Connectivity of sparsified random geometric graphs

Nicolas Broutin, *Inria*

joint work with

Luc Devroye, *McGill*

Gábor Lugosi, ICREA and Pompeu Fabra

Random geometric graphs

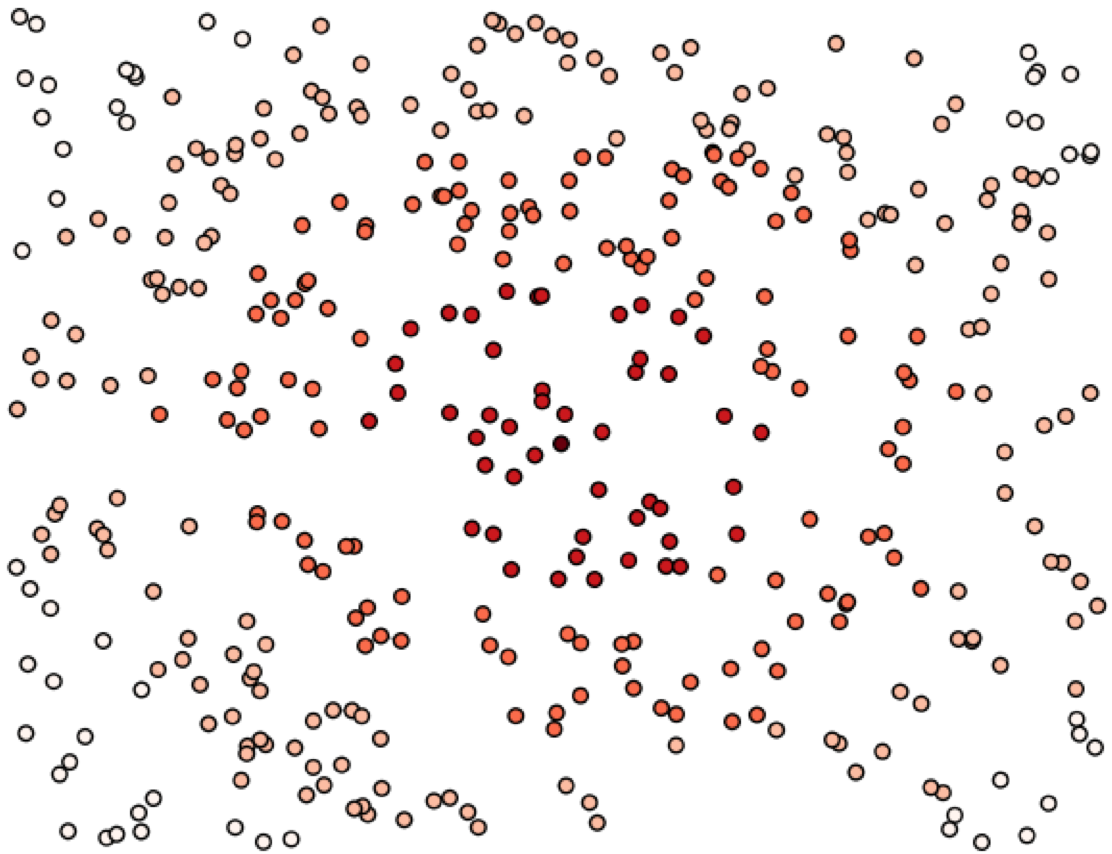
$G(n, r)$ $\quad$ $N =$ Poisson(n) uniform points in $\mathcal{D} \subseteq \mathbb{R}^d$

$i \sim j$ iif $\|X_i - X_j\| < r$

Random geometric graphs

$G(n, r)$ $\qquad N = \text{Poisson}(n)$ uniform points in $\mathcal{D} \subseteq \mathbb{R}^d$

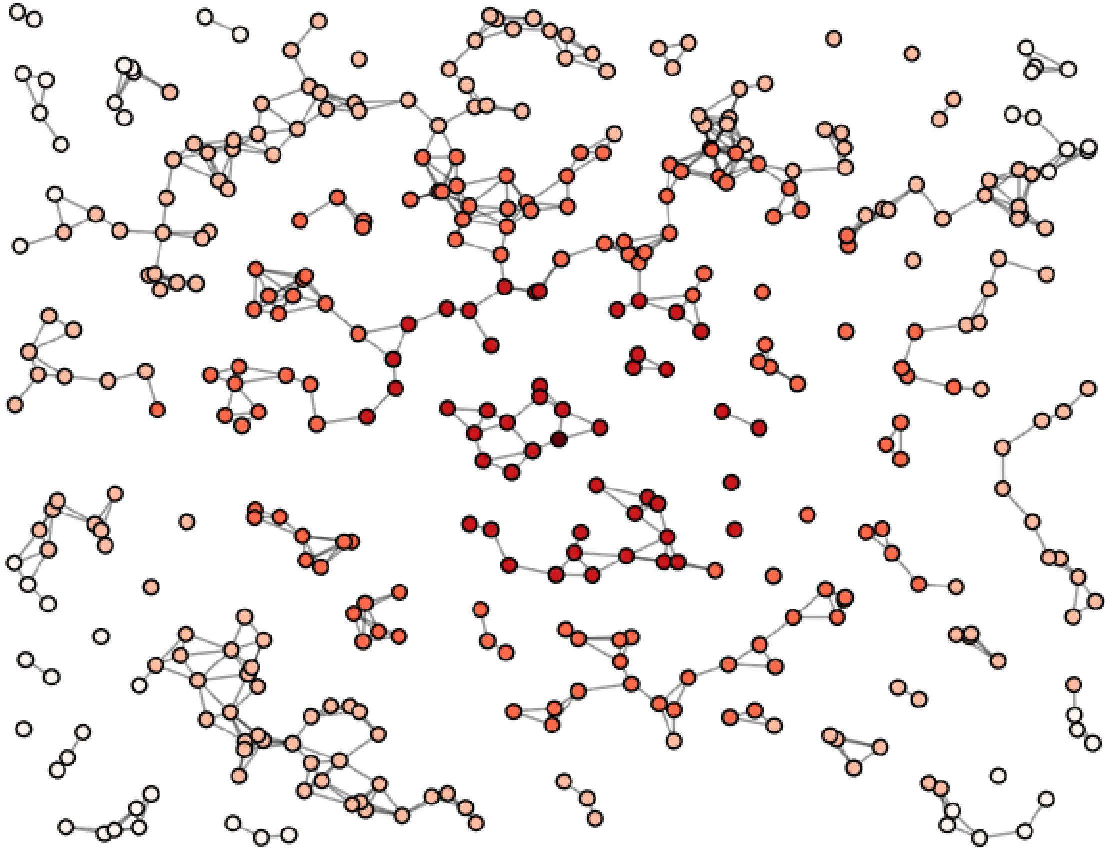
$i \sim j$ iif $\|X_i - X_j\| < r$



Random geometric graphs

$G(n,r)$ $N = \text{Poisson}(n)$ uniform points in $\mathcal{D} \subseteq \mathbb{R}^d$

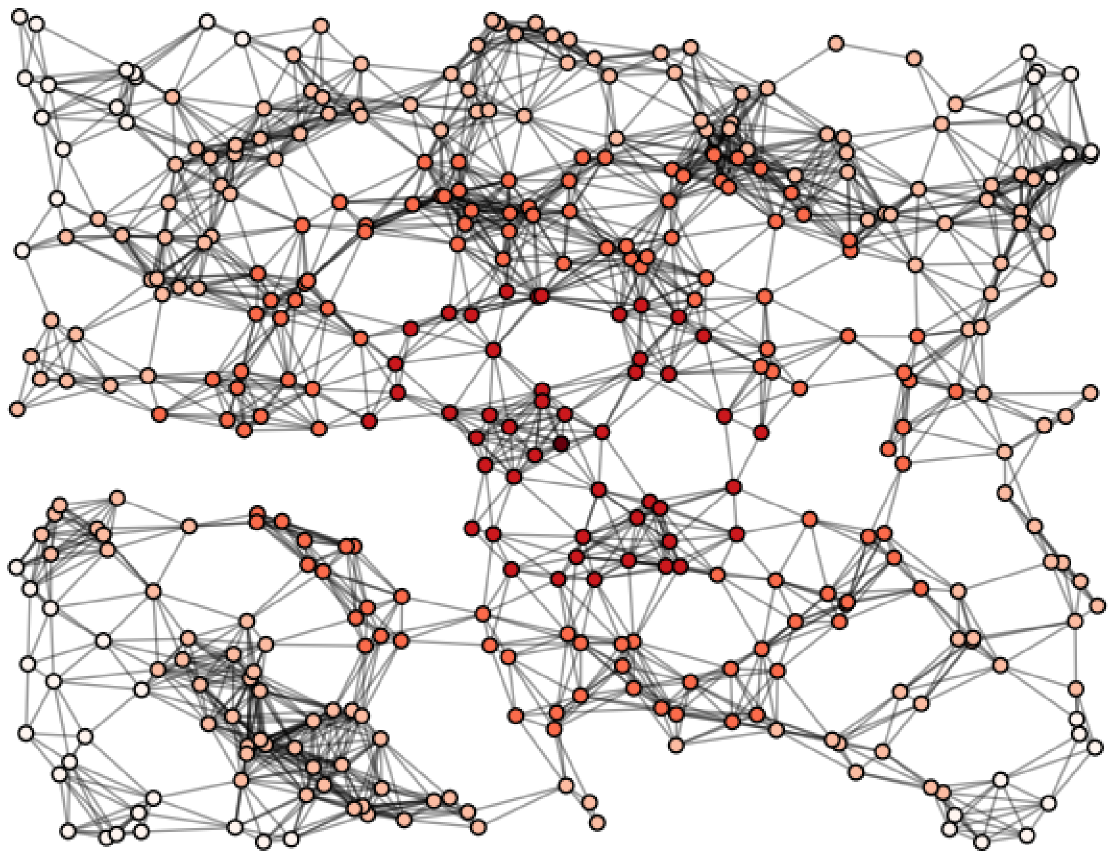
$i \sim j$ iif $\|X_i - X_j\| < r$



Random geometric graphs

$G(n, r)$ $N =$ Poisson(n) uniform points in $\mathcal{D} \subseteq \mathbb{R}^d$

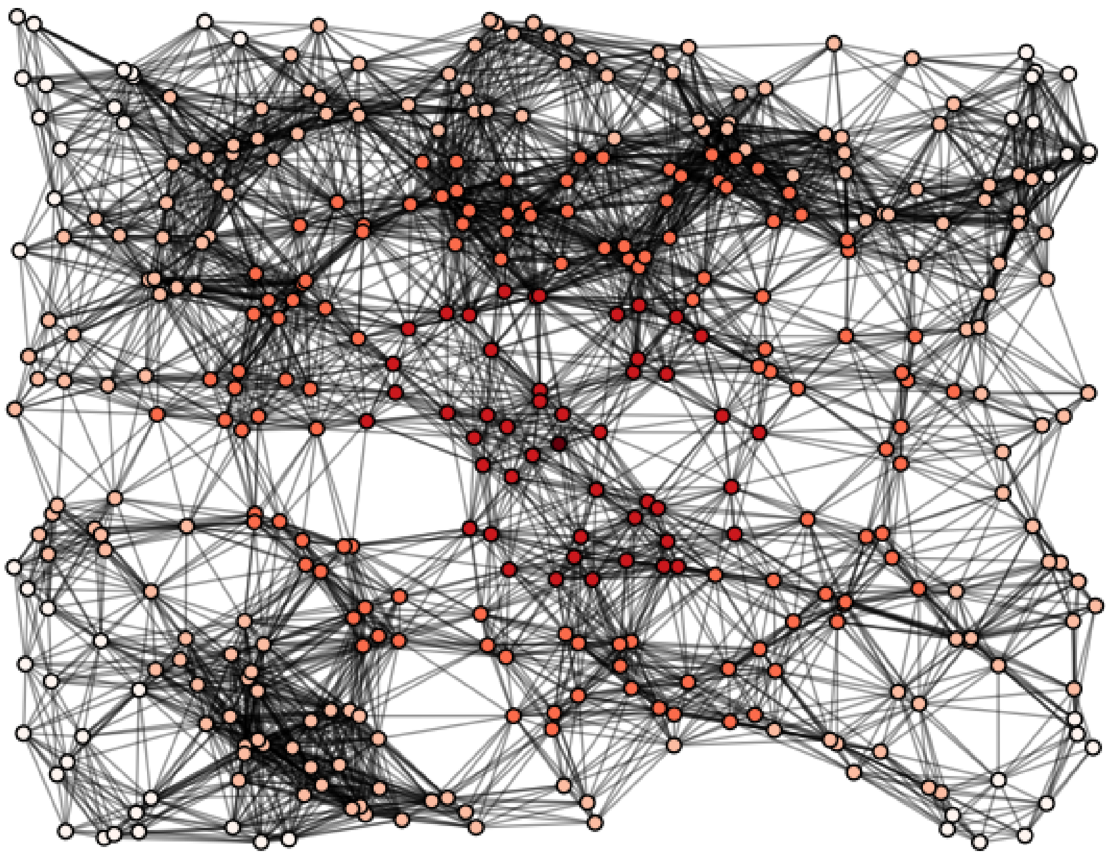
$i \sim j$ iif $\|X_i - X_j\| < r$



Random geometric graphs

$G(n, r)$ $N = \text{Poisson}(n)$ uniform points in $\mathcal{D} \subseteq \mathbb{R}^d$

