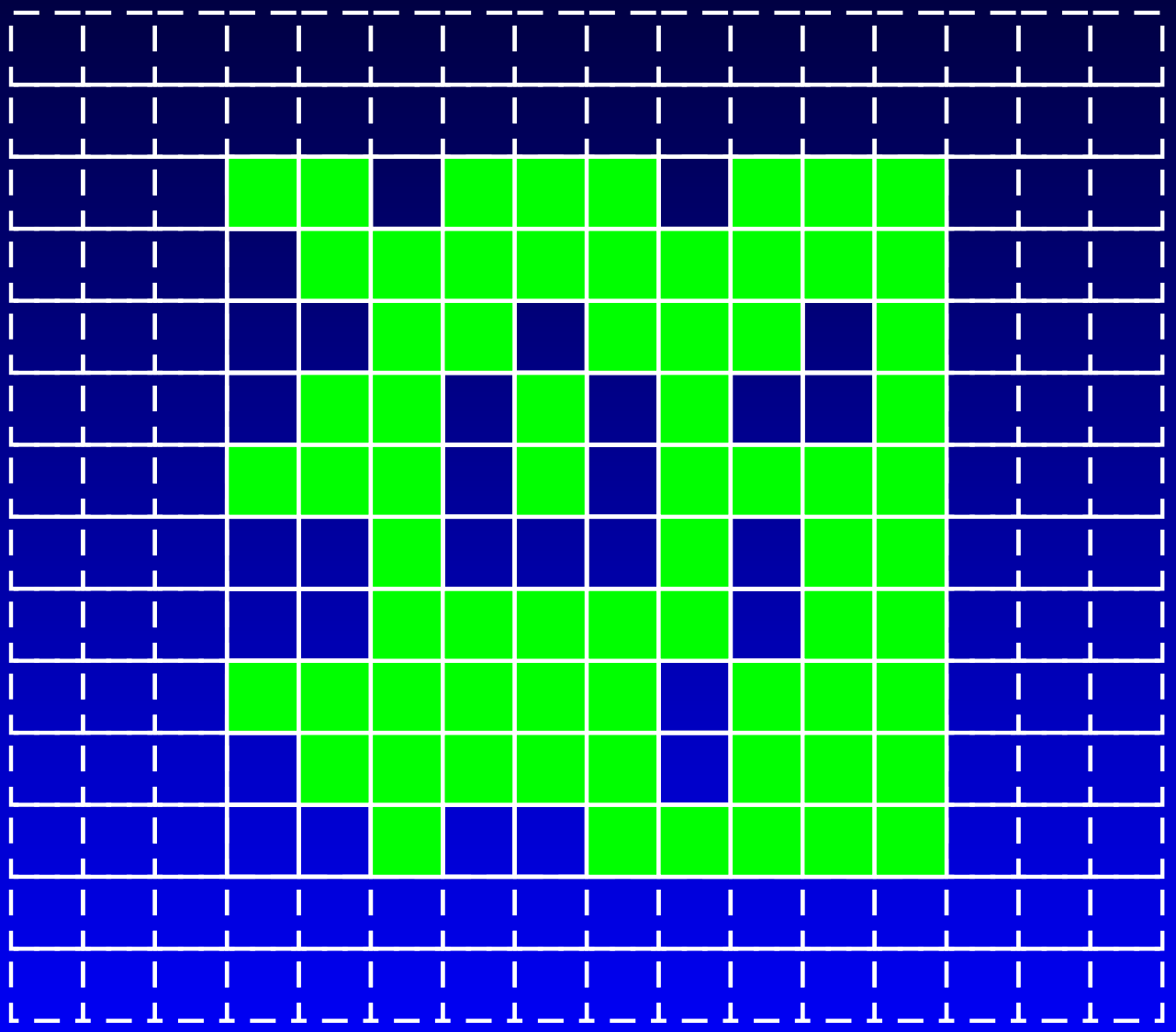
Percolation

Consider an infinite grid of cells, where each cell is (independently) on with probability p_{on}, and off otherwise



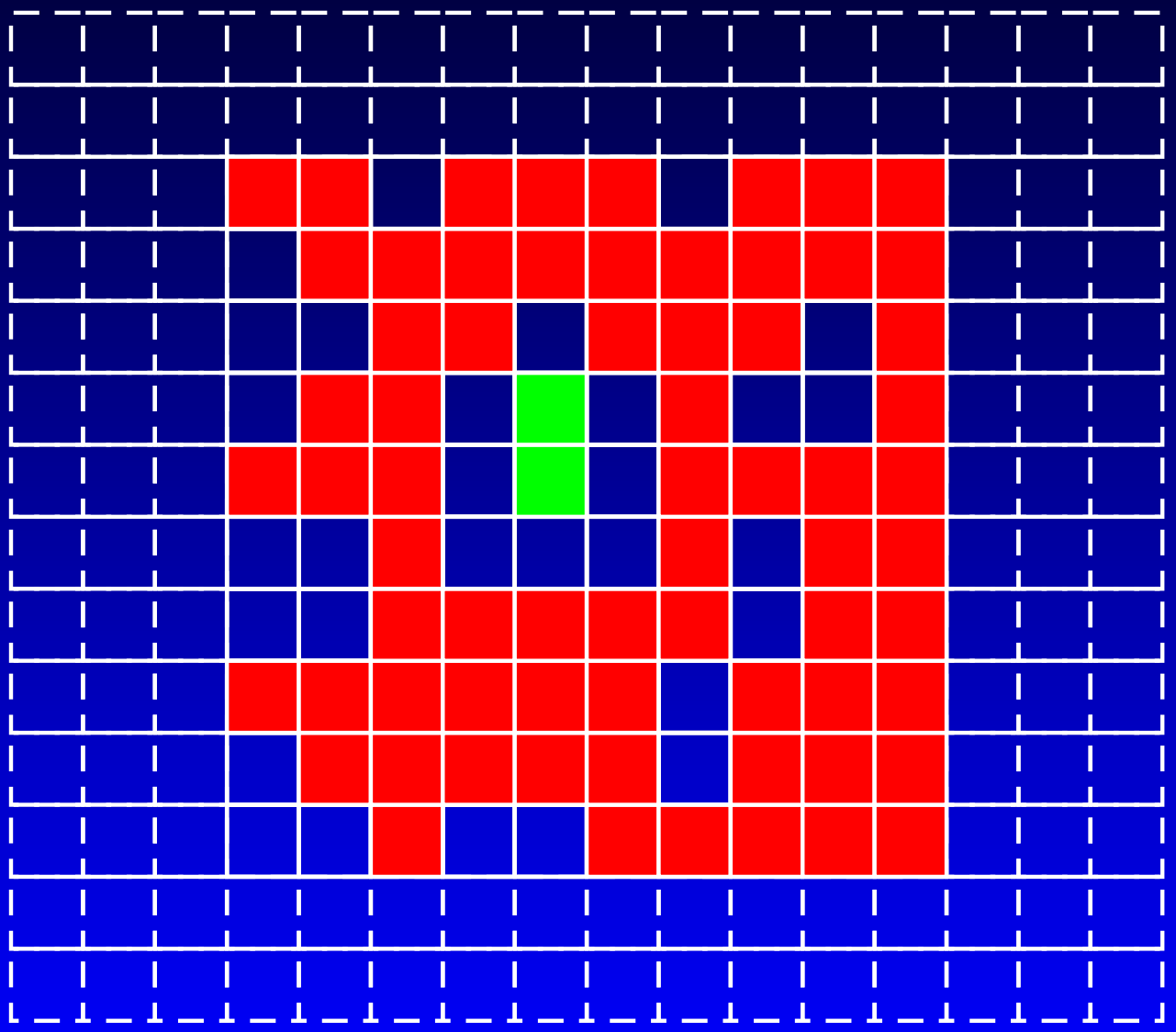
Percolation

A cluster is a connected component of on cells, where each cell is adjacent to the cells to its top, bottom, right, and left.



