# Connectivity and giant component in random irrigation graphs

Gábor Lugosi

ICREA and Pompeu Fabra University, Barcleona

joint work with
Nicolas Broutin (INRIA)
Luc Devroye (McGill)

# irrigation graphs

Start with a connected graph on $n$ vertices.

# irrigation graphs

Start with a connected graph on $n$ vertices.

An irrigation subgraph is obtained when each vertex selects $c$ neighbors at random (without replacement).

# irrigation graphs

Start with a connected graph on $n$ vertices.

An irrigation subgraph is obtained when each vertex selects $c$ neighbors at random (without replacement).

Here the underlying graph is a random geometric graph: vertices are $X_1, \ldots, X_n$ i.i.d. uniform on the torus $[0,1]^2$ and $X_i \sim X_j$ iff $\|X_i - X_j\| < r$.

# irrigation graphs

Start with a connected graph on $n$ vertices.

An irrigation subgraph is obtained when each vertex selects $c$ neighbors at random (without replacement).

Here the underlying graph is a random geometric graph: vertices are $X_1, \ldots, X_n$ i.i.d. uniform on the torus $[0,1]^2$ and $X_i \sim X_j$ iff $\|X_i - X_j\| < r$.

Also called bluetooth graphs.
They are locally sparsified random geometric graphs.

# irrigation graphs

Start with a connected graph on $n$ vertices.

An irrigation subgraph is obtained when each vertex selects $c$ neighbors at random (without replacement).

Here the underlying graph is a random geometric graph: vertices are $X_1, \dots, X_n$ i.i.d. uniform on the torus $[0,1]^2$ and $X_i \sim X_j$ iff $\|X_i - X_j\| < r$.

Also called bluetooth graphs.
They are locally sparsified random geometric graphs.

Introduced by Ferraguto, Mambrini, Panconesi, and Petrioli (2004).

Fenner and Frieze (1982) considered $K_n$ as the underlying graph.

# irrigation graphs

Start with a connected graph on $n$ vertices.

An irrigation subgraph is obtained when each vertex selects $c$ neighbors at random (without replacement).

Here the underlying graph is a random geometric graph: vertices are $X_1, \ldots, X_n$ i.i.d. uniform on the torus $[0,1]^2$ and $X_i \sim X_j$ iff $\|X_i - X_j\| < r$.

Also called bluetooth graphs.
They are locally sparsified random geometric graphs.

Introduced by Ferraguto, Mambrini, Panconesi, and Petrioli (2004).

Fenner and Frieze (1982) considered $K_n$ as the underlying graph.

Question: How large does $c$ need to be for $G(n, r, c)$ to be connected?

# irrigation graphs

$\boldsymbol{G(n,r)}$ needs to be connected.

# irrigation graphs

$G(n, r)$ needs to be connected.

Connectivity threshold is $r^* \sim \sqrt{\log n / (\pi n)}$.

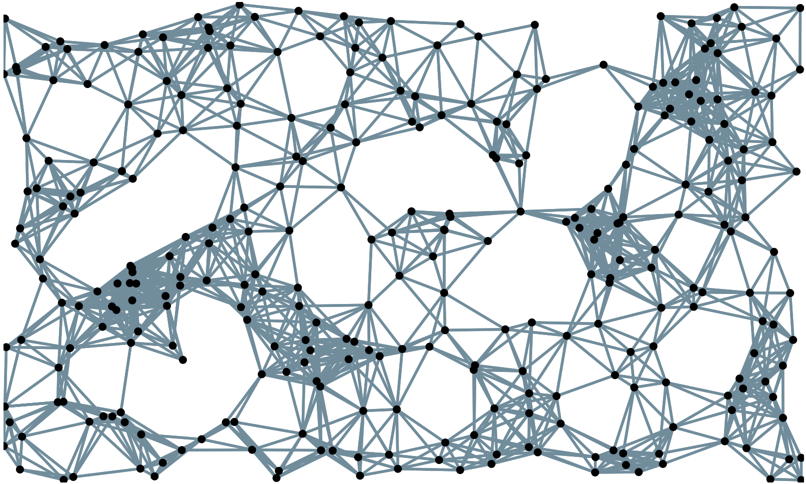
# irrigation graphs

$G(n, r)$ needs to be connected.

