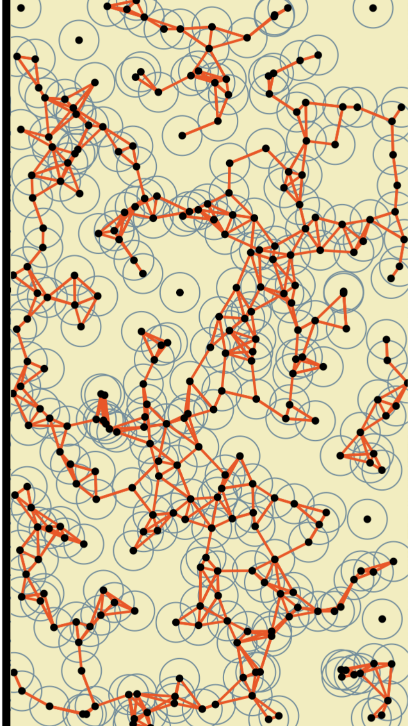
Connectivity of random irrigation networks

Nicolas Broutin, Luc Devroye,
Nicolas Fraiman, and Gábor Lugosi

September 4, 2012



Irrigation graph

Start with a connected graph on n vertices.

Irrigation graph

Start with a connected graph on n vertices.

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

Irrigation graph

Start with a connected graph on n vertices.

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

We consider the case when the underlying graph is a *random geometric graph* such that the vertices are $X_1, \ldots, X_n$ be i.i.d. uniform on $[0,1]^d$ and $X_i \sim X_j$ iff $\|X_i - X_j\| < r$.

Irrigation graph

Start with a connected graph on n vertices.

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

We consider the case when the underlying graph is a *random geometric graph* such that the vertices are $X_1, \ldots, X_n$ be i.i.d. uniform on $[0,1]^d$ and $X_i \sim X_j$ iff $\|X_i - X_j\| < r$.

Such graphs are also called *bluetooth graphs.*
They are *locally sparsified* random geometric graphs.

Irrigation graph

Start with a connected graph on n vertices.

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

We consider the case when the underlying graph is a *random geometric graph* such that the vertices are $X_1, \ldots, X_n$ be i.i.d. uniform on $[0,1]^d$ and $X_i \sim X_j$ iff $\|X_i - X_j\| < r$.

Such graphs are also called *bluetooth graphs.*
They are *locally sparsified* random geometric graphs.

The model was introduced by Ferraguto, Mambrini, Panconesi, and Petrioli *"A new approach to device discovery and scatternet formation in bluetooth networks"* (2004).

Irrigation graph

Start with a connected graph on n vertices.

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

We consider the case when the underlying graph is a *random geometric graph* such that the vertices are $X_1, \ldots, X_n$ be i.i.d. uniform on $[0,1]^d$ and $X_i \sim X_j$ iff $\|X_i - X_j\| < r$.

Such graphs are also called *bluetooth graphs.*
They are *locally sparsified* random geometric graphs.

The model was introduced by Ferraguto, Mambrini, Panconesi, and Petrioli *"A new approach to device discovery and scatternet formation in bluetooth networks"* (2004).

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

Irrigation graph

- $G(n,r,c)$ is a subgraph of a *random geometric graph* $G(n,r)$, so we need $G(n,r)$ to be connected.

Irrigation graph

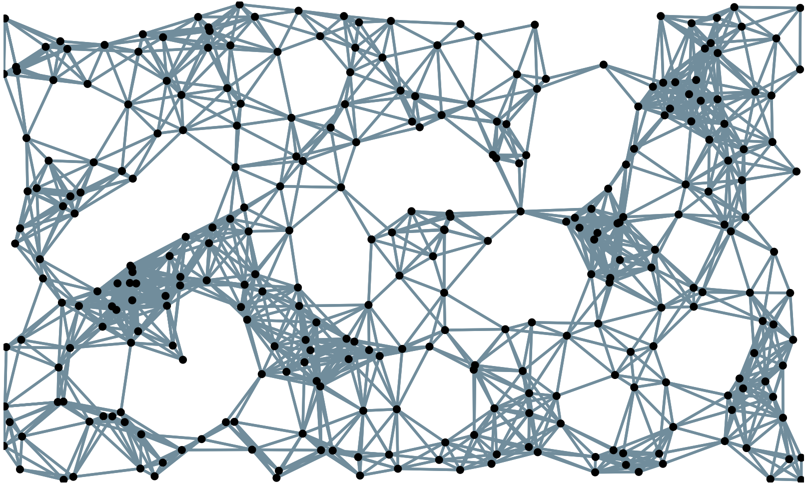
- $G(n,r,c)$ is a subgraph of a *random geometric graph* $G(n,r)$, so we need $G(n,r)$ to be connected.