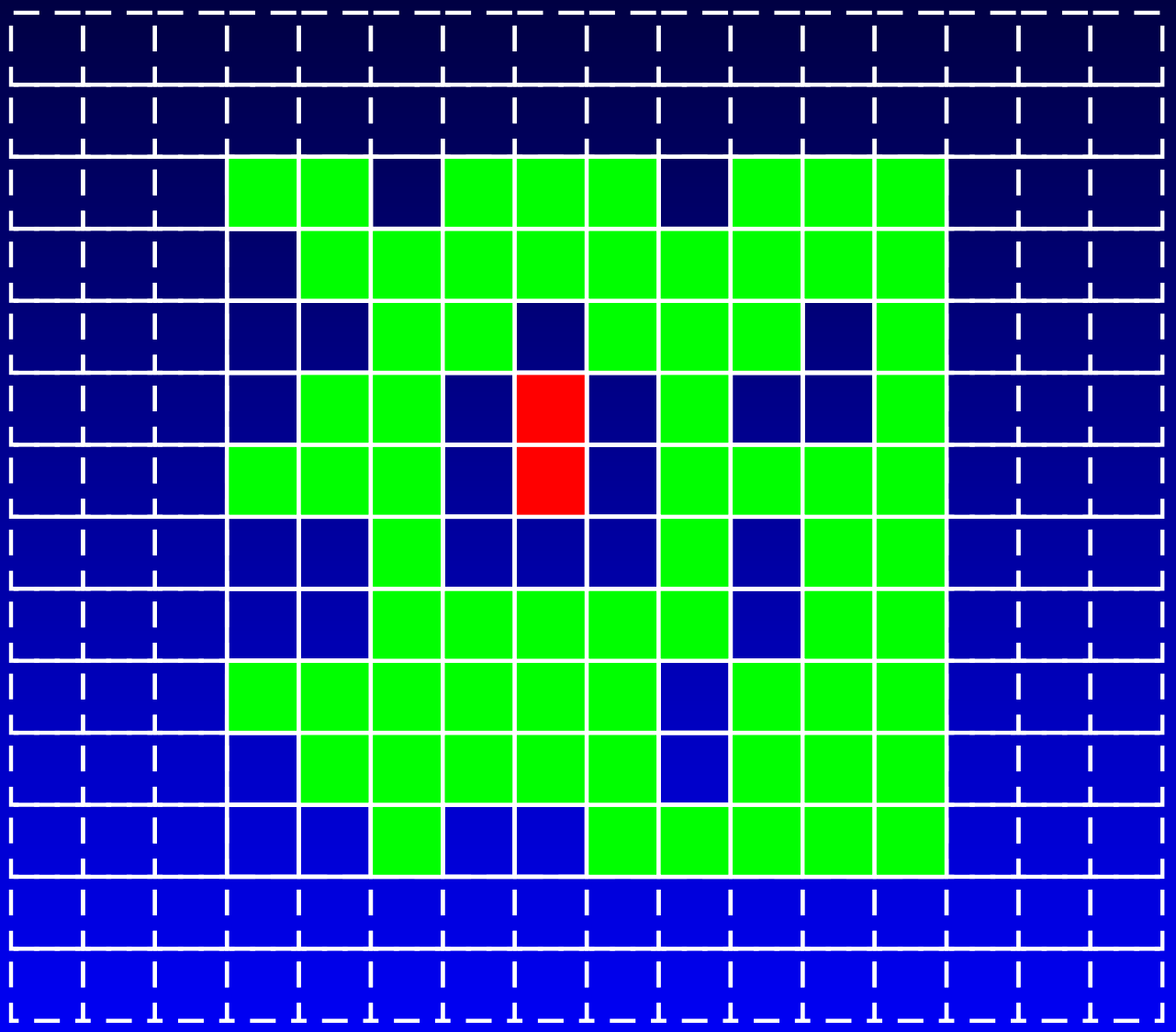
Percolation

A cluster is a connected component of on cells, where each cell is adjacent to the cells to its top, bottom, right, and left.



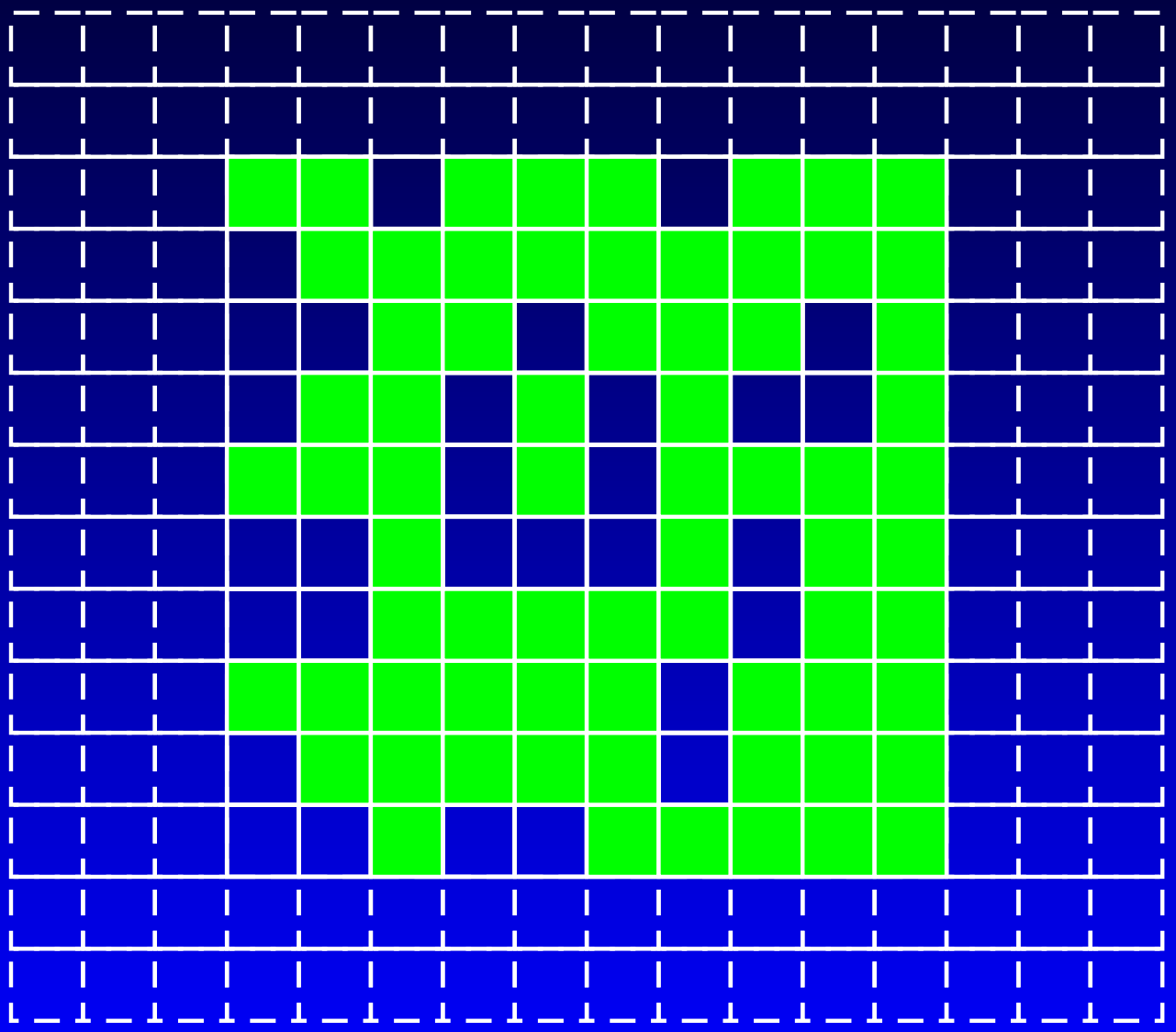
Percolation

A cluster is a connected component of on cells, where each cell is adjacent to the cells to its top, bottom, right, and left.