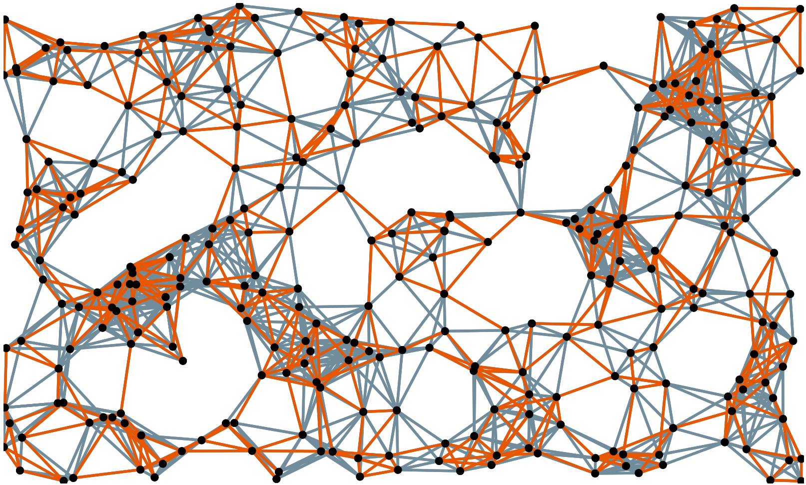
Connectivity threshold is $r^* \sim \sqrt{\log n/(\pi n)}$.

We only consider $r \geq \gamma\sqrt{\log n/n}$ for a sufficiently large $\gamma$.

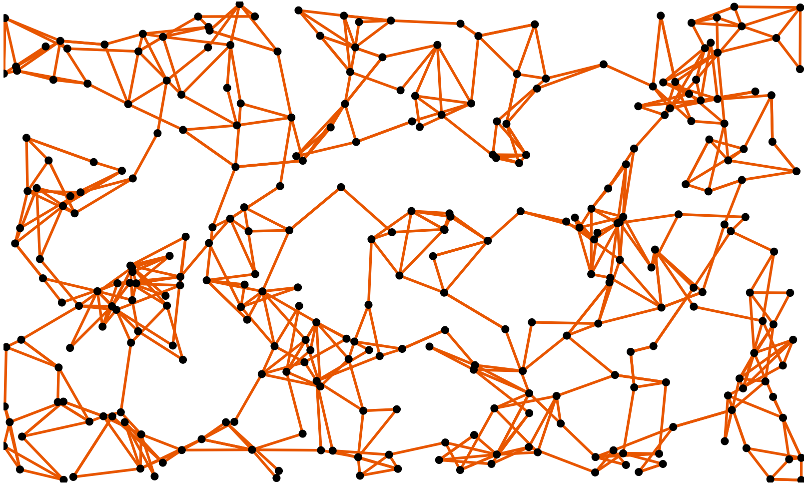
irrigation graphs



# irrigation graphs



# irrigation graphs



# previous results

Fenner and Frieze, 1982: For $r = \infty$, $G(n, r, 2)$ is connected whp.

# previous results

Fenner and Frieze, 1982: For $r = \infty$, $G(n, r, 2)$ is connected whp.

Dubhashi, Johansson, Häggström, Panconesi, Sozio, 2007: For constant $r$ the graph $G(n, r, 2)$ is connected whp.

# previous results

Fenner and Frieze, 1982: For $r = \infty$, $G(n, r, 2)$ is connected whp.

Dubhashi, Johansson, Häggström, Panconesi, Sozio, 2007: For constant $r$ the graph $G(n, r, 2)$ is connected whp.

Crescenzi, Nocentini, Pietracaprina, Pucci, 2009: $\exists\ \alpha, \beta$ such that if

$$r \geq \alpha \sqrt{\frac{\log n}{n}} \qquad \text{and} \qquad c \geq \beta \log(1/r),$$

then $G(n, r, c)$ is connected whp.

This bound is sub-optimal in all ranges of $r$.