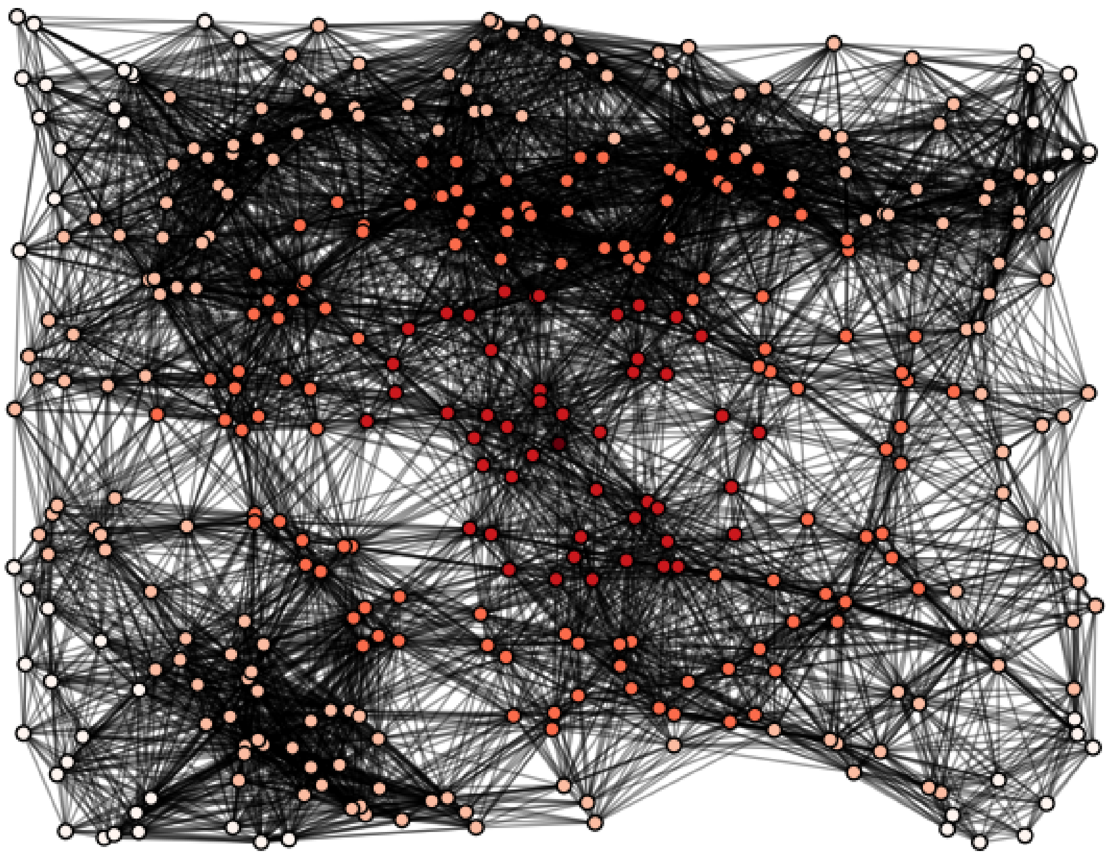
$i \sim j$ iif $\|X_i - X_j\| < r$



Random geometric graphs

$G(n, r)$ $\quad N = \text{Poisson}(n)$ uniform points in $\mathcal{D} \subseteq \mathbb{R}^d$

$i \sim j$ iif $\|X_i - X_j\| < r$



Connectivity of random geometric graphs

Theorem. ϑ_d the volume of the unit ball in $\mathbb{R}^d$

$r_n^\star$ such that $\vartheta_d n (r_n^\star)^d = \log n$

For any $\epsilon \in (0,1)$

$$\lim_{n\to\infty} \mathbf{P}\left(G(n, r_n) \text{ is connected}\right) = \begin{cases} 0 & \text{if } r_n \leq (1-\epsilon) r_n^\star \\ 1 & \text{if } r_n \geq (1+\epsilon) r_n^\star \end{cases}$$

Random irrigation graphs

$G(n, r)$ gets connected when average degree is $\Theta(\log n)$

idea: sparsify the graph
in a distributed way
while ensuring connected

irrigation graphs $S_n(r, c)$

every point "sees" his neighbours in $G(n, r)$

every point keeps c neighbours chosen at random

Random irrigation graphs

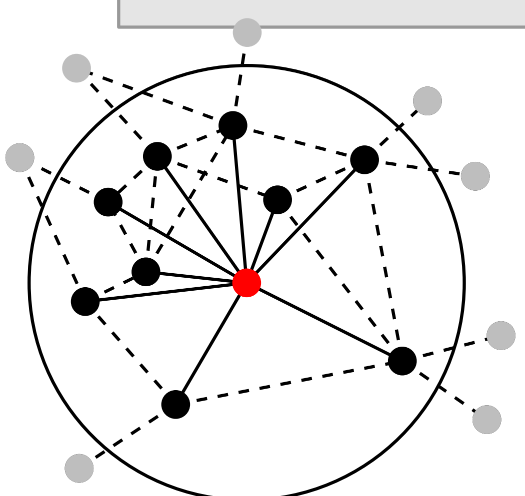
$G(n, r)$ gets connected when average degree is $\Theta(\log n)$

idea: sparsify the graph
in a distributed way
while ensuring connected

irrigation graphs $S_n(r, c)$

every point "sees" his neighbours in $G(n, r)$

every point keeps c neighbours chosen at random



Random irrigation graphs

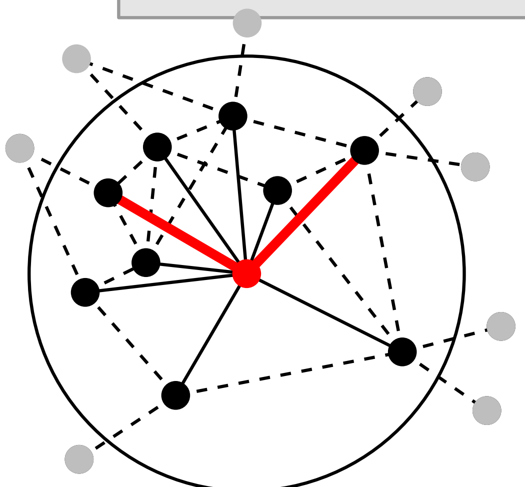
$G(n,r)$ gets connected when average degree is $\Theta(\log n)$

idea: sparsify the graph
in a distributed way
while ensuring connected

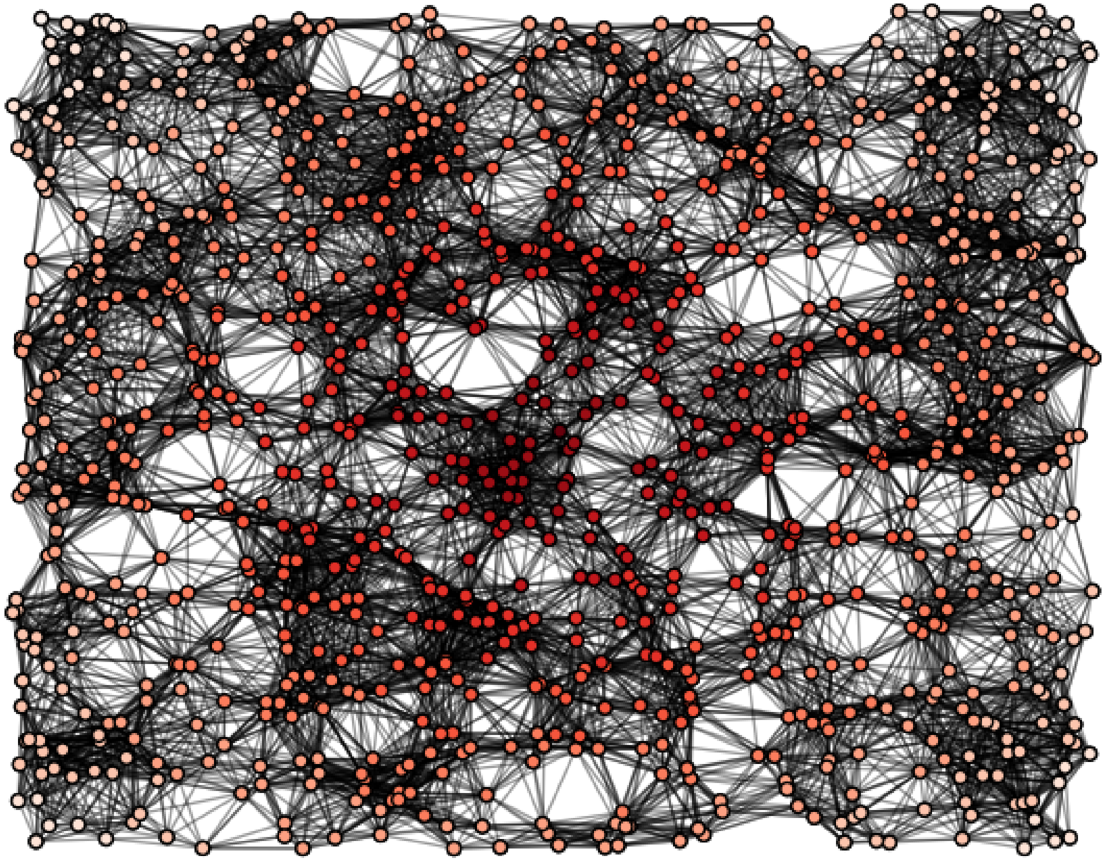
irrigation graphs $S_n(r,c)$

every point "sees" his neighbours in $G(n,r)$

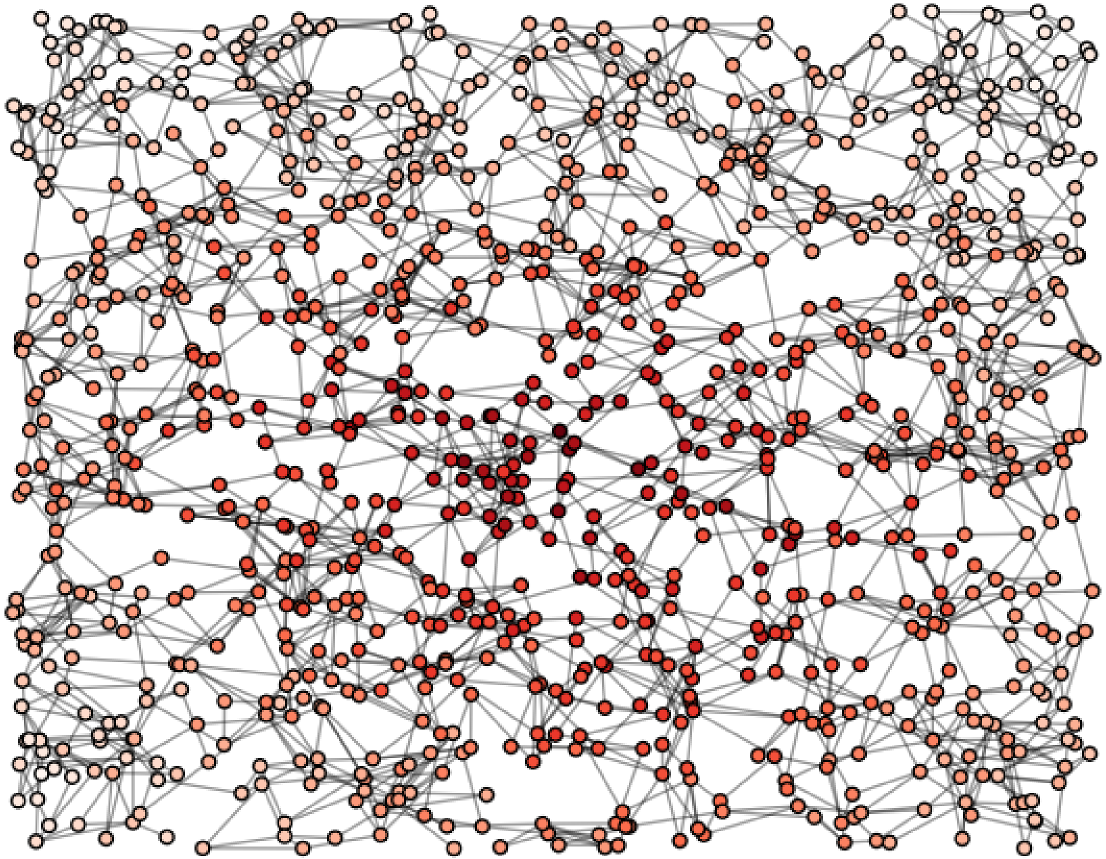
every point keeps c neighbours chosen at random



Random irrigation graphs



Random irrigation graphs



History and results

Previous results:

Dubhashi, Johansson, Häggström, Panconesi, and Sozio

$$r = \Theta(1) \quad \Rightarrow \quad S_n(r,2) \text{ is connected whp}$$

Crescenzi, Nocentini, Pietracaprina, Pucci

$$d = 2 \quad \begin{array}{l} r > \sqrt{\log n / n} \\ c > \gamma_2 \log(1/r) \end{array} \Rightarrow \quad S_n(r,c) \text{ is connected whp}$$

History and results

Previous results:

Dubhashi, Johansson, Häggström, Panconesi, and Sozio

$$r = \Theta(1) \quad \Rightarrow \quad S_n(r, 2) \text{ is connected whp}$$

but $G(n, r)$ is then an expander !