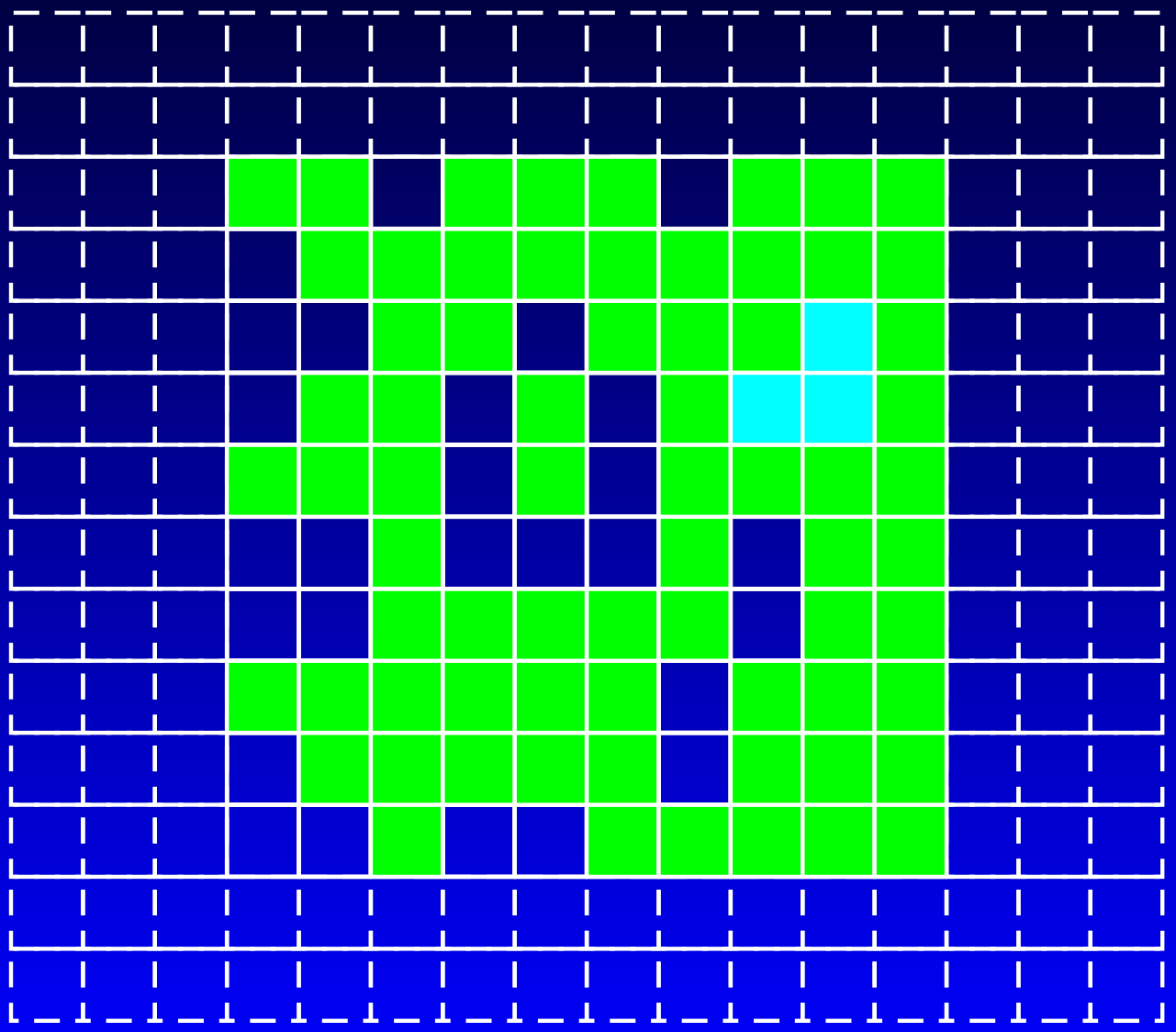
Percolation

A lake is a *-connected component of off cells, together with its interior part.



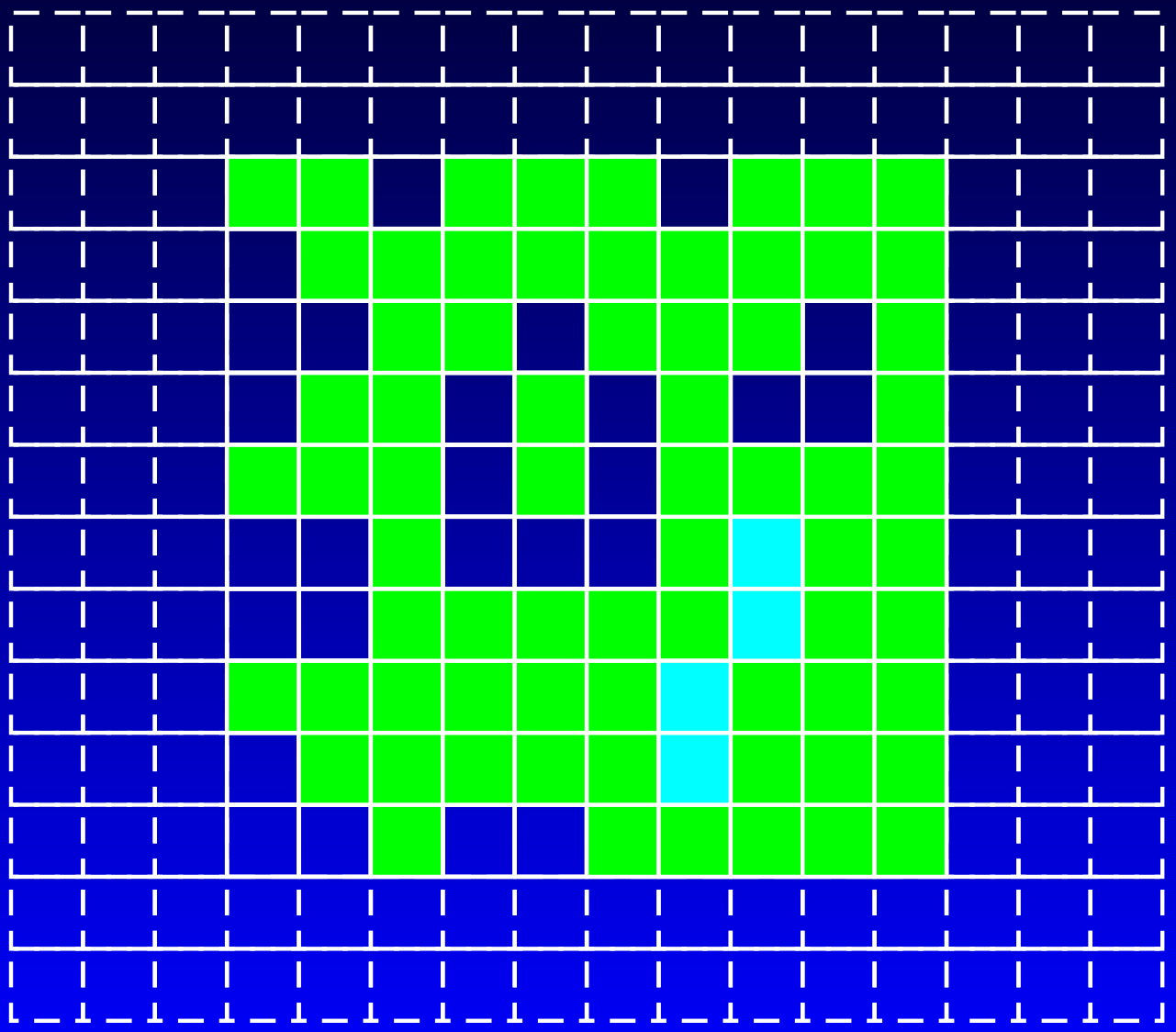
Percolation

A lake is a *-connected component of off cells, together with its interior part.