# previous results

Broutin, Devroye, Fraiman, and Lugosi, 2012:
There exists a constant $\gamma^* > 0$ such that for all $\gamma \geq \gamma^*$ and $\epsilon \in (0, 1)$, if

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

# previous results

Broutin, Devroye, Fraiman, and Lugosi, 2012:
There exists a constant $\gamma^* > 0$ such that for all $\gamma \geq \gamma^*$ and $\epsilon \in (0,1)$, if

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

then

- if $c \geq (1+\epsilon)c_t$ then $G(n,r,c)$ is connected whp.
- if $c \leq (1-\epsilon)c_t$ then $G(n,r,c)$ is disconnected whp.

# previous results

Broutin, Devroye, Fraiman, and Lugosi, 2012:
There exists a constant $\gamma^* > 0$ such that for all $\gamma \geq \gamma^*$ and $\epsilon \in (0,1)$, if

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

then

- if $c \geq (1+\epsilon)c_t$ then $G(n,r,c)$ is connected whp.
- if $c \leq (1-\epsilon)c_t$ then $G(n,r,c)$ is disconnected whp.

$c_t$ does not depend on $\gamma$ or $d$.

We get a significantly sparser graph while preserving connectivity.

In this talk we investigate genuinely sparse graphs with $c$ constant.

# sparse connectivity by enlarging $r$

The lower bound follows from a more general result:

# sparse connectivity by enlarging $r$

The lower bound follows from a more general result:

Let $\epsilon \in (0, 1)$ and $\lambda \in [1, \infty]$ be such that $r > \gamma^* \left(\frac{\log n}{n}\right)^{1/2}$

$$\frac{\log nr^2}{\log \log n} \to \lambda \quad \text{and} \quad c \leq (1 - \epsilon)\sqrt{\left(\frac{\lambda}{\lambda - 1/2}\right) \frac{\log n}{\log nr^2}}.$$

Then $G(n, r, c)$ is disconnected whp.

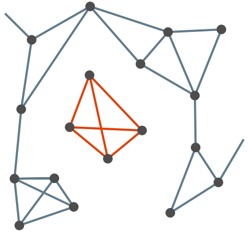
# sparse connectivity by enlarging $r$

In particular, take $r \sim n^{-(1-\delta)/2}$. Then for $c \leq (1-\epsilon)/\sqrt{\delta}$ (constant) the graph is disconnected.

# sparse connectivity by enlarging $r$

In particular, take $r \sim n^{-(1-\delta)/2}$. Then for $c \leq (1-\epsilon)/\sqrt{\delta}$ (constant) the graph is disconnected.

The smallest possible components are cliques of size $c+1$. These appear whp.



# connectivity for constant $c$

The lower bound is not far from the truth: when $r \sim n^{-(1-\delta)/2}$, $c = \sqrt{(1+o(1))/\delta} + const.$ is sufficient for connectivity.

The irrigation graph is connected but genuinely sparse:

# connectivity for constant $c$

The lower bound is not far from the truth: when $r \sim n^{-(1-\delta)/2}$, $c = \sqrt{(1+o(1))/\delta} + const.$ is sufficient for connectivity.

The irrigation graph is connected but genuinely sparse:

Let $\delta \in (0,1)$, $\gamma > 0$. Suppose that $r \sim \gamma n^{-(1-\delta)/2}$. There exists a constant such that $G(n,r,c)$ is connected whp. One may take $c = c_1 + c_2 + c_3 + 1$, where

$$c_1 = \left\lceil \sqrt{(1+\epsilon)/\delta} \right\rceil ,$$

and $c_2, c_3$ are absolute constants.

# Supercritical $r$, constant $c$

Sketch of proof:

• First show that $X_1, \ldots, X_n$ are sufficiently regular whp. Once the $X_i$ are fixed, randomness comes from the edge choices only.

# Supercritical $r$, constant $c$

Sketch of proof:

- First show that $X_1, \ldots, X_n$ are sufficiently regular whp. Once the $X_i$ are fixed, randomness comes from the edge choices only.
- Partition $[0,1]^2$ into congruent squares of side length $1/(2\sqrt{2}r)$

# Supercritical $r$, constant $c$

Sketch of proof:

• First show that $X_1, \ldots, X_n$ are sufficiently regular whp. Once the $X_i$ are fixed, randomness comes from the edge choices only.