- Penrose (1997) showed that $\forall\,\varepsilon > 0$, $G(n,r)$ is connected whp if $r \geq (1+\varepsilon)r_t$ where

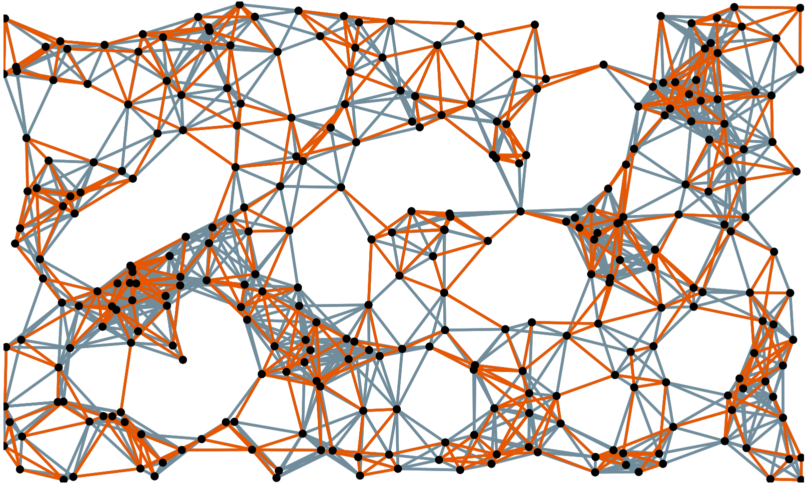
$$r_t = \theta_d \left(\frac{\log n}{n}\right)^{1/d} \qquad \text{and} \qquad \theta_d = \frac{2}{(2d \operatorname{Vol} B(0,1))^{1/d}}.$$

  We only consider values of r above this level.

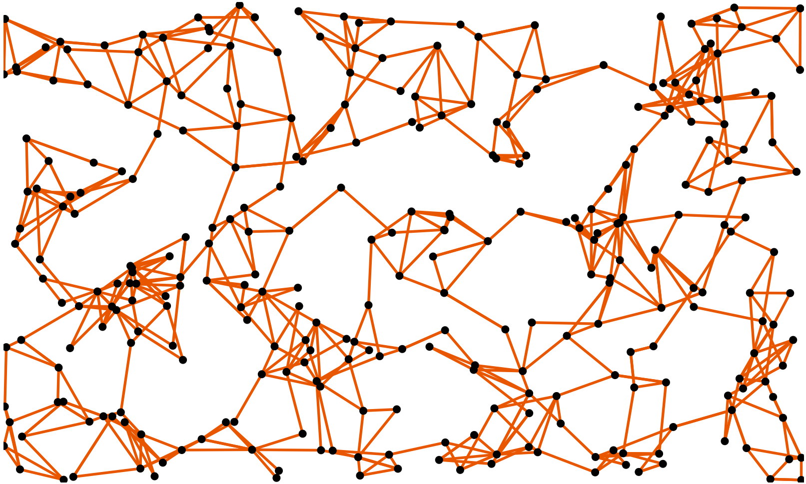
Irrigation graph



Irrigation graph



Irrigation graph



Previous results

Theorem (Fenner and Frieze, 1982)

For $r = \infty$, the graph $G(n, r, 2)$ (the random 2-out graph) is connected whp.

Previous results

Theorem (Fenner and Frieze, 1982)

For $r = \infty$, the graph $G(n, r, 2)$ (the random 2-out graph) is connected whp.