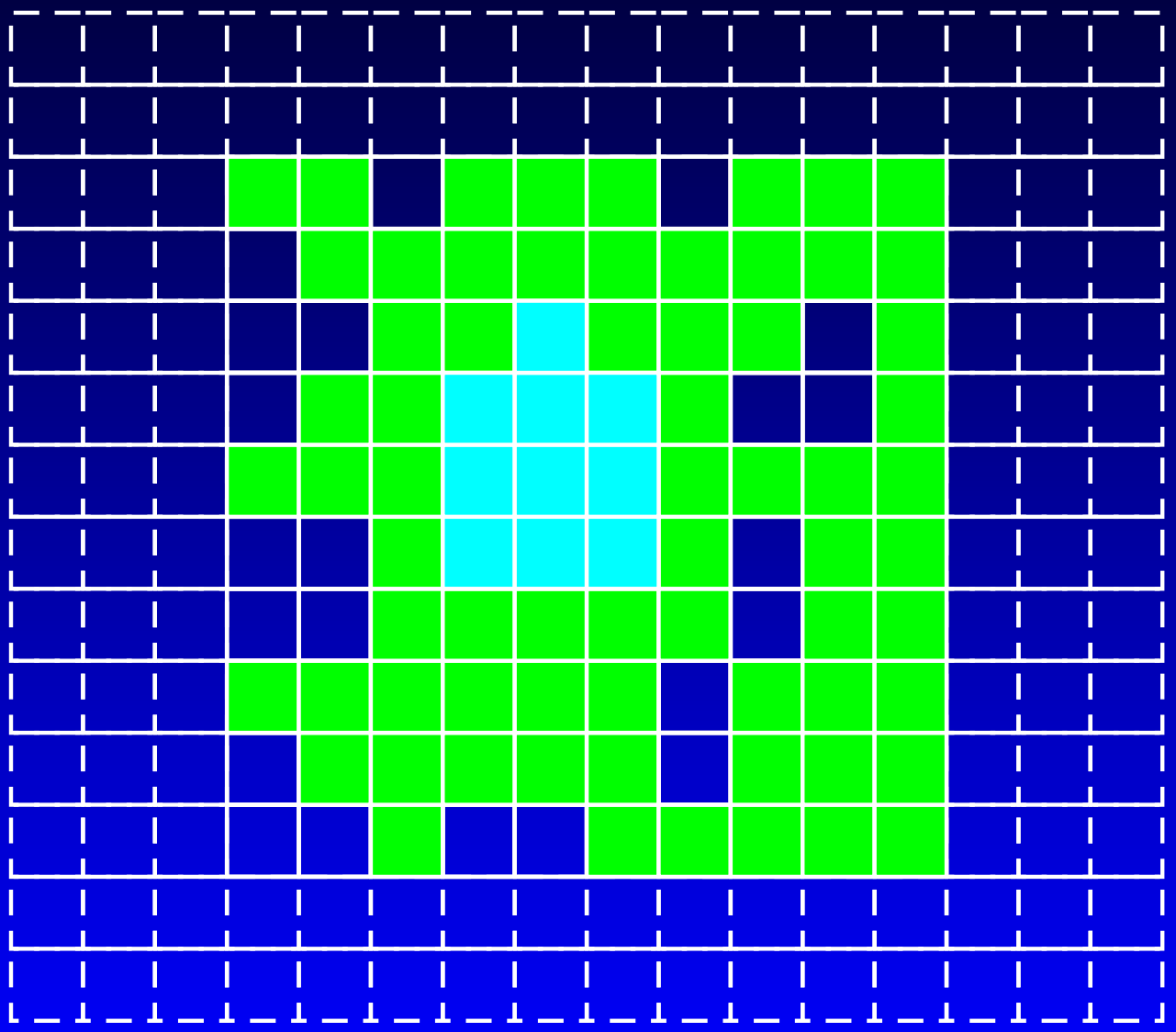
Percolation

A lake is a *-connected component of off cells, together with its interior part.



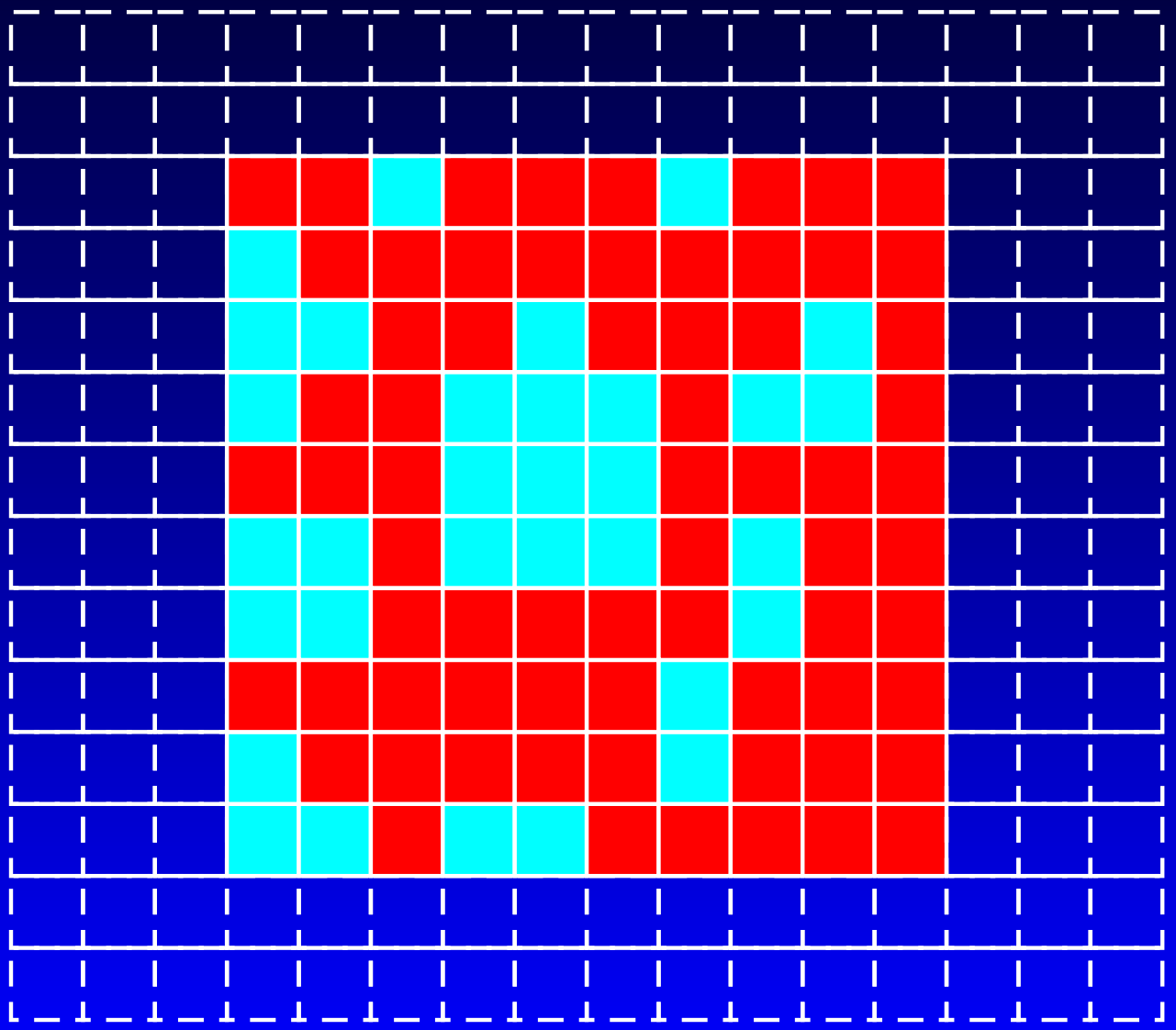
Percolation

A lake is a *-connected component of off cells, together with its interior part.