• Partition $[0,1]^2$ into congruent squares of side length $1/(2\sqrt{2}r)$

• We add edges in four phases. In the first we start from $X_1$, and using $c_1$ choices of each vertex, we go for $\delta^2 \log_{c_1} n$ generations. There exists a cube in the grid that contains a connected component of size $n^{\text{const.}\delta^2}$.

# Supercritical $r$, constant $c$

Sketch of proof:

• First show that $X_1, \ldots, X_n$ are sufficiently regular whp. Once the $X_i$ are fixed, randomness comes from the edge choices only.

• Partition $[0,1]^2$ into congruent squares of side length $1/(2\sqrt{2}r)$

• We add edges in four phases. In the first we start from $X_1$, and using $c_1$ choices of each vertex, we go for $\delta^2 \log_{c_1} n$ generations. There exists a cube in the grid that contains a connected component of size $n^{\text{const.}\delta^2}$.

• Second, we add $c_2$ new connections to each vertex in the component. At least one of the grid cells has a positive fraction of its points in a connected component.

# Supercritical $r$, constant $c$

Sketch of proof:

• First show that $X_1, \ldots, X_n$ are sufficiently regular whp. Once the $X_i$ are fixed, randomness comes from the edge choices only.

• Partition $[0,1]^2$ into congruent squares of side length $1/(2\sqrt{2}r)$

• We add edges in four phases. In the first we start from $X_1$, and using $c_1$ choices of each vertex, we go for $\delta^2 \log_{c_1} n$ generations. There exists a cube in the grid that contains a connected component of size $n^{\text{const.}\delta^2}$.

• Second, we add $c_2$ new connections to each vertex in the component. At least one of the grid cells has a positive fraction of its points in a connected component.

• Third, using $c_3$ new connections of each vertex, we obtain a connected component that contains a constant fraction of the points in every cell of the grid, whp.

# Supercritical $r$, constant $c$

Sketch of proof:

- First show that $X_1, \dots, X_n$ are sufficiently regular whp. Once the $X_i$ are fixed, randomness comes from the edge choices only.

- Partition $[0,1]^2$ into congruent squares of side length $1/(2\sqrt{2}r)$

- We add edges in four phases. In the first we start from $X_1$, and using $c_1$ choices of each vertex, we go for $\delta^2 \log_{c_1} n$ generations. There exists a cube in the grid that contains a connected component of size $n^{\text{const.}\delta^2}$.

- Second, we add $c_2$ new connections to each vertex in the component. At least one of the grid cells has a positive fraction of its points in a connected component.

- Third, using $c_3$ new connections of each vertex, we obtain a connected component that contains a constant fraction of the points in every cell of the grid, whp.

- Finally, add just one more connection per vertex so that the entire graph becomes connected.

# the giant component

Consider now $r > \lambda\sqrt{\log n/n}$ for a sufficiently large $\lambda$.

# the giant component

Consider now $r > \lambda\sqrt{\log n/n}$ for a sufficiently large $\lambda$.

It is not difficult to see that for $c = 1$ the size of the largest component is at most poly-logarithmic.

# the giant component

Consider now $r > \lambda\sqrt{\log n / n}$ for a sufficiently large $\lambda$.

It is not difficult to see that for $c = 1$ the size of the largest component is at most poly-logarithmic.

It is $o(n)$ even for much larger values of $r$.

# the giant component

Consider now $r > \lambda\sqrt{\log n/n}$ for a sufficiently large $\lambda$.

It is not difficult to see that for $c = 1$ the size of the largest component is at most poly-logarithmic.

It is $o(n)$ even for much larger values of $r$.

To study the phase transition, we generalize the model. Each vertex $x_i$ draws a bounded independent integer-valued random variable $\xi_i \geq 1$ and selects $\xi_i$ neighbors at random (without replacement).

# the giant component

Consider now $r > \lambda\sqrt{\log n/n}$ for a sufficiently large $\lambda$.

It is not difficult to see that for $c = 1$ the size of the largest component is at most poly-logarithmic.

It is $o(n)$ even for much larger values of $r$.

To study the phase transition, we generalize the model. Each vertex $x_i$ draws a bounded independent integer-valued random variable $\xi_i \geq 1$ and selects $\xi_i$ neighbors at random (without replacement).

Main result: for any $\epsilon > 0$, if $\mathbb{E}\xi_i \geq 1 + \epsilon$, then the size of the largest component is $n(1 - o(1))$ whp.