Theorem (Dubhashi, Johansson, Häggström, Panconesi, Sozio, 2007)

For constant r the graph $G(n, r, 2)$ is connected whp.

Previous results

Theorem (Fenner and Frieze, 1982)

For $r = \infty$, the graph $G(n, r, 2)$ (the random 2-out graph) is connected whp.

Theorem (Dubhashi, Johansson, Häggström, Panconesi, Sozio, 2007)

For constant r the graph $G(n, r, 2)$ is connected whp.

Theorem (Crescenzi, Nocentini, Pietracaprina, Pucci, 2009)

In dimension $d = 2$, $\exists\, \alpha, \beta$ 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.

Main result

Theorem

There exists a constant $\gamma^* > 0$ *such that for all* $\gamma \geq \gamma^*$ *and* $\varepsilon \in (0,1)$*, if*

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

Main result

Theorem

There exists a constant $\gamma^* > 0$ such that for all $\gamma \geq \gamma^*$ and $\varepsilon \in (0,1)$, if

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

then

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

Main result

Theorem

There exists a constant $\gamma^* > 0$ such that for all $\gamma \geq \gamma^*$ and $\varepsilon \in (0,1)$, if

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

then

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

c_t does not depend on γ or d.

Below the threshold

Theorem

Let $\gamma \geq \gamma^*$ *and* $\varepsilon \in (0,1)$*. If* $r = \gamma \left(\frac{\log n}{n}\right)^{1/d}$ *and* $c \leq (1-\varepsilon) c_t$ *then* $G(n,r,c)$ *is disconnected whp.*

Below the threshold

Theorem

Let $\gamma \geq \gamma^*$ *and* $\varepsilon \in (0,1)$. *If* $r = \gamma \left(\frac{\log n}{n}\right)^{1/d}$ *and* $c \leq (1-\varepsilon)c_t$ *then* $G(n,r,c)$ *is disconnected whp.*

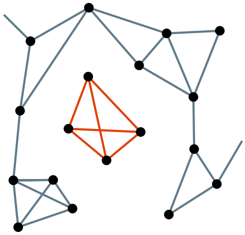
- The smallest possible components are cliques of size $c+1$.

Below the threshold

Theorem

Let $\gamma \geq \gamma^*$ and $\varepsilon \in (0,1)$. If $r = \gamma\left(\frac{\log n}{n}\right)^{1/d}$ and $c \leq (1-\varepsilon)c_t$ then $G(n,r,c)$ is disconnected whp.

- The smallest possible components are cliques of size $c+1$.



Isolated (c+1)-cliques

- We show that there exists an isolated $(c+1)$-clique whp.

Isolated (c+1)-cliques

- We show that there exists an isolated $(c+1)$-clique whp.
- Let $\mathscr{F}$ be the random family of subsets of $\{1,\ldots,n\}$ given by

$$\mathscr{F} = \Big\{ Q \subset \{1,\ldots,n\} : |Q| = c+1,\ \|X_i - X_j\| < r\ \forall\, i,j \in Q \Big\}.$$

Isolated (c+1)-cliques

- We show that there exists an isolated $(c+1)$-clique whp.

- Let $\mathscr{F}$ be the random family of subsets of $\{1,\ldots,n\}$ given by

$$\mathscr{F} = \Big\{Q \subset \{1,\ldots,n\} : |Q| = c+1,\ \|X_i - X_j\| < r\ \forall\, i,j \in Q\Big\}.$$

- Let $I(Q)$ be the indicator of the event that Q is an isolated clique. Then $N = \sum_{Q\in\mathscr{F}} I(Q)$ is the number of isolated $(c+1)$-cliques.