Percolation

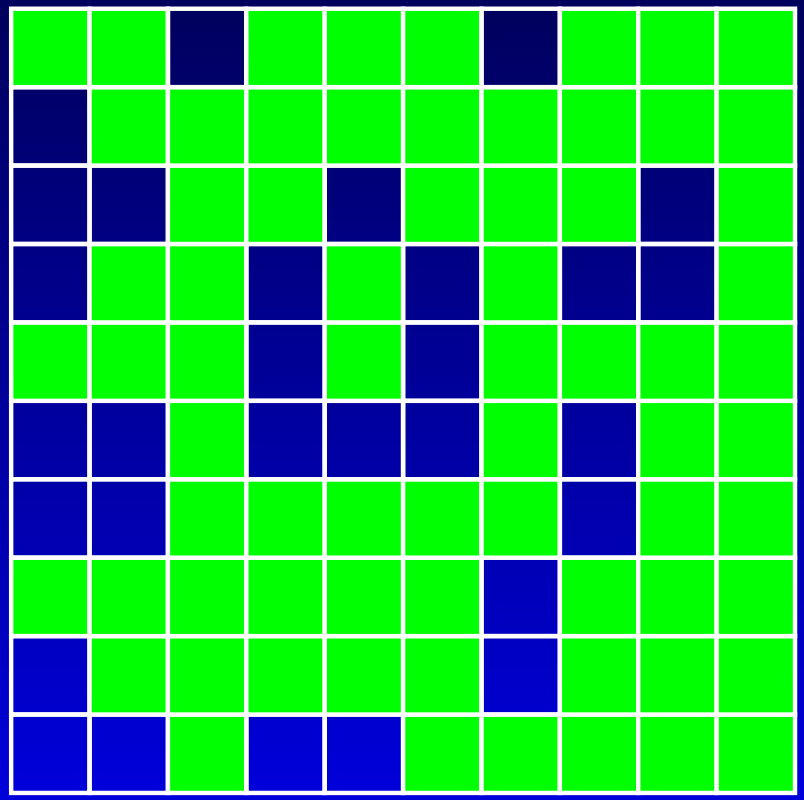
Thr: [Grimmet'89] For $p_{on} > p^* < 1$, almost surely there is a unique ∞ cluster, and $Pr[\,|lake| > k\,] \leq e^{-\gamma\sqrt{k}}$.



Box-percolation

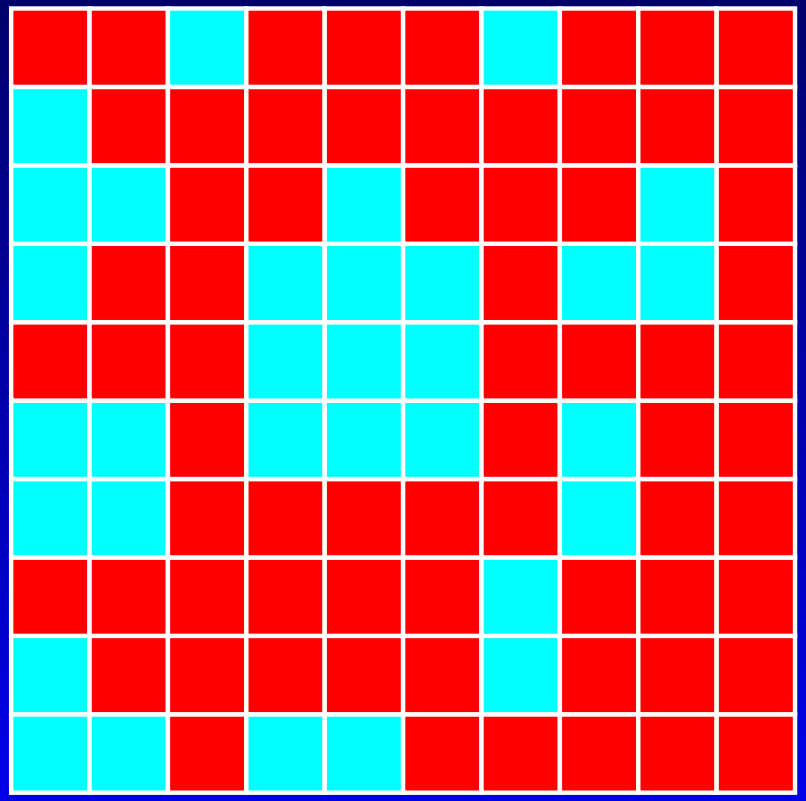
For our purposes, infinite grids are useless.

$\Rightarrow$ consider a square box of m cells.



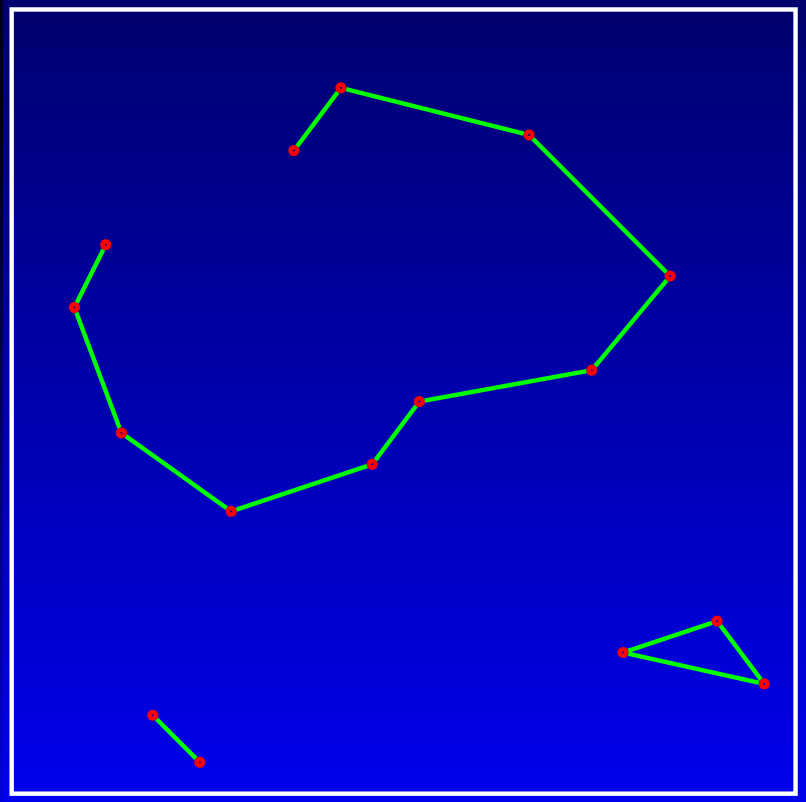
Box-percolation

Thr: [Deuschel,Pisztora'96] For $p_{on} > p^*$, whp the box contains a unique giant cluster of $\Theta(m)$ cells, and all the lakes have size $O(\log^2 m)$.

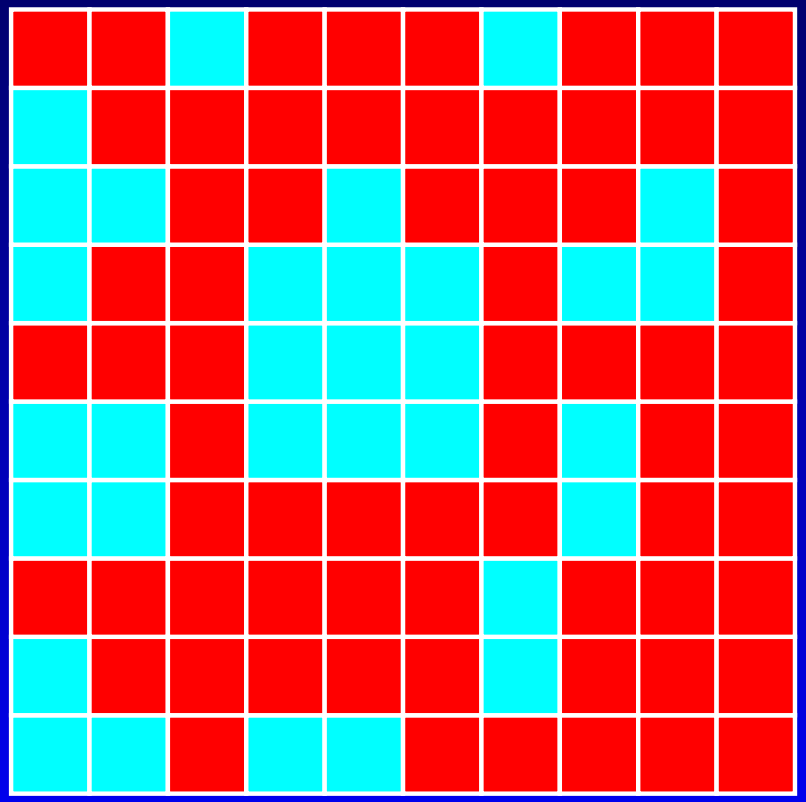


Reduction

$\Rightarrow$ Idea: "map" giant cluster into giant component and "trap" small components into lakes.



$\Leftrightarrow$?