Crescenzi, Nocentini, Pietracaprina, Pucci

$$d = 2 \quad \begin{array}{l} r > \sqrt{\log n / n} \\ c > \gamma_2 \log(1/r) \end{array} \Rightarrow \quad S_n(r, c) \text{ is connected whp}$$

but for $r < n^{-\epsilon}$, $c = \Theta(\log n)$!

Connectivity for small visibility radius

Theorem. (B-Devroye-Lugosi-Fraiman) $\mathcal{D} = [0,1]^d$

Suppose γ large enough

Set

$$r \sim \gamma \left(\frac{\log n}{n} \right)^{1/d} \qquad \text{and} \qquad c_n^\star = \sqrt{\frac{2 \log n}{\log \log n}}$$

Then, for any $\epsilon \in (0,1)$

$$\lim_{n\to\infty} \mathbf{P}\left(S_n(r_n, c_n) \text{is connected}\right) = \begin{cases} 0 & \text{if } c_n \leq (1-\epsilon)c_n^\star \\ 1 & \text{if } c_n \geq (1+\epsilon)c_n^\star \end{cases}$$

Connectivity for large visibility radius

Take

$$r_n \sim \gamma n^{-(1-\delta)/d} \qquad \text{with} \qquad \delta \in (0,1)$$

Theorem. There exists a constant $c = c(\delta)$ such that

$$\lim_{n\to\infty} \mathbf{P}\left(S_n(r_n, c) \text{ is connected}\right) = 1$$

Theorem. Define:

$$\underline{c^\star}(\delta) := \sup\left\{c : \lim_{n\to\infty} \mathbf{P}\left(S_n(r_n, c) \text{ is connected}\right) = 0\right\}$$

$$\overline{c^\star}(\delta) := \inf\left\{c : \lim_{n\to\infty} \mathbf{P}\left(S_n(r_n, c) \text{ is connected}\right) = 1\right\}$$

Then, for any $\epsilon \in (0,1)$ and δ small enough,

$$(1-\epsilon)\delta^{-1/2} \le c^\star(\delta) \le (1+\epsilon)\delta^{-1/2}$$

Giant component

Refined model: $(d = 2)$

$(\xi_i)_{i \geq 1}$ i.i.d. copies of a random variable $\xi \geq 1$

Every node i chooses ξ_i neighbours

$\mathscr{C}_1(\cdot)$ the size of the largest connected component

Theorem. Suppose $\mathbf{E}[\xi] > 1$. For every $\delta \in (0,1)$, there exists a $\gamma > 0$ such that, for $r_n \geq \gamma\sqrt{n^{-1}\log n}$

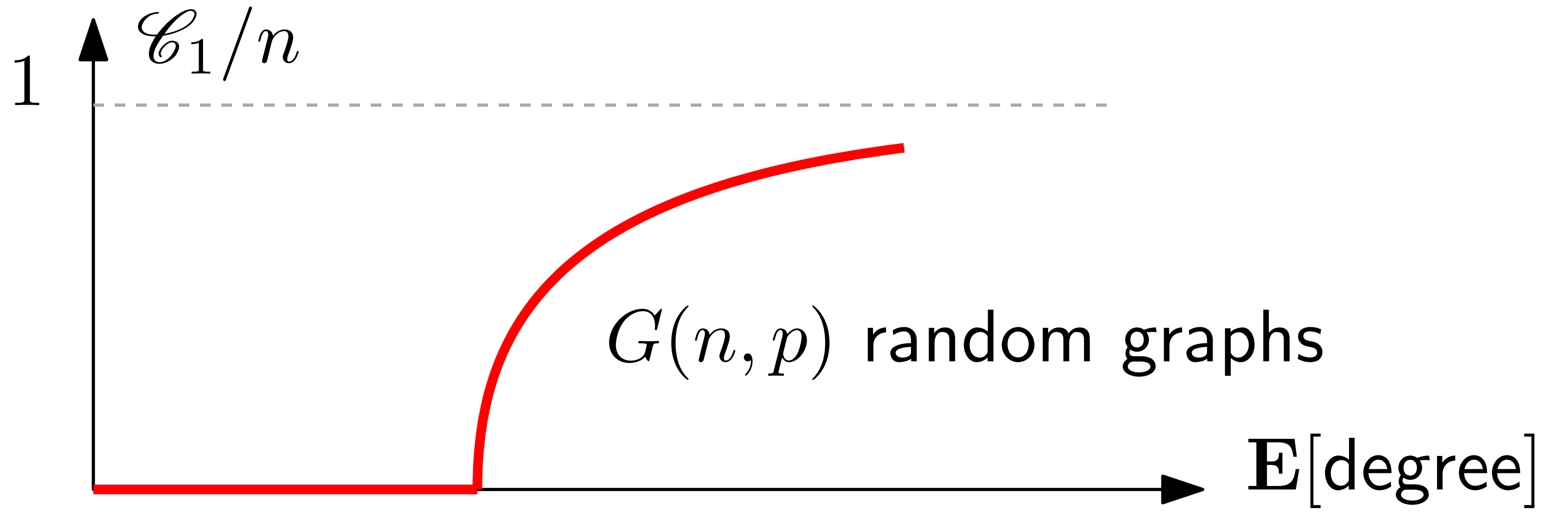
$$\lim_{n\to\infty} \mathbf{P}\left(\mathscr{C}_1(S_n(r_n,\xi)) \geq (1-\delta)n\right) = 1$$

Theorem. Suppose $r_n = o\big((n\log n)^{-1/3}\big)$. Then $\forall\delta > 0$

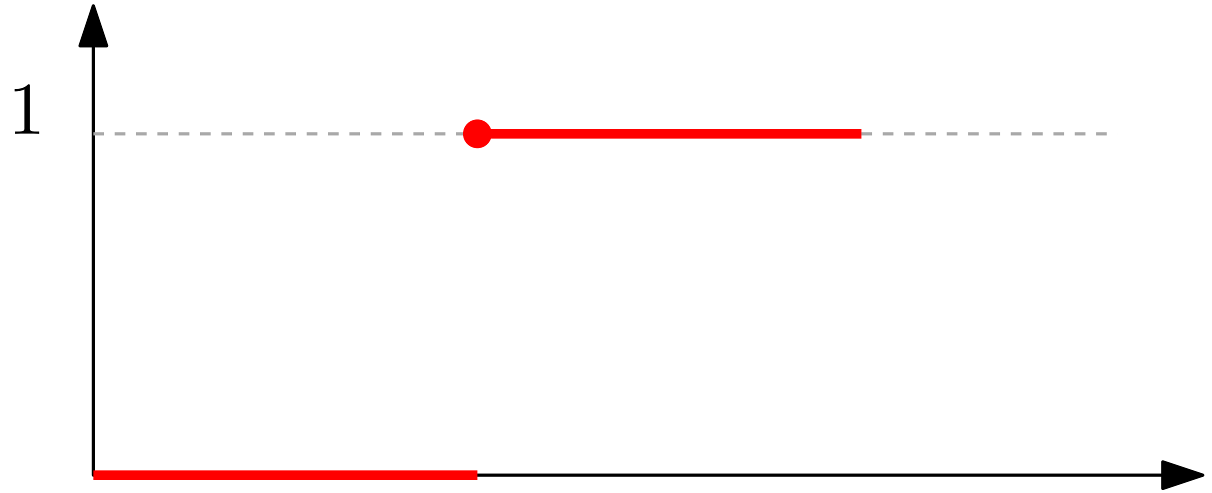
$$\lim_{n\to\infty} \mathbf{P}\left(\mathscr{C}_1(S_n(r_n,1)) \geq \delta n\right) = 0$$

"Explosive" phase transition

Usually: phase transition is continuous



Here: as discontinuous as it can be...



Intuition from the mean-field model

Assume $r_n = \infty$

Theorem. (Fenner–Frieze '82)

$$\lim_{n\to\infty} \mathbf{P}\left(S_n(\infty, 2) \text{ is connected}\right) = 1$$

Theorem. There exists a constant $\delta > 0$ such that

$$\lim_{n\to\infty} \mathbf{P}\left(\mathscr{C}_1(S_n(\infty, 1)) \geq \delta n\right) > 0$$

Intuition from the mean-field model

Assume $r_n = \infty$

Theorem. (Fenner–Frieze '82)

$$\lim_{n\to\infty} \mathbf{P}\left(S_n(\infty, 2) \text{ is connected}\right) = 1$$

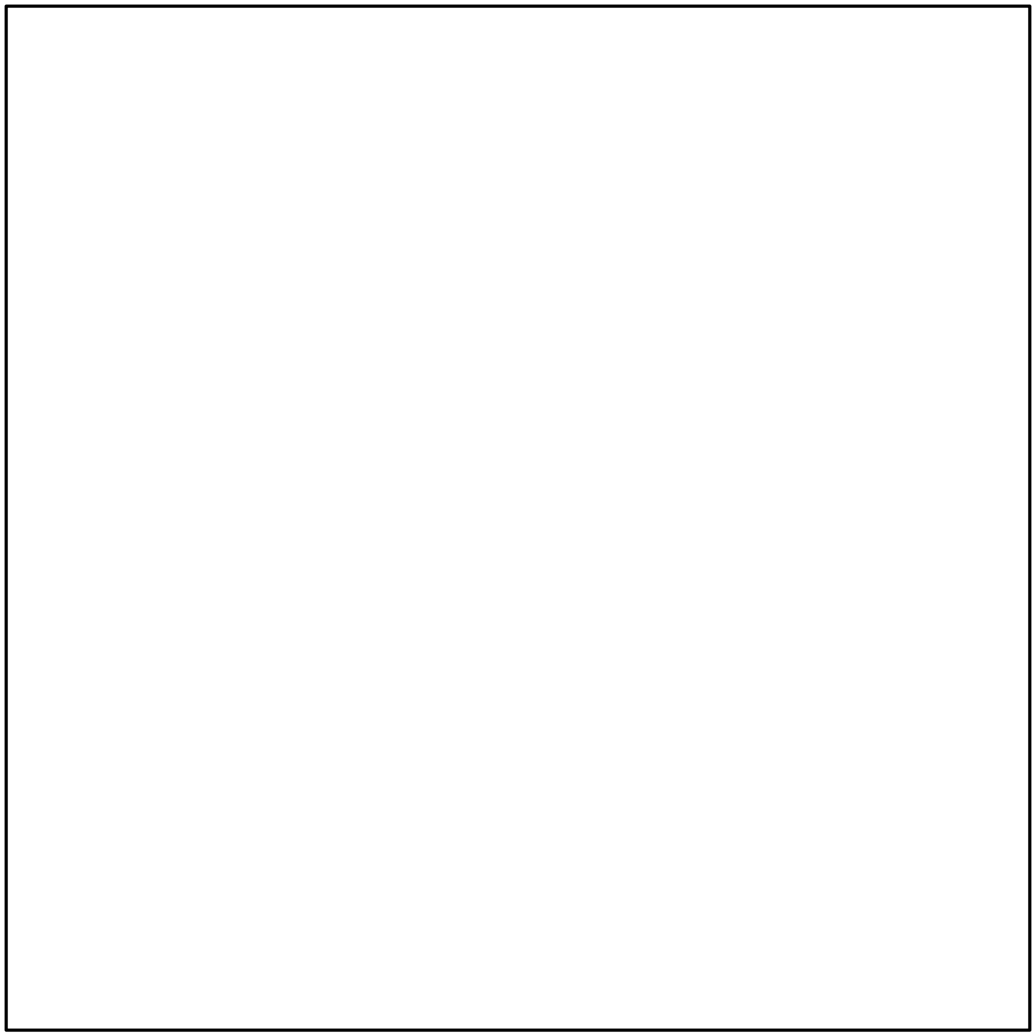
Theorem. There exists a constant $\delta > 0$ such that

$$\lim_{n\to\infty} \mathbf{P}\left(\mathscr{C}_1(S_n(\infty, 1)) \geq \delta n\right) > 0$$