Isolated (c+1)-cliques

- We need some regularity on the uniformly distributed points. For every $1 \leq j \leq n$

$$\alpha n r^2 < \#\{i : X_i \in B(X_j, r)\} < \beta n r^2.$$

Isolated (c+1)-cliques

- We need some regularity on the uniformly distributed points. For every $1 \le j \le n$

$$\alpha n r^2 < \#\{i : X_i \in B(X_j, r)\} < \beta n r^2.$$

- Let D be the event described above. We use the second-moment method and prove that

$$\mathbf{P}\{N 1_D > 0\} \ge \frac{\mathbf{E}\{N 1_D\}^2}{\mathbf{E}\left\{N^2 1_D\right\}} \to 1.$$

Above the threshold

Theorem

Let $\gamma \geq \gamma^*$ *and* $\varepsilon \in (0,1)$. *If* $r = \gamma \left(\frac{\log n}{n}\right)^{1/d}$ *and* $c \geq (1+\varepsilon)c_t$ *then* $G(n,r,c)$ *is connected whp.*

Gridding and percolation

We tile the unit square $[0, 1]^2$ into cells of side length r.

Gridding and percolation

We tile the unit square $[0,1]^2$ into cells of side length r.

- Two cells are connected if they are adjacent and there is an edge between one vertex of each cell.

Gridding and percolation

We tile the unit square $[0,1]^2$ into cells of side length r.

- Two cells are connected if they are adjacent and there is an edge between one vertex of each cell.
- Two cells are $*$-connected if they share at least a corner and there is an edge between one vertex of each cell.

Gridding and percolation

We tile the unit square $[0,1]^2$ into cells of side length r.

- Two cells are connected if they are adjacent and there is an edge between one vertex of each cell.
- Two cells are $*$-connected if they share at least a corner and there is an edge between one vertex of each cell.
- A cell is colored black if all the vertices in it are connected to each other without using an edge that leaves the cell. The other cells are initially colored white.

Gridding and percolation

The following properties hold whp:

Gridding and percolation

The following properties hold whp:

1. Every cell in the grid contains at most $\lambda \log n$ vertices for some $\lambda = \lambda(\gamma)$.

Gridding and percolation

The following properties hold whp:

1. Every cell in the grid contains at most $\lambda \log n$ vertices for some $\lambda = \lambda(\gamma)$.

2. Every cell in the grid connects to its adjacent cells.

Gridding and percolation

The following properties hold whp:

1. Every cell in the grid contains at most $\lambda \log n$ vertices for some $\lambda = \lambda(\gamma)$.
2. Every cell in the grid connects to its adjacent cells.
3. Every $*$-connected component of white cells has size at most $q = 2(\log n)^{2/3}$.

Gridding and percolation

The following properties hold whp:

1. Every cell in the grid contains at most $\lambda \log n$ vertices for some $\lambda = \lambda(\gamma)$.
2. Every cell in the grid connects to its adjacent cells.
3. Every $*$-connected component of white cells has size at most $q = 2(\log n)^{2/3}$.
4. Every connected component of G has size at least $s = \exp((\log n)^{1/3})$.

The four properties

1. Every cell in the grid contains at most $\lambda \log n$ vertices.

The four properties

1. Every cell in the grid contains at most $\lambda \log n$ vertices.

- Concentration of number of points in cells.
- $\mathbf{E}\{\# C\} = \Theta(nr^2) = \Theta(\log n)$.

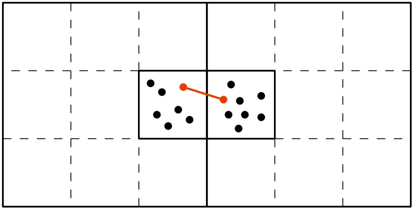
The four properties

2. Every cell in the grid connects to its adjacent cells.

The four properties

2. Every cell in the grid connects to its adjacent cells.

- Subdivide the cell and find an edge bewteen two squares in the border.



The four properties

3. Every $*$-connected component of white cells has size at most $q = 2(\log n)^{2/3}$.

The four properties

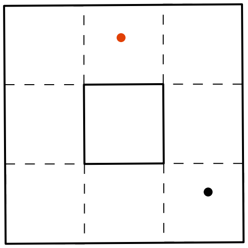
3. Every $*$-connected component of white cells has size at most $q = 2(\log n)^{2/3}$.

- $\#\{*\text{-connected comp. of size } k\} \leq n(8e)^k$.