Reduction

Our mapping requires a few intermediate steps:

- Scenario $\mathcal{A}$: Poisson process.
- Scenario $\mathcal{B}$: independent box-percolation, bad mapping.
- Scenario $\mathcal{C}$: h-dependent box-percolation, good mapping.
- Scenario $\mathcal{D}$: independent box-percolation, good mapping.

Scenario $\mathcal{A}$

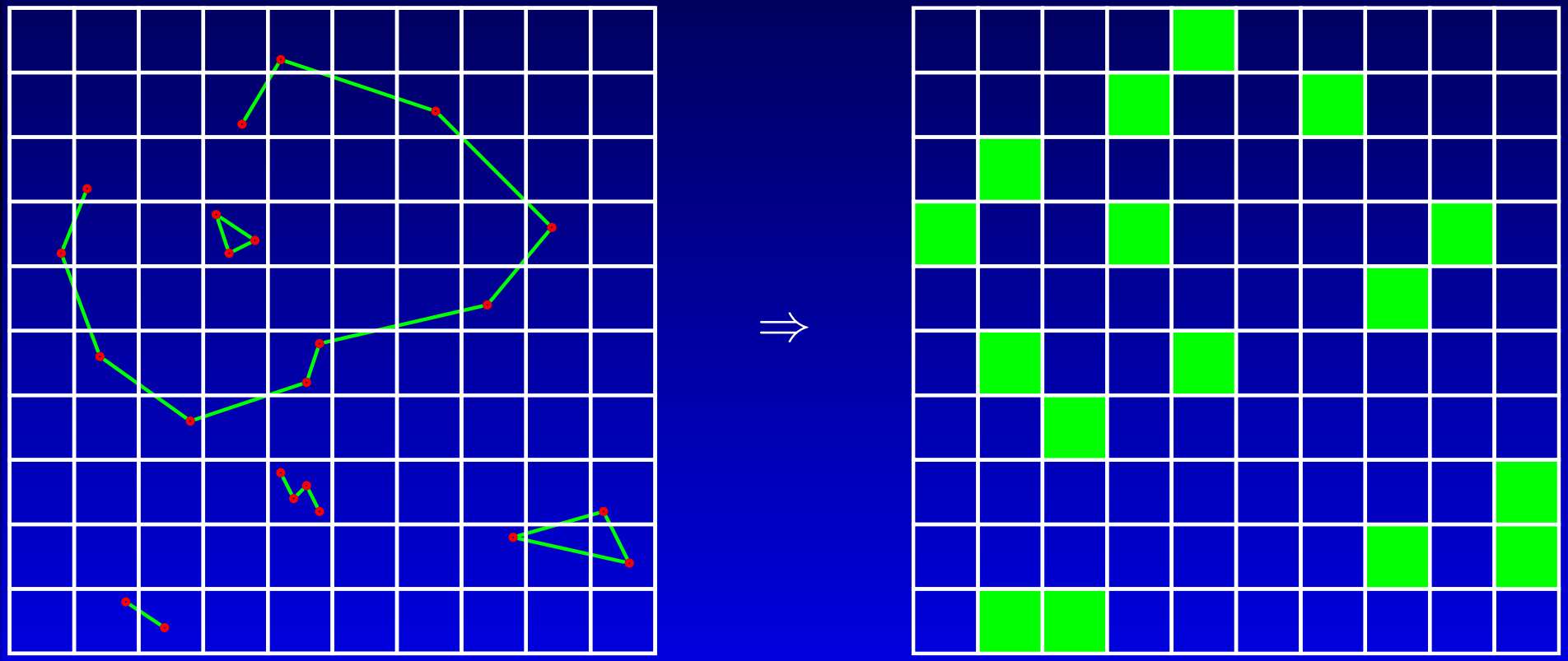
- Consider a Poisson process of mean n.
  - ◇ The number of points in a given region A is a Poisson variable of mean $|A|n$.
  - ◇ The number of points in disjoint regions are independent.
- What happens with probability at most p in Scenario $\mathcal{A}$, happens with probability at most $\sqrt{n} \cdot p$ in the original problem [Kucera'05].
  - ◇ It can be improved to $p + e^{-\gamma n}$.

Scenario $\mathcal{B}$

- Partition the unit square into a grid of $m = n/\alpha$ cells, where α is a (large) constant.
  - α plays a crucial role in the analysis, though it does not appear in the algorithm.
- A cell is on (good) if it contains between $\alpha/2$ and $3\alpha/2$ points.
  - $p_{on}^{\mathcal{B}} \to 1$ for $\alpha \to \infty$.

Scenario $\mathcal{B}$

For $n = 200$ and $\alpha = 2$ ($m = 100$):

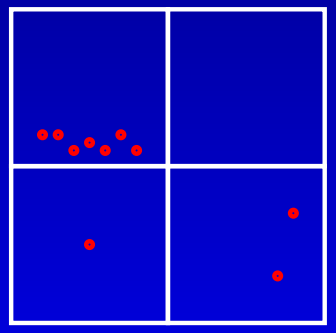


Scenario $\mathcal{B}$

:-) we have independent box-percolation.

- ◇ here we use the properties of Poisson processes.
- ◇ for a sufficiently large constant α, there is a giant cluster of good cells, and lakes are small.

:-(The points of the giant cluster do not form a giant component for any constant K.



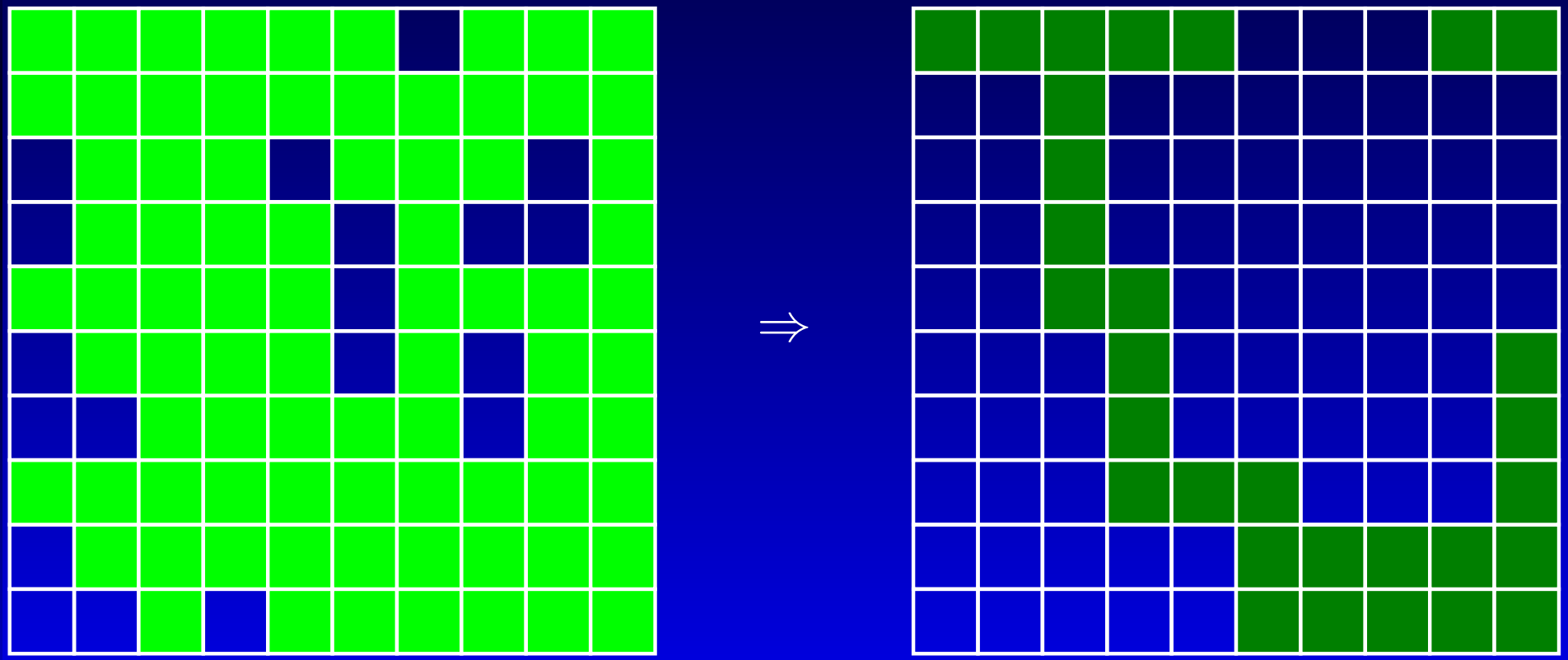
$\alpha = 2$

Scenario $\mathcal{C}$

- Consider the same partition into cells as in Scenario $\mathcal{B}$.
- A cell is on (open) in scenario $\mathcal{C}$ if in Scenario $\mathcal{B}$ the same cell is good together with all the cells at *-distance at most $D = O(1)$.
  - $p_{on}^{\mathcal{C}} \to 1$ for $p_{on}^{\mathcal{B}} \to 1$ $(\alpha \to \infty)$.