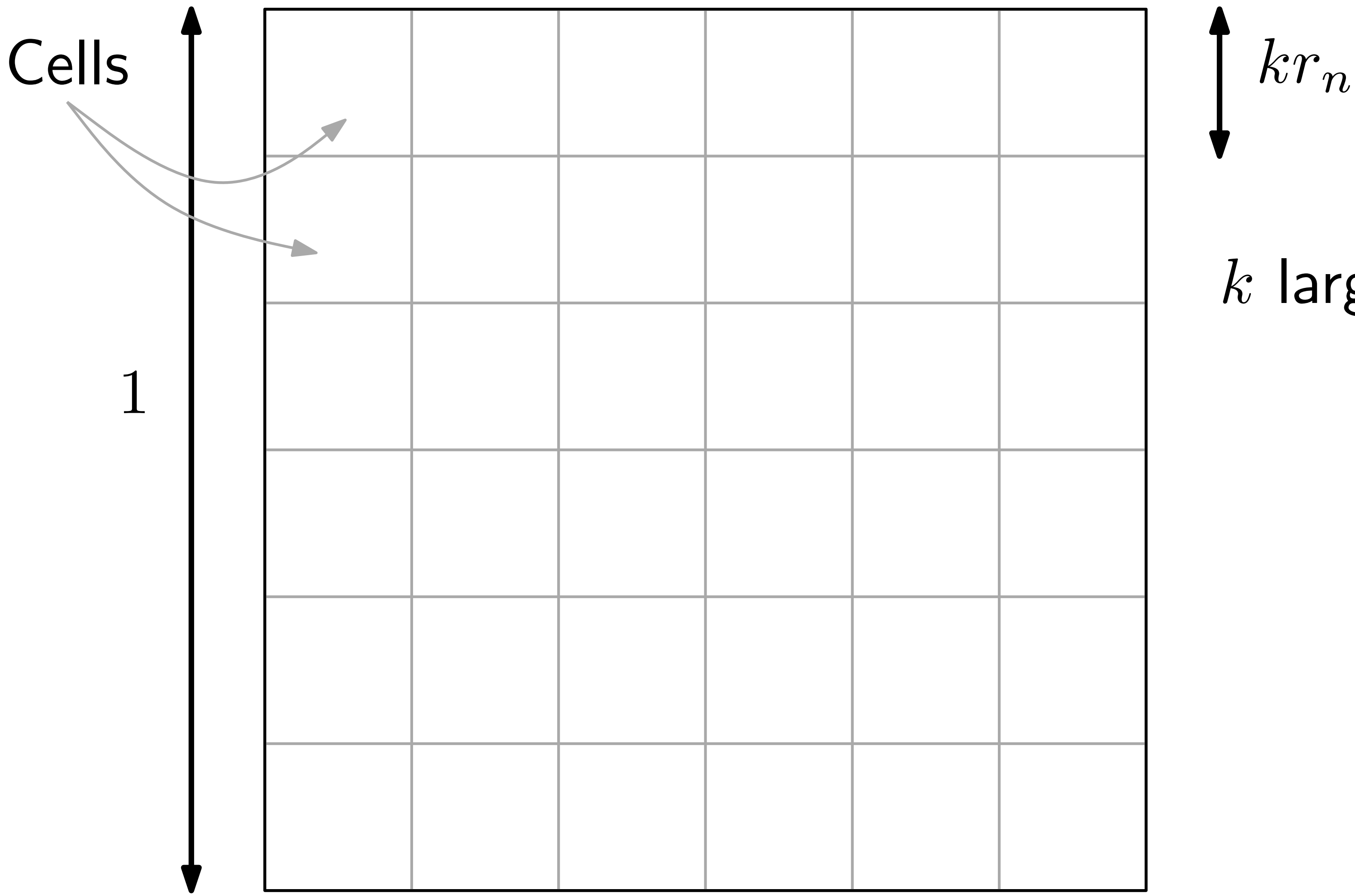
Idea: $S_n(\infty, 1)$ is a random mapping

$S_n(\infty, 2)$ is the union of 2 random mappings

Discretization I

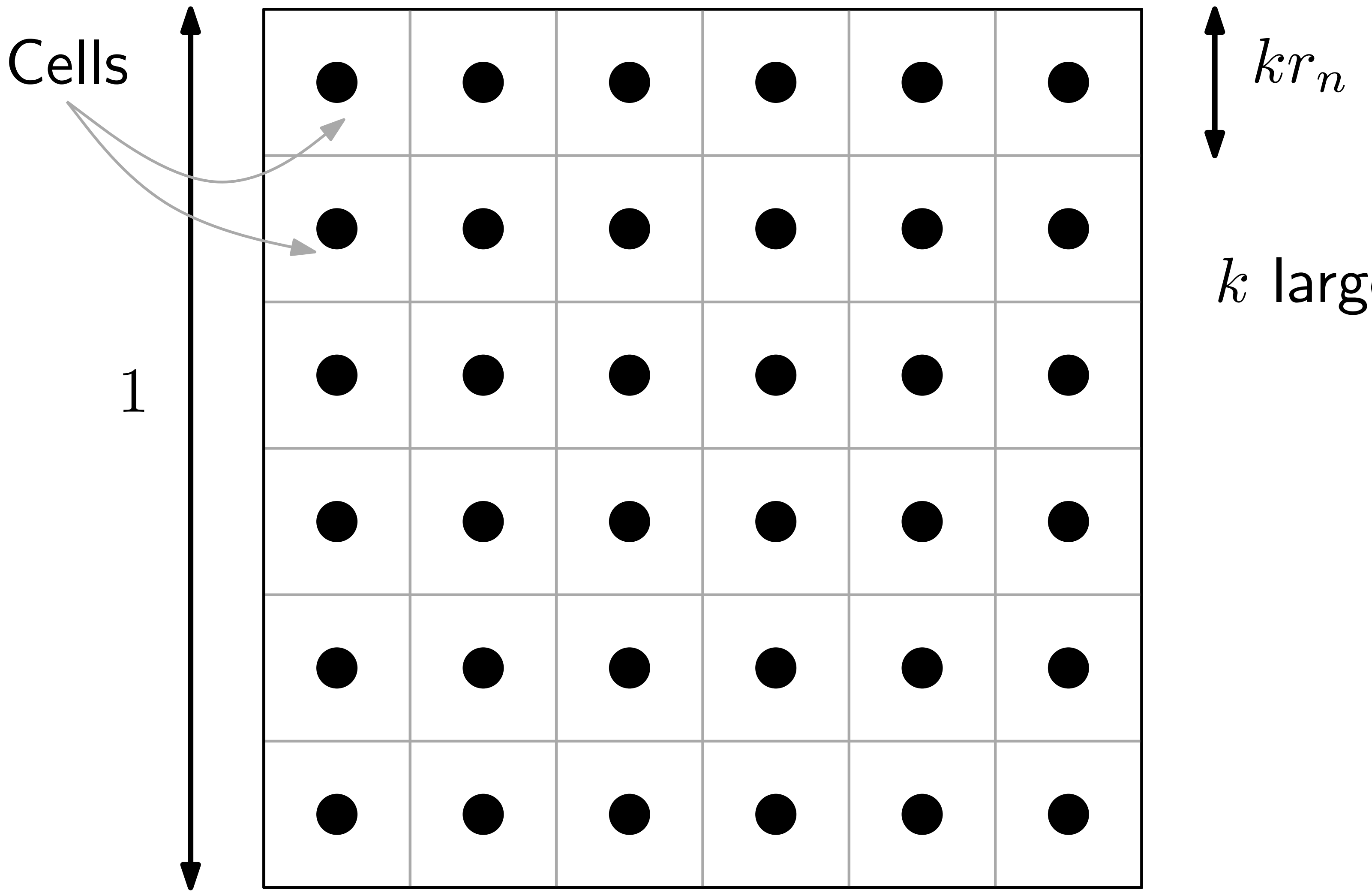


Discretization I



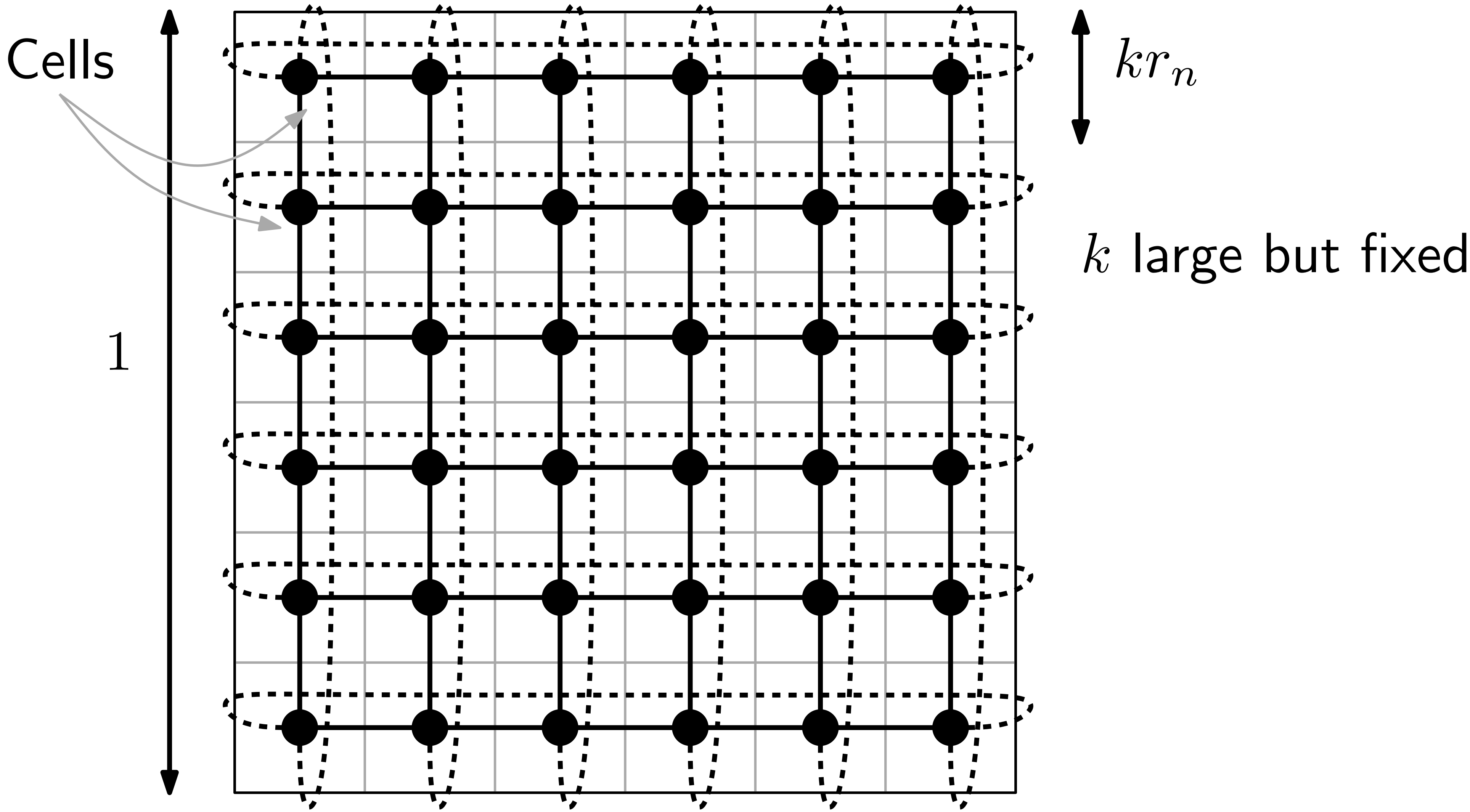
k large but fixed

Discretization I



k large but fixed

Discretization I



Associate : events to nodes/cells
events to bonds/faces } percolation

Uniformity of the point set

Definition. A cell is δ-good if for all boxes B

$$(1-\delta)\frac{nr_n^2}{4d^2} \leq |B \cap \mathbf{X}| \leq (1+\delta)\frac{nr_n^2}{4d^2}$$