# the giant component

Consider now $r > \lambda\sqrt{\log n/n}$ for a sufficiently large $\lambda$.

It is not difficult to see that for $c = 1$ the size of the largest component is at most poly-logarithmic.

It is $o(n)$ even for much larger values of $r$.

To study the phase transition, we generalize the model. Each vertex $x_i$ draws a bounded independent integer-valued random variable $\xi_i \geq 1$ and selects $\xi_i$ neighbors at random (without replacement).

Main result: for any $\epsilon > 0$, if $\mathbb{E}\xi_i \geq 1 + \epsilon$, then the size of the largest component is $n(1 - o(1))$ whp.

Explosive percolation: the phase transition is discontinuous. We have even more: the proportion of vertices in the giant component jumps from $0$ to $1$. We have super-explosive percolation.

# explosive percolation

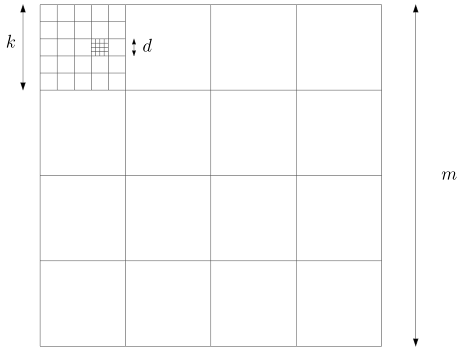
# the giant component: formal statement

For every $\delta \in (0, 1)$ there exists a $\gamma > 0$ such that for every $\epsilon > 0$, if $r \geq \gamma\sqrt{\log n/n}$ and $\mathbb{E}\xi > 1 + \epsilon$, then the largest component has size at least $n(1 - \delta)$ whp.

# the giant component: proof

The proof is a mix of branching process and percolation arguments.

We start with discretizing the torus $[0,1]^2$ into cells of side length $kr/2$. Each cell is further divided into boxes of side length $r/(2d)$. $k, d$ are large odd (constant) integers.



# uniformity lemma

During the entire proof we fix the vertex set $X$.

We need that they are sufficiently regularly placed.

# uniformity lemma

During the entire proof we fix the vertex set $\boldsymbol{X}$.

We need that they are sufficiently regularly placed.

One can prove that if $\gamma > 12d^2/\delta^2$ then whp every box $\boldsymbol{B}$ is $\delta$-good:

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

# uniformity lemma

During the entire proof we fix the vertex set $X$.

We need that they are sufficiently regularly placed.

One can prove that if $\gamma > 12d^2/\delta^2$ then whp every box $B$ is $\delta$-good:

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

This gives us a condition on $r$:

$$r \geq \gamma\sqrt{\frac{\log n}{n}}$$

# the web

In the first phase of the proof we prove the existence of a web:

# the web

In the first phase of the proof we prove the existence of a web:

Let $\boldsymbol{\epsilon > 0}$. There exist $\boldsymbol{\delta, k, d}$ such that if all cells are $\boldsymbol{\delta}$-good, then with probability at least $\boldsymbol{1 - \epsilon}$, $\boldsymbol{G(n, r, c)}$ has a connected component such that $\boldsymbol{1 - \epsilon}$ fraction of all boxes contain $\mathbb{E}[\xi]^{k^2/2}$ vertices of the component.

# the web

In the first phase of the proof we prove the existence of a web:

Let $\epsilon > 0$. There exist $\delta, k, d$ such that if all cells are $\delta$-good, then with probability at least $1 - \epsilon$, $G(n, r, c)$ has a connected component such that $1 - \epsilon$ fraction of all boxes contain $\mathbb{E}[\xi]^{k^2/2}$ vertices of the component.

This is the heart of the proof. We set up an exploration process and then couple it with a percolation model.

# node events

The cells define naturally an $\boldsymbol{m} \times \boldsymbol{m}$ grid.

We call the vertices nodes and the directed edges links

# node events

The cells define naturally an $\boldsymbol{m} \times \boldsymbol{m}$ grid.

We call the vertices nodes and the directed edges links

A node event occurs if, starting from a vertex in the central box of the cell, after $\boldsymbol{k^2}$ generations of edges, without exiting the cell, each box has at least $(\mathbb{E}[\xi])^{k^2/2}$ vertices.