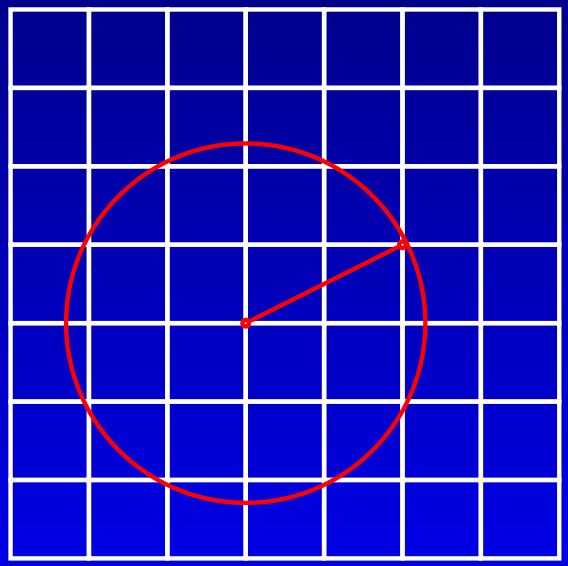
Scenario $\mathcal{C}$

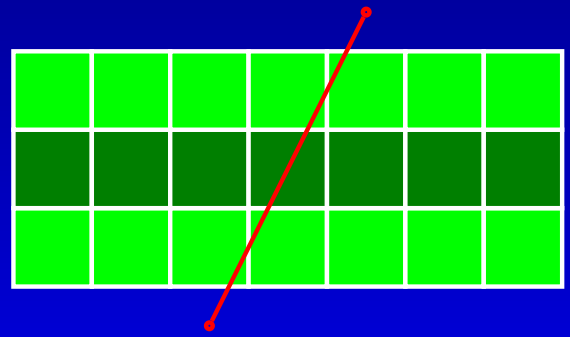
For $D = 1$:



Scenario $\mathcal{C}$

- for $D \geq 3$ and $K \geq 7^2(3\alpha/2)$, points in adjacent on cells form a clique (cells with too many points are far away).
- for $2(D-3)(\alpha/2) > K$, points inside distinct lakes cannot reach each other without reaching the giant component before (the giant component is thick).



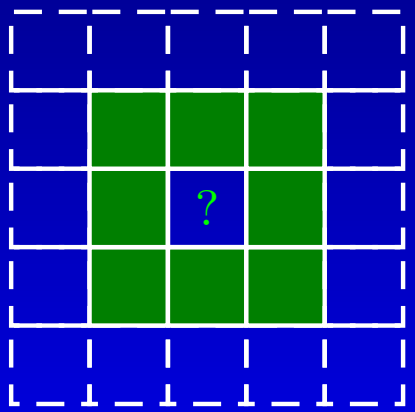


Scenario $\mathcal{C}$

:-) Good mapping

- A giant cluster of cells translates into a giant component of points.
- Small components contained in distinct lakes cannot merge without reaching the giant component first.

:-(We have no independent box-percolation.



Scenario $\mathcal{D}$

- [Ligget,Schonmann,Stacey'97] There is a coupling between two probability distributions of on cells $\mathcal{C}'$ and $\mathcal{D}'$ such that:
  - ◇ if a cell is on in $\mathcal{D}'$, then it is on also in $\mathcal{C}'$;
  - ◇ the marginal distribution of $\mathcal{C}'$ is the distribution of $\mathcal{C}$;
  - ◇ the marginal distribution of $\mathcal{D}'$ is an independent box-percolation of parameter $p_{on}^{\mathcal{D}}$;
  - ◇ $p_{on}^{\mathcal{D}'} \to 1$ for $p_{on}^{\mathcal{C}'} \to 1$.
- The marginal distribution of $\mathcal{D}'$ is our Scenario $\mathcal{D}$.
  - ◇ $\mathcal{C}$ stochastically dominates $\mathcal{D}$.

:-) the same good mapping as in Scenario $\mathcal{C}$.

:-) we can use all the standard results from independent box-percolation.

Unfolding

For a sufficiently large constant α, whp

- There is a giant cluster in $\mathcal{D}$.

$\Rightarrow$ Same in $\mathcal{C}$.

$\Rightarrow$ There is a giant component in $\mathcal{A}$.

- All the lakes of $\mathcal{D}$ have size $O(\log^2 n)$.

$\Rightarrow$ Same in $\mathcal{C}$.

$\Rightarrow$ Since small components of $\mathcal{A}$ are trapped inside lakes of $\mathcal{C}$, by standard large deviation all such components have size $O(\log n) * O(\log^2 n) = O(\log^3 n)$

$\Rightarrow$ ... $O(\log^2 n)$ with a slightly more complicated analysis.

Slightly more complicated analysis

Lem: Let Z_i be the number of points in cell i, and let L be a lake in scenario $\mathcal{C}$ with n points. Then, for large enough number of points n, there is a constant $\gamma > 0$ such that

$$\Pr(\sum_{i \in L} Z_i > h) \leq e^{-\gamma\sqrt{h}}.$$

$\square$.

Let $B := (B_1, \ldots, B_m)$ be the random vector denoting which cells are good or bad, and $b = (b_1, \ldots, b_m)$ any particular such configuration. Then

$$\begin{aligned}
\Pr(\sum_{i \in L} Z_i > h) &= \sum_k \Pr\left(\sum_{i \in L} Z_i > h \mid |L| = k\right) \Pr(|L| = k) \\
&= \sum_k \sum_b \Pr\left(\sum_{i \in L} Z_i > h \mid |L| = k, B = b\right) \Pr(B = b \mid |L| = k) \Pr(|L| = k) \\
&= \sum_k \sum_b \Pr(\sum_{i \in L} Z_i > h \mid B = b) \Pr(B = b \mid |L| = k) \Pr(|L| = k).
\end{aligned}$$

The last equality follows, since if we know B we also know the size of L. We now focus on the term $\Pr(\sum_{i \in L} Z_i > h \mid B = b)$. We will show that we can replace the variables $(Z_i \mid B = b)$ with a set of i.i.d. variables that stochastically dominate them and that obey the large deviation principle. The Poisson process can be realized as the product of m independent Poisson processes, each operating inside a cell. This implies that if we have a set of events E_i where each event depends only on what happens in cell i, then $\Pr(\cap_i E_i) = \prod_i \Pr(E_i)$. Thus, we have

$$\Pr(\cap_i \{Z_i = h_i\} | B = b) = \frac{\Pr(\cap_i \{Z_i = h_i, B_i = b_i\})}{\Pr(\cap_i \{B_i = b_i\})} = \frac{\prod_i \Pr(Z_i = h_i, B_i = b_i)}{\prod_i \Pr(B_i = b_i)} = \prod_i \Pr(Z_i = h_i | B_i = b_i).$$

Slightly more complicated analysis

If we define $X_i = (Z_i \,|\, B_i = \text{good})$ and $Y_i = (Z_i \,|\, B_i = \text{bad})$, it follows that $\sum_i (Z_i \,|\, B)$ has the same law of the sum of independent variables each of which is X_i or Y_i depending on whether cell i is good or bad. Let us define a collection of i.i.d. random variables W_i's each of which has the distribution of $(Z_i \,|\, Z_i > 2\alpha)$. Each W_i stochastically dominates both X_i and Y_i so that