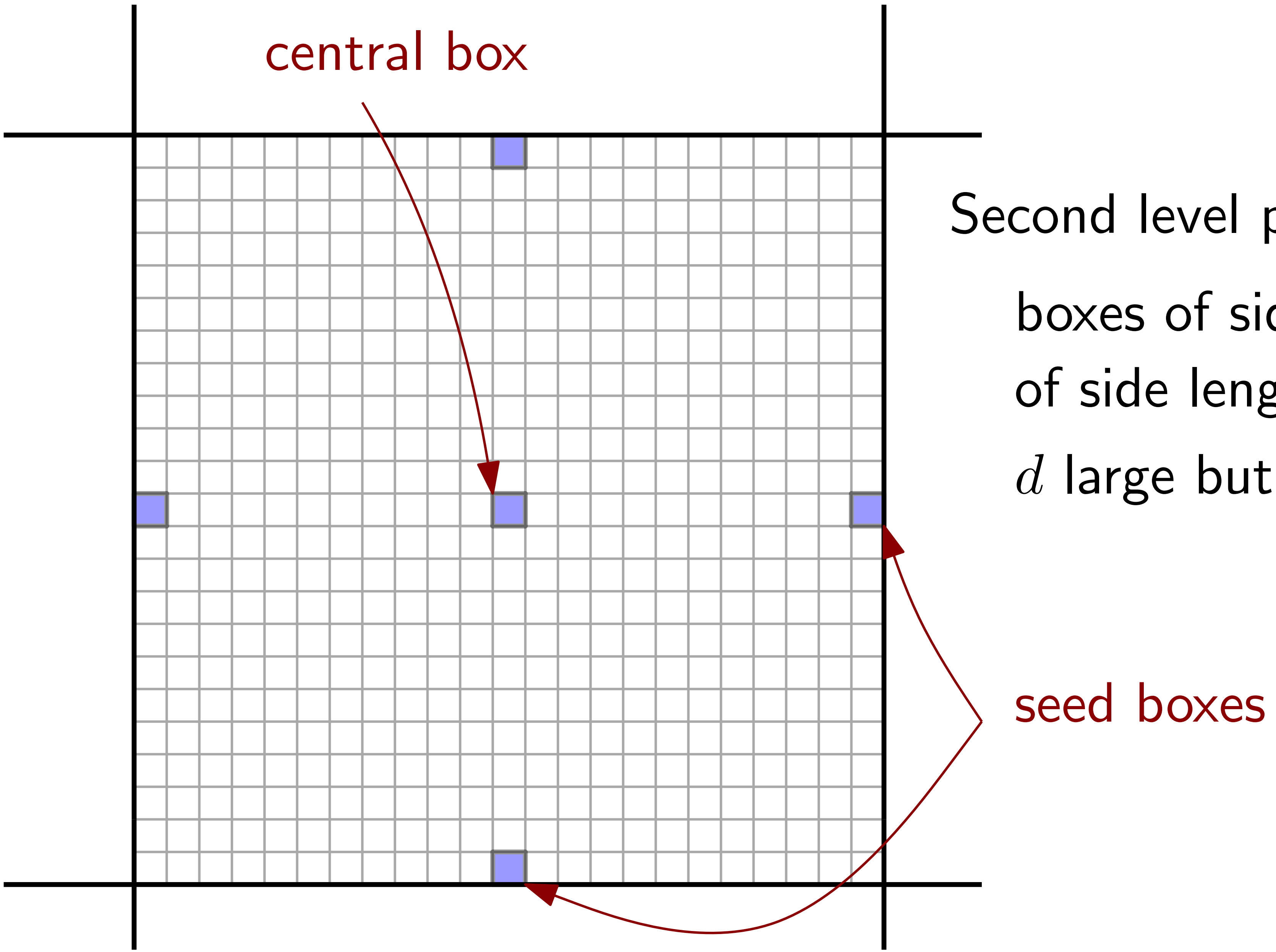
Lemma. Suppose $r_n \geq \gamma\sqrt{\log n/n}$. For any $\delta > 0$, there exists $\gamma_0 > 0$ such that if $\gamma > \gamma_0$

$$\lim_{n\to\infty} \mathbf{P}\left(\text{every cell is } \delta\text{-good}\right) = 1$$

Lemma. $\exists \alpha, \beta > 0$ such that

$$\lim_{n\to\infty} \mathbf{P}\left(\forall x : \alpha n r_n^2 \leq |B(x, r_n) \cap \mathbf{X}| \leq \beta n r_n^2\right) = 1$$

Node/Cell events



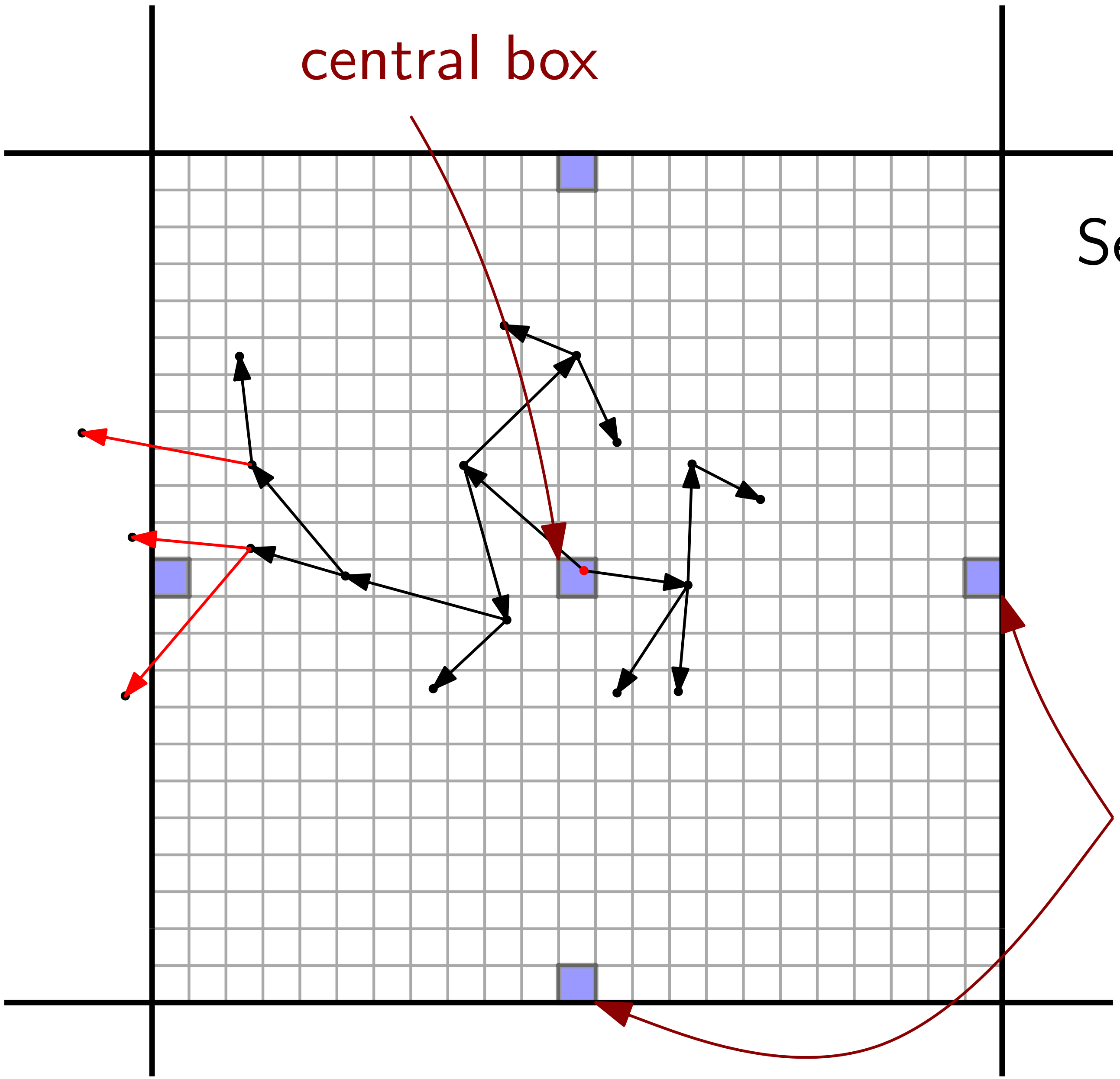
Second level partition:

boxes of side length

of side length $r/(2d)$

d large but fixed

Node/Cell events



Second level partition:

boxes of side length

of side length $r/(2d)$

d large but fixed

Node events II

Exploration: in a given cell

Start from a point $x \in \mathbf{X}$ in the central box

Explore the neighborhoods for k^2 generations

Kill the paths that leave the cell

Define:

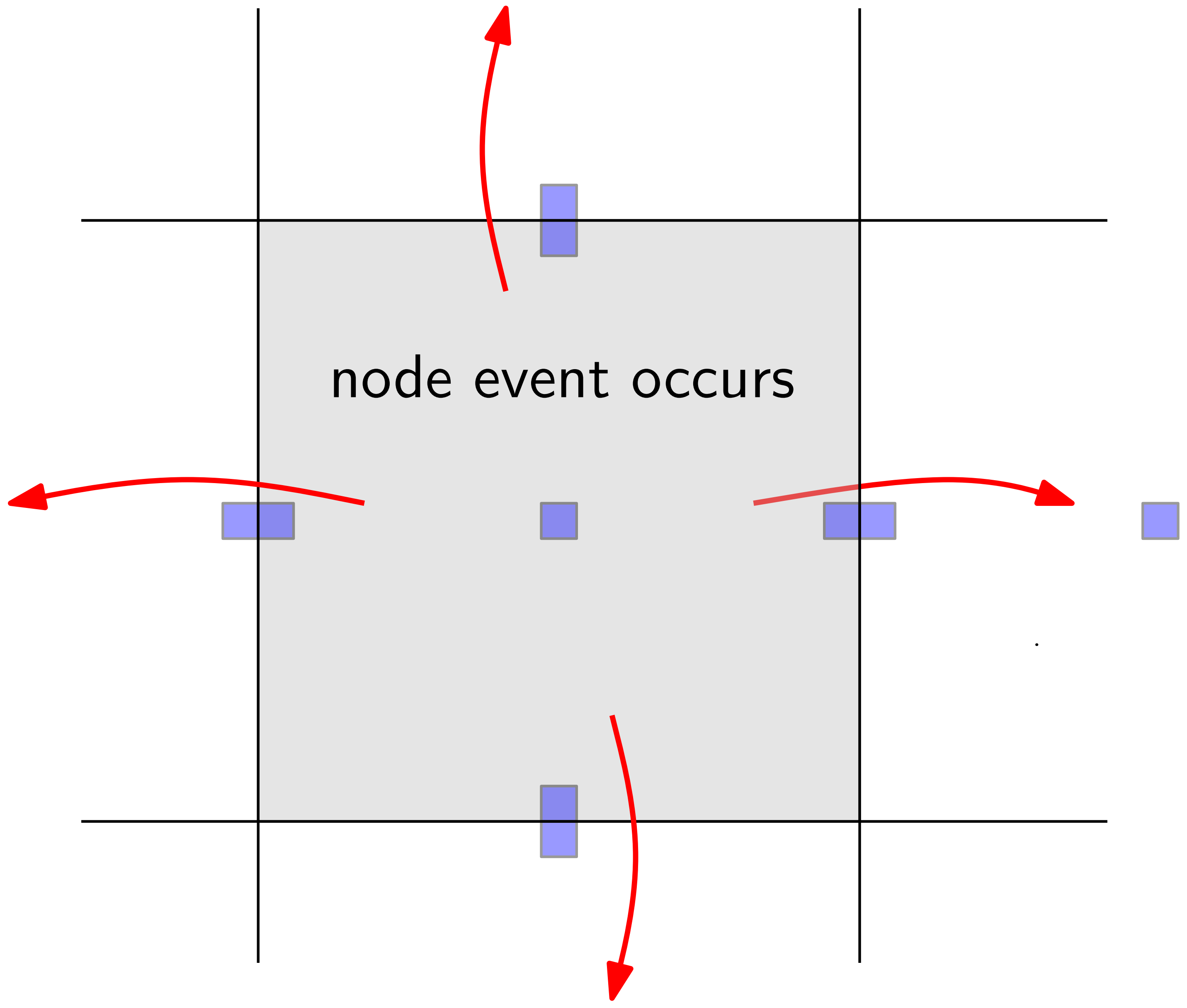
$\Delta_x(i)$ the collection of points at generation i

$$G_x = \{\forall \text{ box } B, |\Delta_x(k^2) \cap B| \geq \mathbf{E}[\xi]^{k^2/2}\}$$