The four properties

3. Every $*$-connected component of white cells has size at most $q = 2(\log n)^{2/3}$.

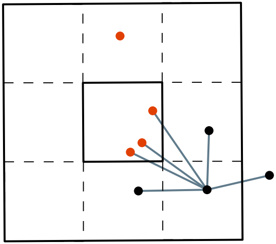
- $\#\{*\text{-connected comp. of size } k\} \le n(8e)^k$.
- It suffices to show that $\mathbf{P}\{\text{Cell is white}\} \le p = \exp(-(\log n)^{2/3})$.



The four properties

3. Every $*$-connected component of white cells has size at most $q = 2(\log n)^{2/3}$.

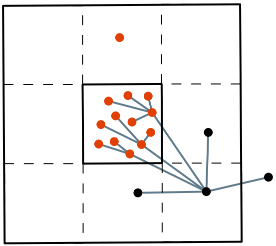
- $\#\{*\text{-connected comp. of size } k\} \le n(8e)^k$.
- It suffices to show that $\mathbf{P}\{\text{Cell is white}\} \le p = \exp(-(\log n)^{2/3})$.



The four properties

3. Every $*$-connected component of white cells has size at most $q = 2(\log n)^{2/3}$.

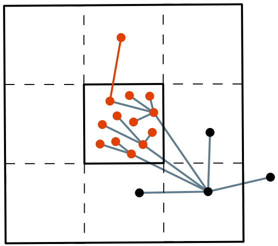
- $\#\{*\text{-connected comp. of size } k\} \le n(8e)^k$.
- It suffices to show that $\mathbf{P}\{\text{Cell is white}\} \le p = \exp(-(\log n)^{2/3})$.



The four properties

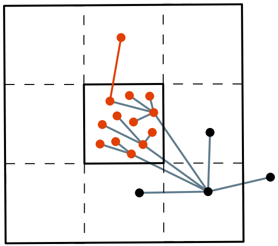
3. Every $*$-connected component of white cells has size at most $q = 2(\log n)^{2/3}$.

- $\#\{*\text{-connected comp. of size } k\} \leq n(8e)^k$.
- It suffices to show that $\mathbf{P}\{\text{Cell is white}\} \leq p = \exp(-(\log n)^{2/3})$.



The four properties

3. Every $*$-connected component of white cells has size at most $q = 2(\log n)^{2/3}$.



- $\#\{*\text{-connected comp. of size } k\} \le n(8e)^k$.
- It suffices to show that $\mathbf{P}\{\text{Cell is white}\} \le p = \exp(-(\log n)^{2/3})$.
- If $k > q$ then $n(8e)^k p^k \to 0$.

The four properties

4. Every connected component of G has size at least $s = \exp((\log n)^{1/3})$.

The four properties

4. Every connected component of G has size at least $s = \exp((\log n)^{1/3})$.

- Save $(\varepsilon/2)c_t$ edge choices.

The four properties

4. Every connected component of G has size at least $s = \exp((\log n)^{1/3})$.

- Save $(\varepsilon/2)c_t$ edge choices.
- No small components with $(1+\varepsilon/2)c_t$ choices.

The four properties

4. Every connected component of G has size at least $s = \exp((\log n)^{1/3})$.

- Save $(\varepsilon/2)c_t$ edge choices.
- No small components with $(1+\varepsilon/2)c_t$ choices.
- Use extra edges iteratively to double the size of components.

Gridding and percolation

1. Every cell in the grid contains at most $\lambda \log n$ vertices.
2. Every cell in the grid connects to its adjacent cells.
3. Every $*$-connected component of white cells has size at most q cells.
4. Every connected component of G has size at least s.

If all properties hold, then the whole graph is connected.

Everything is connected

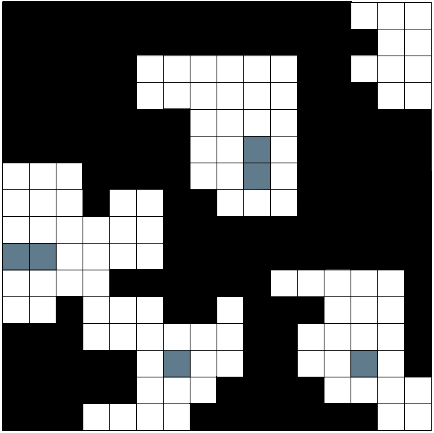
- Black connector: There exists a connected component of black cells that links two opposite sides of $[0,1]^2$.