# node events

The cells define naturally an $\boldsymbol{m} \times \boldsymbol{m}$ grid.

We call the vertices nodes and the directed edges links

A node event occurs if, starting from a vertex in the central box of the cell, after $\boldsymbol{k^2}$ generations of edges, without exiting the cell, each box has at least $(\mathbb{E}[\xi])^{k^2/2}$ vertices.

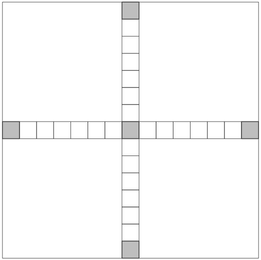
By coupling the growth process to a branching random walk, we show that a node event occurs with probability close to $\mathbf{1}$.

# link events

If a node event occurs, we try to “infect” the neighboring cells starting from the seed boxes:

# link events

If a node event occurs, we try to “infect” the neighboring cells starting from the seed boxes:

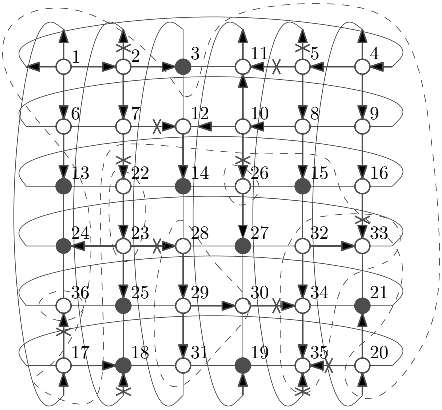


Of the $(\mathbb{E}[\xi])^{k^2/2}$ vertices in a seed box, at least one will connect to a vertex in the central box of the neighboring cell via a path of length $kd$ that always stays on the ladder.

This happens with probability near $\mathbf{1}$.

# exploration process

Three sets of nodes: explored, active, unseen.




Oriented connected components correspond to connected components of the web.

# mixed site/bond percolation

After the exploration process, all node events and some link events are defined. (All independent!)

# mixed site/bond percolation

After the exploration process, all node events and some link events are defined. (All independent!)

We assign independent Bernoulli variables to all undefined oriented links.

# mixed site/bond percolation

After the exploration process, all node events and some link events are defined. (All independent!)

We assign independent Bernoulli variables to all undefined oriented links.

We declare a bond open if both oriented links are open.

# mixed site/bond percolation

After the exploration process, all node events and some link events are defined. (All independent!)

We assign independent Bernoulli variables to all undefined oriented links.

We declare a bond open if both oriented links are open.

This defines a dense mixed site/bond percolation process on the grid. Any open component is an oriented connected component.

# mixed site/bond percolation

After the exploration process, all node events and some link events are defined. (All independent!)

We assign independent Bernoulli variables to all undefined oriented links.

We declare a bond open if both oriented links are open.

This defines a dense mixed site/bond percolation process on the grid. Any open component is an oriented connected component.

Using results of Deuschel and Pisztora (1996) for high-density site percolation, we conclude that there is an open component containing $\mathbf{1-\epsilon}$ fraction of the nodes. This gives us the web.

# connecting to the web

We constructed the web by revealing the edge choices of only a constant number ($\boldsymbol{(\mathbb{E}[\xi])^{k^2/2}}$) of vertices per cell.

# connecting to the web

We constructed the web by revealing the edge choices of only a constant number ($(\mathbb{E}[\xi])^{k^2/2}$) of vertices per cell.

Once the web is built, we connect almost all unseen vertices.

Take such a vertex. Build a new web starting from this point. The two webs will "see" each other in $\Theta(1/r^2)$ boxes and connect up with probability $1/(nr^2)$ at each point.

# connecting to the web

We constructed the web by revealing the edge choices of only a constant number ($(\mathbb{E}[\xi])^{k^2/2}$) of vertices per cell.

Once the web is built, we connect almost all unseen vertices.

Take such a vertex. Build a new web starting from this point. The two webs will "see" each other in $\Theta(1/r^2)$ boxes and connect up with probability $1/(nr^2)$ at each point.

The probability that any vertex is connected to the web is $1 - o(1)$.