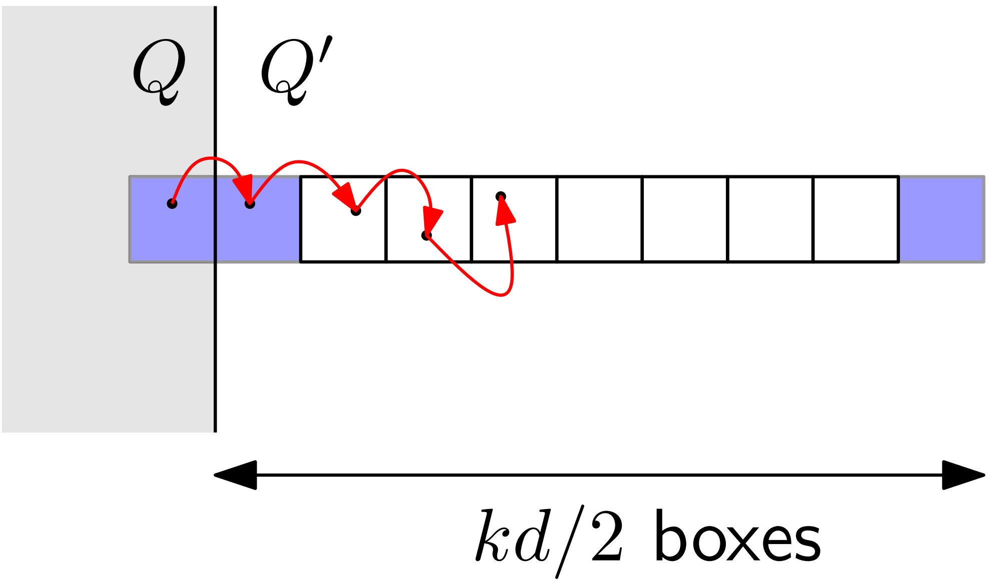
Lemma. $\forall \eta > 0$, can choose all the constants such that for all n large

$$\mathbf{P}(G_x \mid \mathbf{X}, \delta\text{-good}) \geq 1 - \eta$$

Link events



Link events II



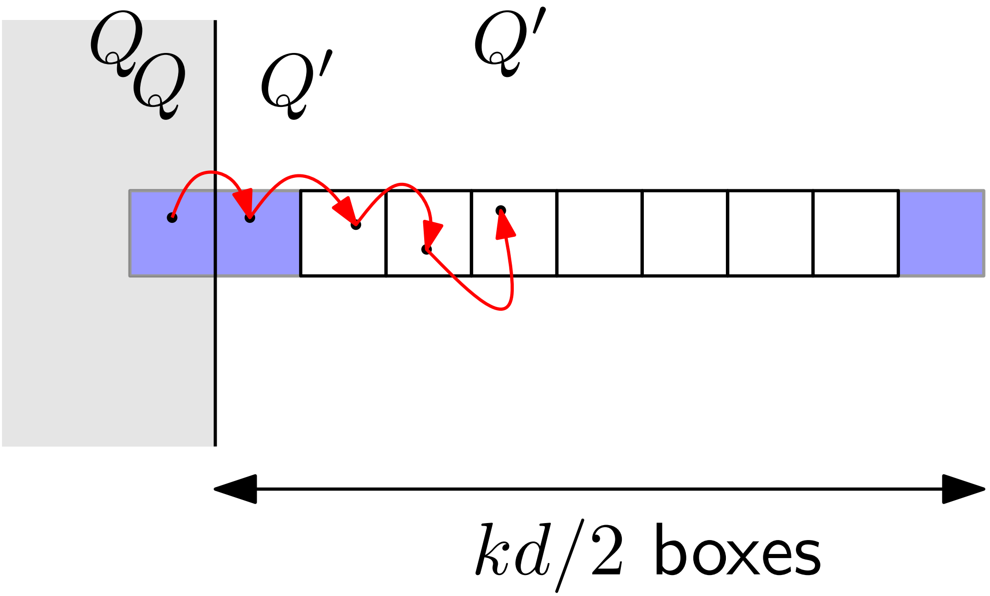
$J_x(Q, Q')$ the sequence of first neighbors of x reaches the central box in a ladder fashion

$J_R = \cap_{x \in R} J_x$, with R the population of the seed box in Q

Lemma.

$$\mathbf{P}(\, J_R(Q, Q') \mid \mathbf{X}, R) \geq 1 - \exp\left(\frac{|R|}{(10\beta d^2)^{kd}} \right)$$

Link events II



Recall $|R| \geq \mathbf{E}\left[\xi\right]^{k^2/2}$ if node event occurs in Q

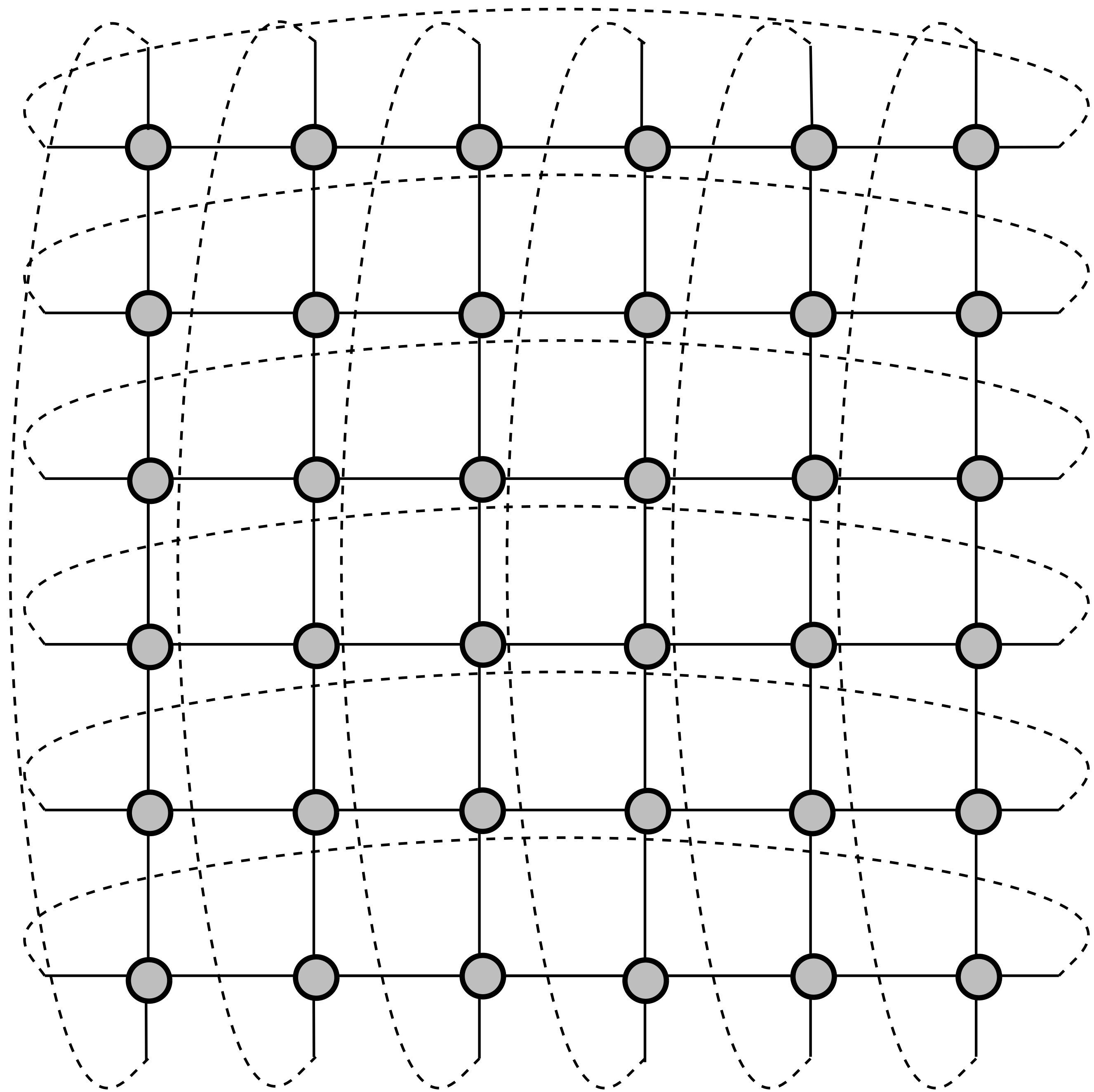
$J_x(Q, Q')$ the sequence of first neighbors of x reaches the central box in a ladder fashion

$J_R = \cap_{x \in R} J_x$, with R the population of the seed box in Q

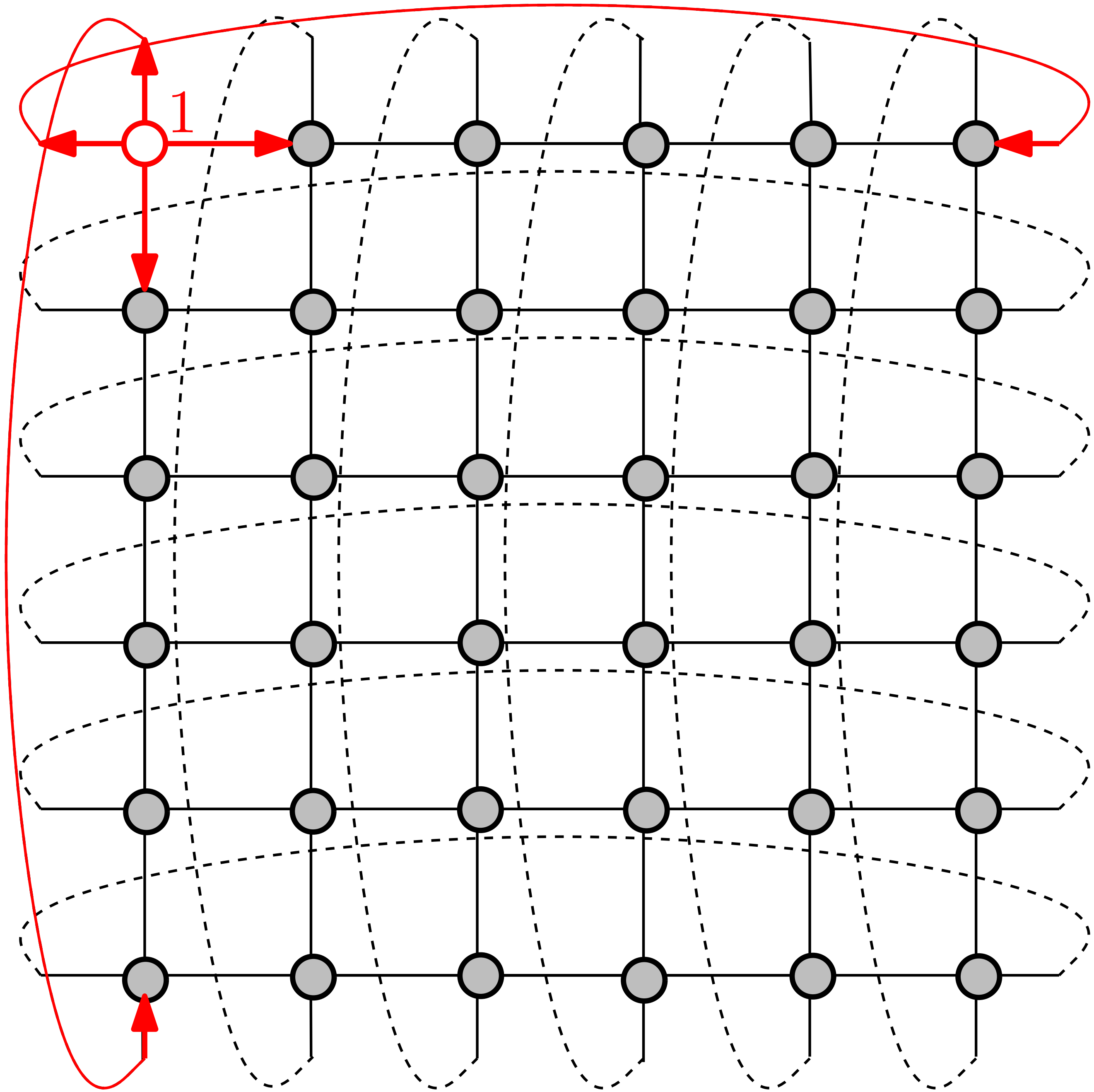
Lemma.

$$\mathbf{P}(\, J_R(Q, Q') \mid \mathbf{X}, R) \geq 1 - \exp\left(\frac{|R|}{(10\beta d^2)^{kd}} \right)$$