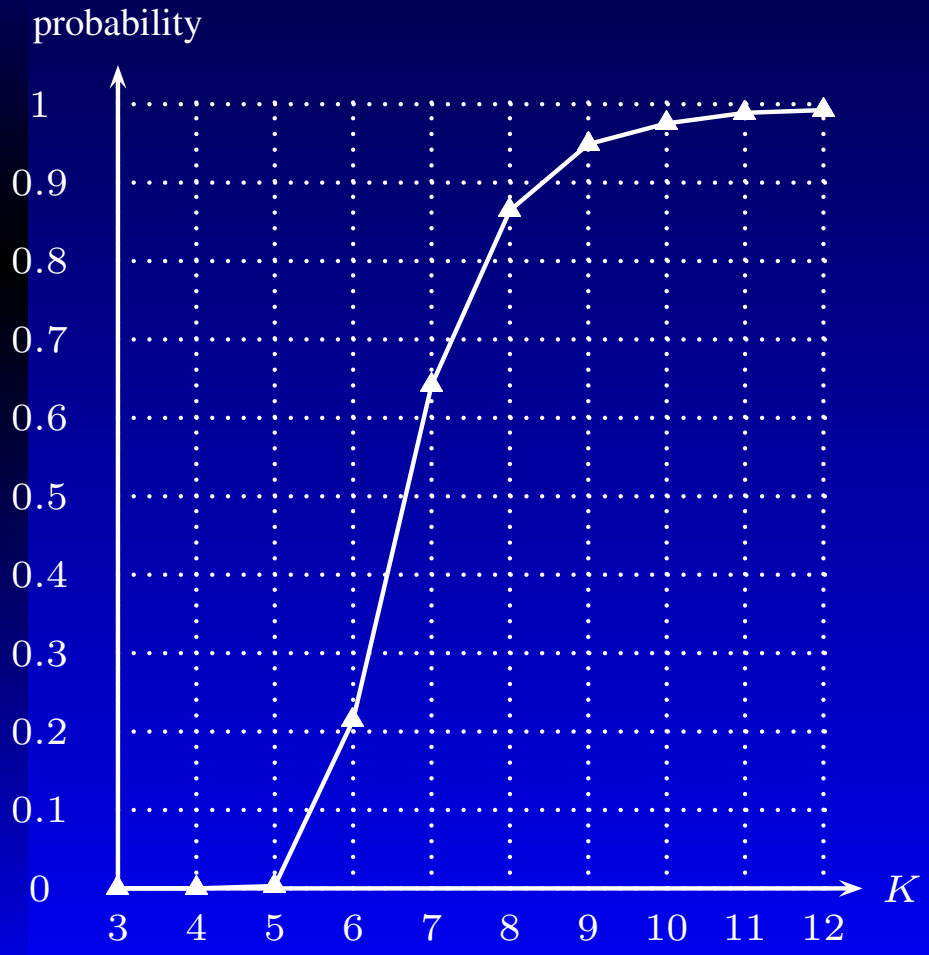
$$\Pr(\sum_{i \in L} Z_i > h \mid B = b) \leq \Pr(\sum_{i \in L} W_i > h),$$

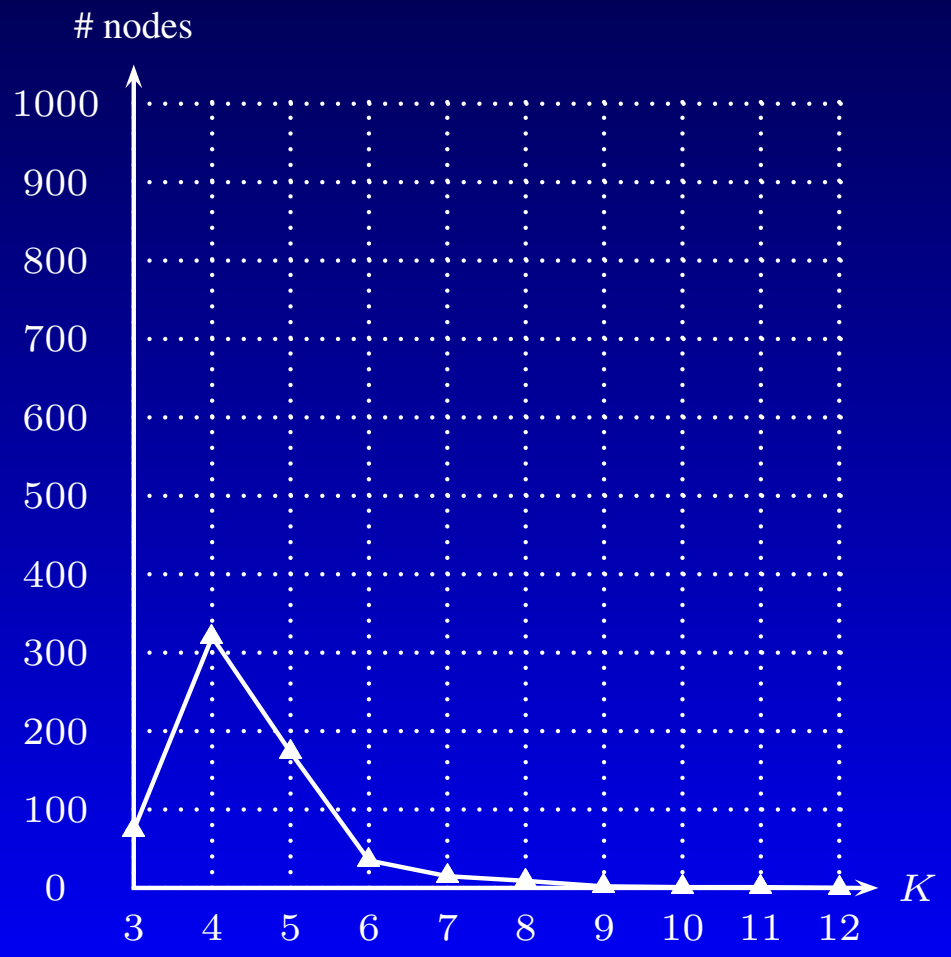
for each configuration b. Moreover the W_i obey the large deviation principle, i.e. the probability of large deviations from the mean is exponentially small. We thus have, for $\beta < 1/E[W_1]$,

$$\begin{aligned}
\Pr(\sum_{i \in L} Z_i > h) &= \sum_k \sum_b \Pr(\sum_{i \in L} Z_i > h \mid B = b) \Pr(B = b \mid |L| = k) \Pr(|L| = k) \\
&\leq \sum_k \sum_b \Pr(\sum_{i \leq k} W_i > h) \Pr(B = b \mid |L| = k) \Pr(|L| = k) \\
&= \sum_k \Pr(\sum_{i \leq k} W_i > h) \Pr(|L| = k) \\
&= \sum_{k \leq \beta h} \Pr(\sum_{i \leq k} W_i > h) \Pr(|L| = k) + \sum_{k > \beta h} \Pr(\sum_{i \leq k} W_i > h) \Pr(|L| = k) \\
&\leq \sum_{k \leq \beta h} \Pr(\sum_{i \leq \beta h} W_i > h) \Pr(|L| = k) + \sum_{k > \beta h} \Pr(\sum_{i \leq k} W_i > h) \Pr(|L| = k) \\
&\leq \sum_{k \leq \beta h} \Pr(\sum_{i \leq \beta h} W_i > h) + \sum_{k > \beta h} \Pr(|L| = k) \\
&= \beta h \Pr(\sum_{i \leq \beta h} W_i > h) + \sum_{k > \beta h} \Pr(|L| = k) \\
&\leq \beta h \, e^{-\gamma_1 h} + \sum_{k > \beta h} e^{-\gamma_2 \sqrt{k}} \leq e^{-\gamma \sqrt{h}}.
\end{aligned}$$

Experiments

On the left, probability of achieving connectivity directly in the first phase, for $n = 1000$ and for different values of K. On the right, 99-th percentile of the maximum size of the second largest component in the same scenario.





Open problems

- Can we decrease the running time from $O(\log^3 n)$ to $O(\log n)$?
  - Small components contain $O(\log^2 n)$ nodes. Which is their diameter?
  - Which is the shape of lakes?