Everything is connected

- Black connector: There exists a connected component of black cells that links two opposite sides of $[0,1]^2$.
- Black giant: Black components of size less than $1/r$ are now recolored gray. All remaining black cells are connected. The corresponding vertices of G belong to the same connected component.
- Connectivity: Each vertex connects to at least one vertex of the black giant.

Everything is connected

K



K'

Spanning ratio and diameter

An important feature of a geometric graph is the *spanning ratio*

$$\sup_{i,j} \frac{\mathrm{dist}(X_i, X_j)}{\|X_i - X_j\|}$$

where $\mathrm{dist}(X_i, X_j)$ is the shortest (Euclidean) distance of X_i and X_j over the edges of the graph. Ideally, this should be small.

Spanning ratio and diameter

An important feature of a geometric graph is the *spanning ratio*

$$\sup_{i,j} \frac{\mathrm{dist}(X_i, X_j)}{\|X_i - X_j\|}$$

where $\mathrm{dist}(X_i, X_j)$ is the shortest (Euclidean) distance of X_i and X_j over the edges of the graph. Ideally, this should be small.

Unfortunately, this can be large if X_i and X_j are very close.

Spanning ratio and diameter

An important feature of a geometric graph is the *spanning ratio*

$$\sup_{i,j} \frac{\mathrm{dist}(X_i, X_j)}{\|X_i - X_j\|}$$

where $\mathrm{dist}(X_i, X_j)$ is the shortest (Euclidean) distance of X_i and X_j over the edges of the graph. Ideally, this should be small.

Unfortunately, this can be large if X_i and X_j are very close.
However, for c slightly larger than critical, we have

Theorem

$\exists\, K, \mu > 0$ *such that if* $\gamma > \gamma^*$, $\quad r = \gamma \left(\frac{\log n}{n}\right)^{1/d}$ *and* $c \geq \mu\sqrt{\log n}$ *then*

$$\sup_{i,j : \|X_i - X_j\| > r} \frac{\mathrm{dist}(X_i, X_j)}{\|X_i - X_j\|} \leq K, \quad \textit{whp.}$$

Spanning ratio and diameter

This implies that the diameter of G satisfies $\operatorname{diam}(G) \le K\sqrt{d}/r$. This is optimal, up to a constant factor.

Spanning ratio and diameter

This implies that the diameter of G satisfies $\mathrm{diam}(G) \leq K\sqrt{d}/r$. This is optimal, up to a constant factor.

Idea of the proof: Partition the unit cube into a grid of cells of side length $\ell = (1/3)\lfloor 1/r \rfloor^{-1}$.

With high probability, any two points i and j, such that X_i and X_j fall in the same cell, are connected by a path of length at most five.

On the other hand, with high probability, any two neighboring cells contain two points, one in each cell, that are connected by an edge of S_n.

Supercritical radii

The proof of disconnectedness may be generalized easily for the entire range of values of r.

Supercritical radii

The proof of disconnectedness may be generalized easily for the entire range of values of r.

Theorem

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

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

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

Supercritical radii

The proof of disconnectedness may be generalized easily for the entire range of values of r.

Theorem

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

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

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

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

Supercritical r, constant c

We can show that the lower bound is not far from the truth: when $r \sim n^{-(1-\delta)/d}$, constant c is sufficient for connectivity.

$c = \sqrt{5/\delta} + c(d)$ is sufficient for connectivity.

The irrigation graph is genuinely sparse.

Supercritical r, constant c

Theorem

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

$$c_1 = \lceil \sqrt{5/(\delta - \delta^2)} \rceil ,$$

and c_2, c_3 depend on d 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, all randomness comes from the edge choices.

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, all randomness comes from the edge choices.

• 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, all randomness comes from the edge choices.

• 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, all randomness comes from the edge choices.

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

• 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.

Thank you