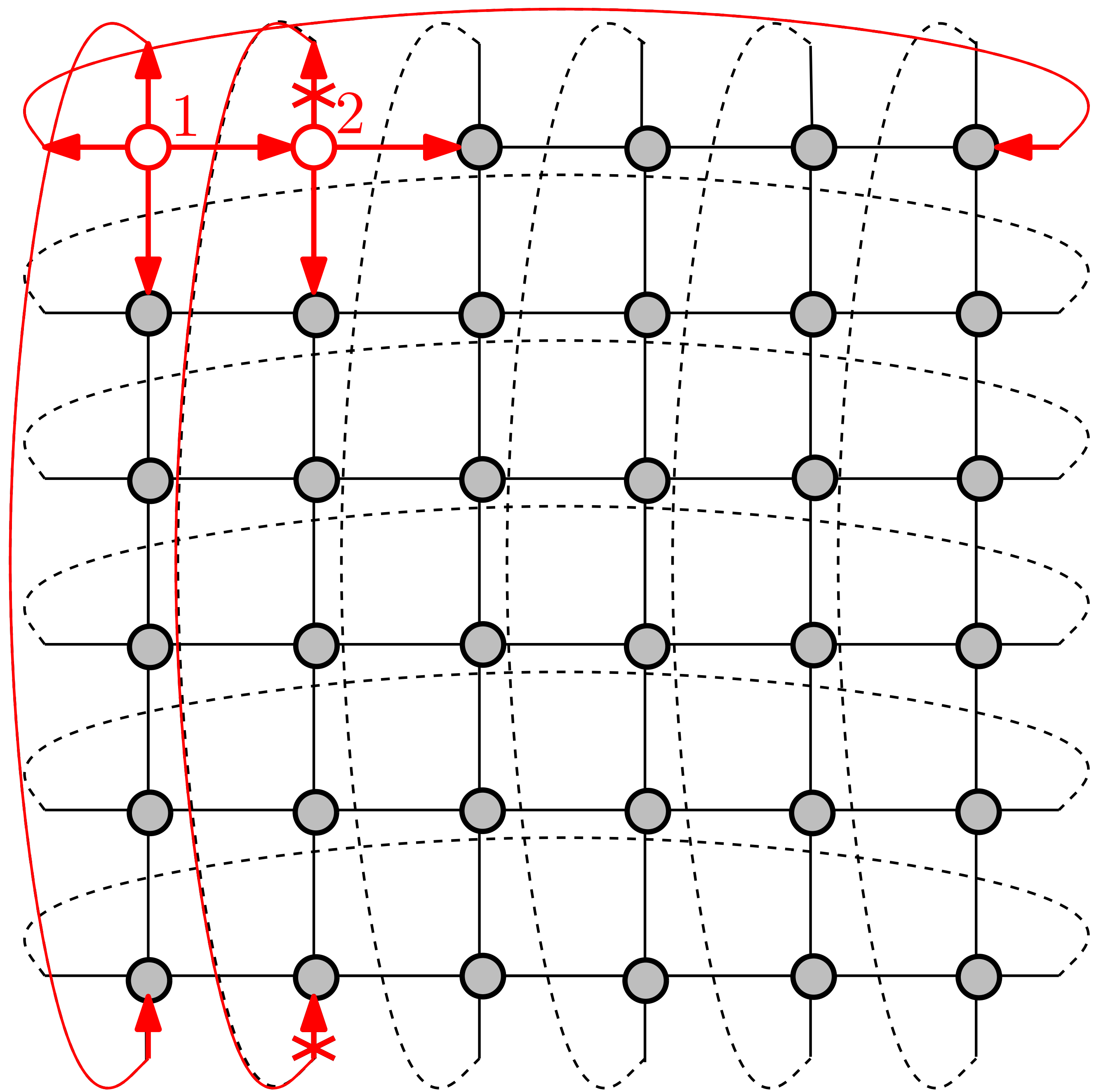
Coupling with a (site/bond) percolation process



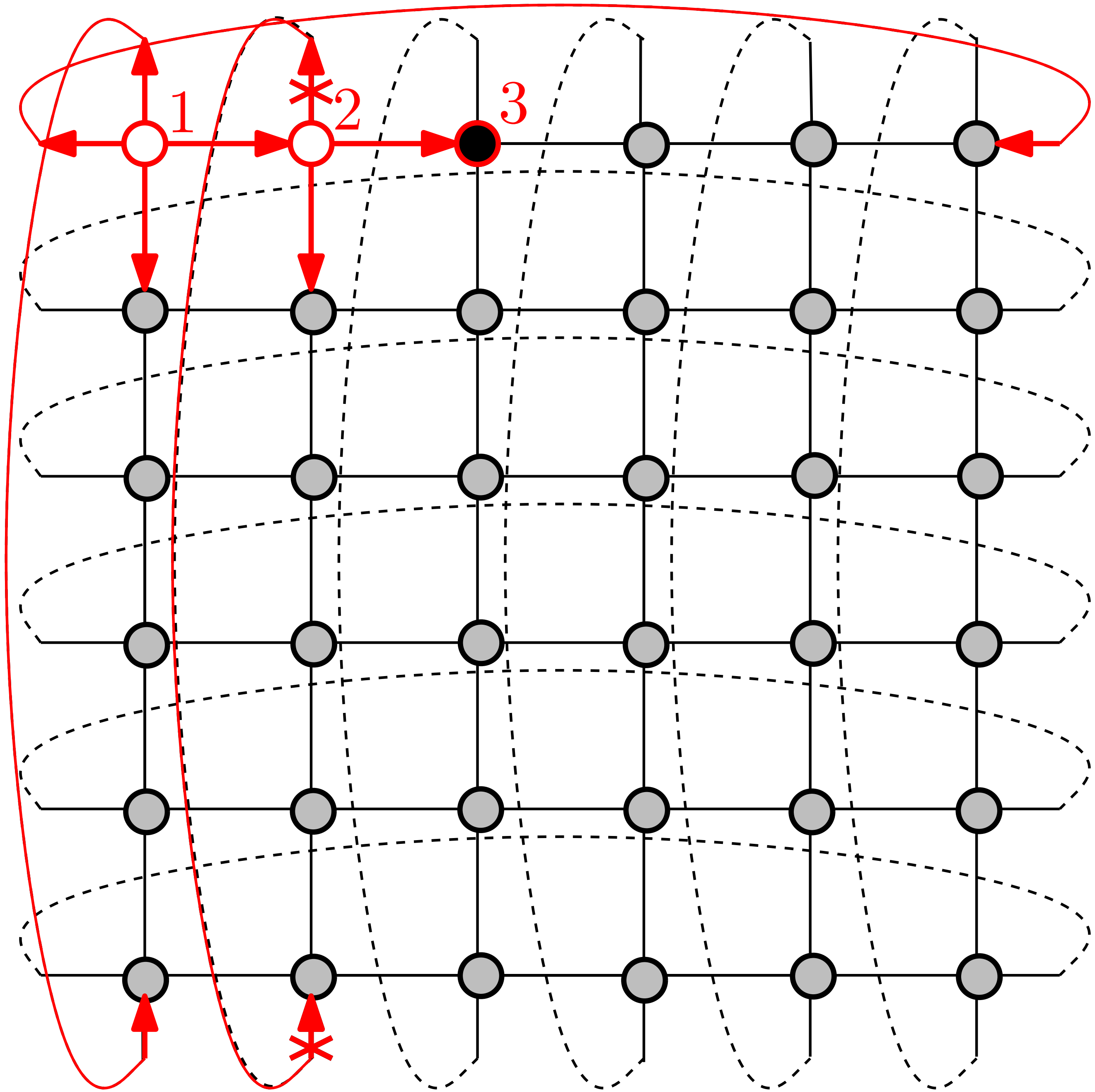
Coupling with a (site/bond) percolation process



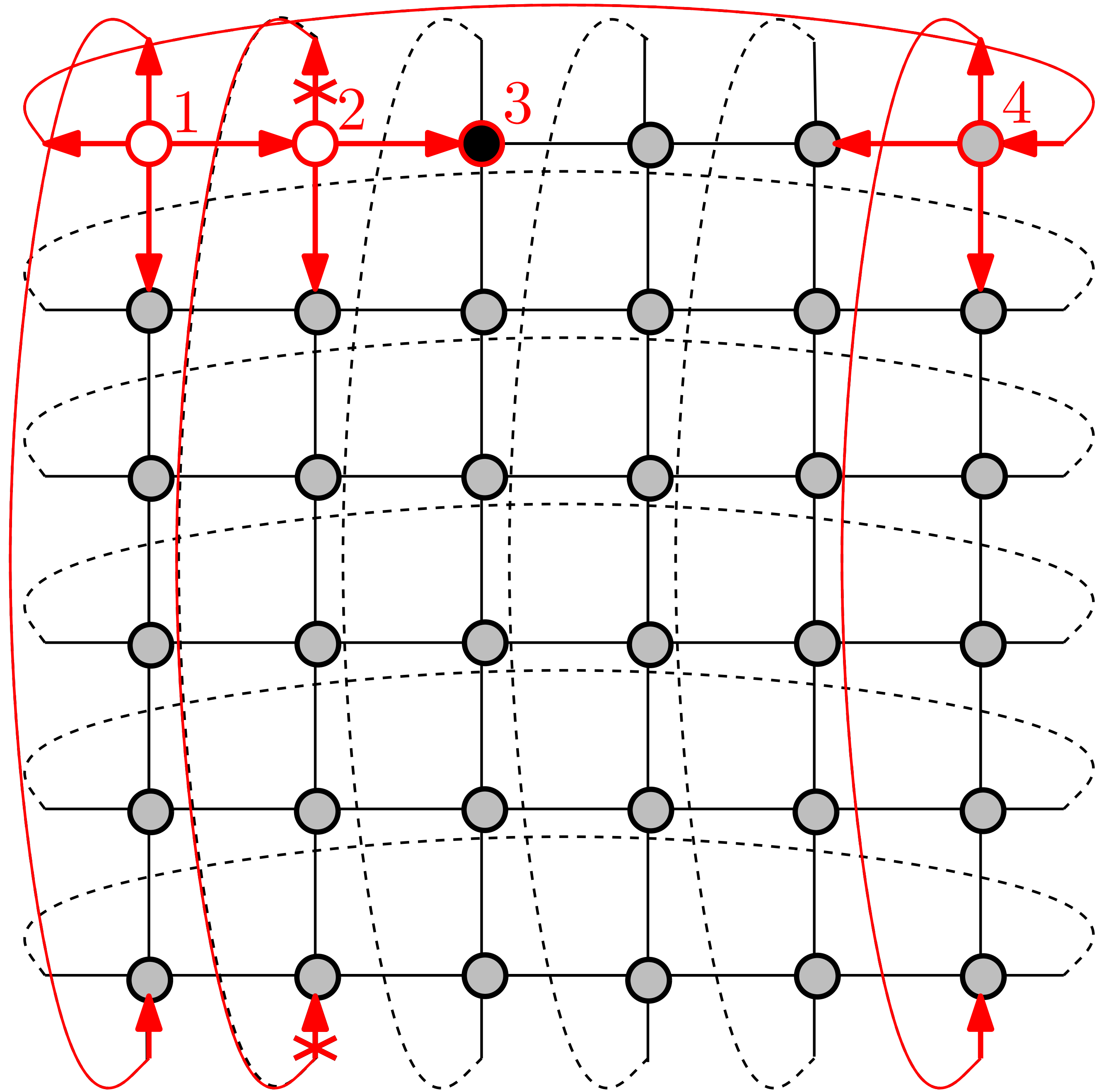
Coupling with a (site/bond) percolation process



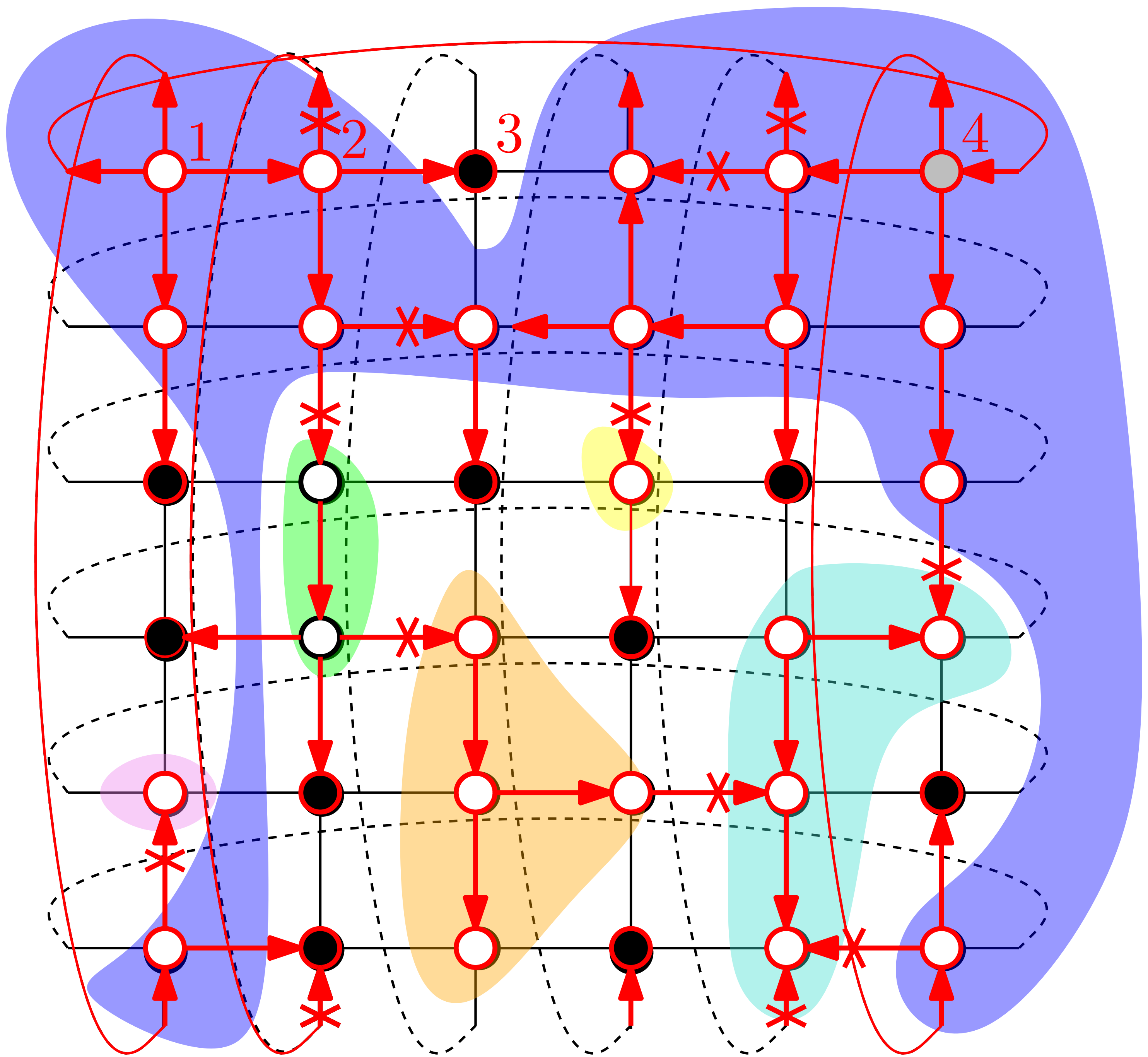
Coupling with a (site/bond) percolation process



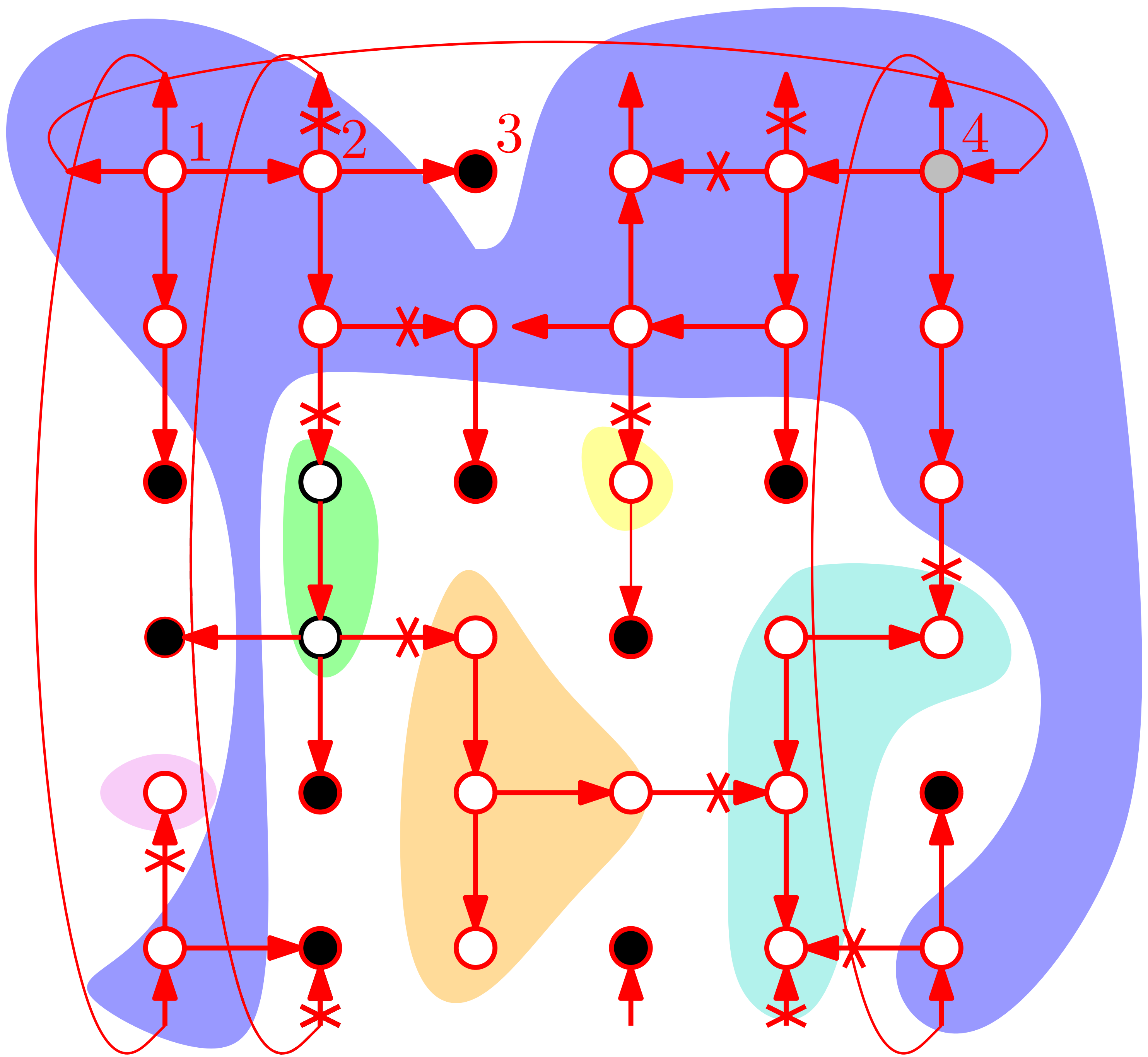
Coupling with a (site/bond) percolation process



Coupling with a (site/bond) percolation process



Coupling with a (site/bond) percolation process



Coupling with a (site/bond) percolation process

percolation process = random configuration, where

each edge open with probability p_e } independently !
each node open with probability p_n

Definition of the coupling:

1. Exploration of clusters $\Longrightarrow$ partial percolation process on torus $[m] \times [m]$

2. For unassigned values: complete with i.i.d. Bernoulli

Observation. If H is a cluster of cells in the percolation process, then, there is a connected component of $S_n(r_n, \xi)$ which "sees" every single point of S

Cluster built is ubiquitous but too sparse!

Lemma. For any $\epsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}\left(\exists H : |H| \geq (1-\epsilon)m^2\right) = 1$$

Problems:

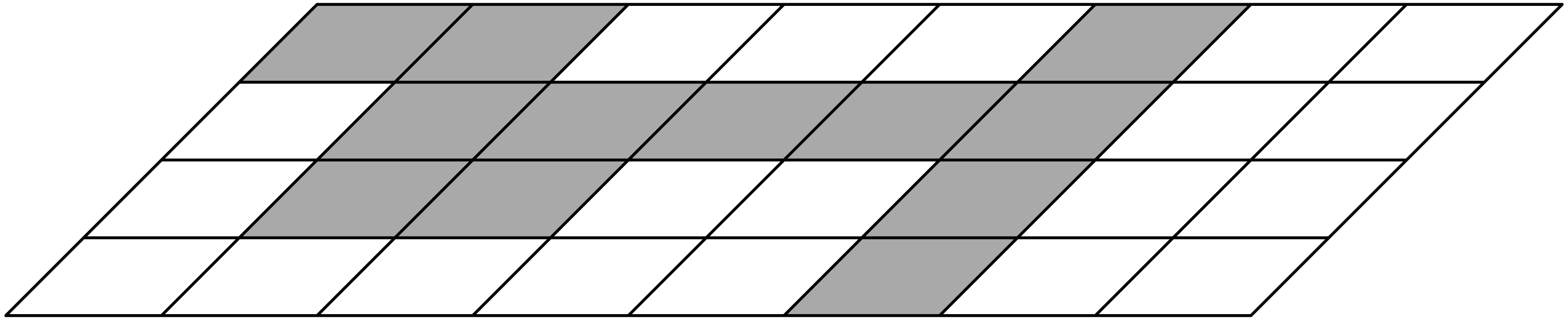
1) in $S_n(r_n, \xi)$ only constant number of points per cell so only $O(m^2) = O(n/\log n)$ nodes in total !

2) close points do not connect locally !

- after ℓ generations about $\mathbf{E}\left[\xi\right]^{\ell}$ points
- need $\ell = \log\log n$ to have a proportion in a cell
- but these $\Omega(\log n)$ are spread at distance $\Omega(\sqrt{\log\log n})$

Finale: gathering the remaining points

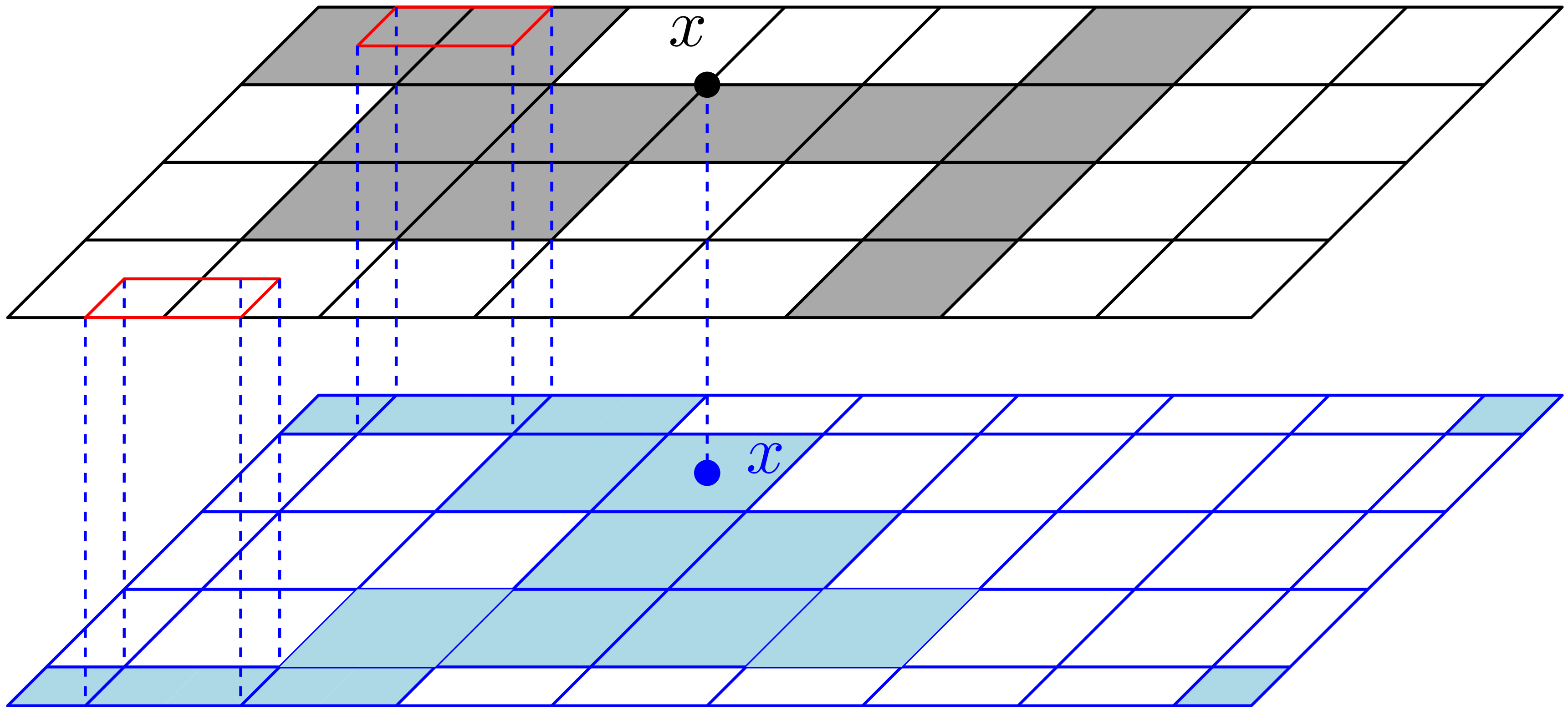
Only "used up" a vanishing proportion of nodes per cell
$\Rightarrow$ can play the same game again !



Finale: gathering the remaining points

Only "used up" a vanishing proportion of nodes per cell
$\Rightarrow$ can play the same game again !

Consider $x \in \mathbf{X}$: make its box central in the discretization



Cells may not be aligned, but boxes are aligned

Hooking up left over points

Observations.

1) $\mathbf{P}(x \text{ in second giant}) \geq 1-\epsilon$
2) the two giants overlap on $(1-2\epsilon)$ proportion of the boxes
   $\Rightarrow \Omega(n/\log n)$ such boxes

The too giants are bound to hook up !

in each of the overlapping boxes: about $\mathbf{E}[\xi]^{k^2/2}$ points whose neighbors have not yet been chosen !

$\mathbf{P}(\text{one node fails to hook}) \leq (1 - O(1/\log n))$

$\mathbf{P}(\text{all node fail to hook}) = (1-\Omega(1/\log n))^{n/\log n} = e^{-\Omega(n)}$