Bootstrap percolation on random graphs

Svante Janson
(with Tomasz Łuczak, Tatyana Turova, Thomas Vallier)

INFORMS APS 2011, Stockholm, 8 July 2011

Bootstrap percolation

Bootstrap percolation is a process of spread of *activation* (or *infection*) on a graph G.
Fix a constant threshold $r \geq 2$. (Typically, $r = 2$.)

We start with some set $\mathcal{A}_0$ of the vertices *active*. At each step every inactive vertex with at least r active neighbours becomes active. This is repeated until no more vertices become active. (Active vertices never become inactive.)

Bootstrap percolation

Bootstrap percolation is a process of spread of *activation* (or *infection*) on a graph G.
Fix a constant threshold $r \geq 2$. (Typically, $r = 2$.)

We start with some set $\mathcal{A}_0$ of the vertices *active*. At each step every inactive vertex with at least r active neighbours becomes active. This is repeated until no more vertices become active. (Active vertices never become inactive.)

What is the size of the final active set $\mathcal{A}^*$?

Bootstrap percolation

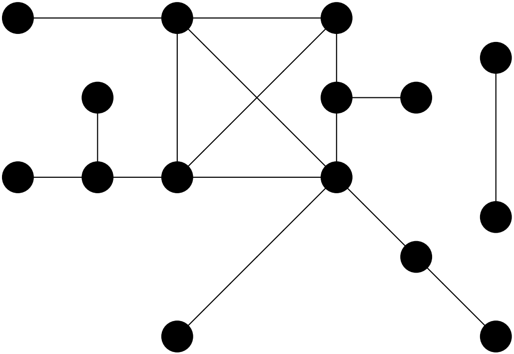
Bootstrap percolation is a process of spread of *activation* (or *infection*) on a graph G.
Fix a constant threshold $r \geq 2$. (Typically, $r = 2$.)

We start with some set $\mathcal{A}_0$ of the vertices *active*. At each step every inactive vertex with at least r active neighbours becomes active. This is repeated until no more vertices become active. (Active vertices never become inactive.)

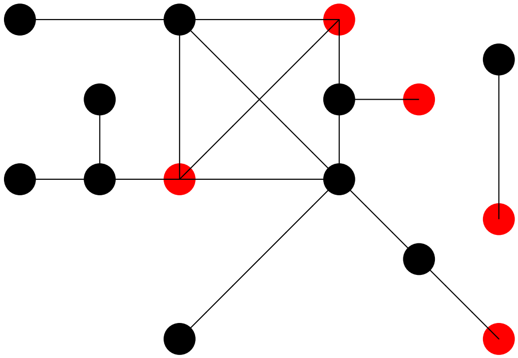
What is the size of the final active set $\mathcal{A}^*$?

In particular: Will eventually all vertices be active?
("$\mathcal{A}_0$ percolates")

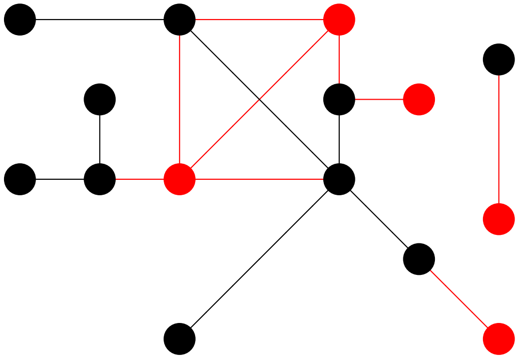
Example



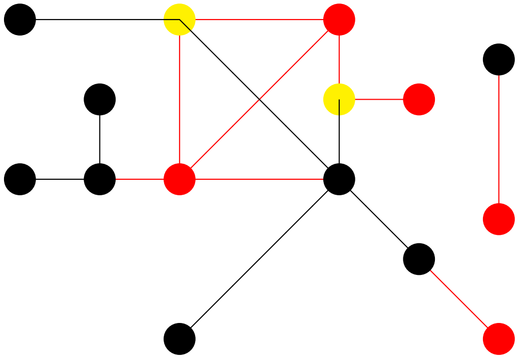
Bootstrap percolation



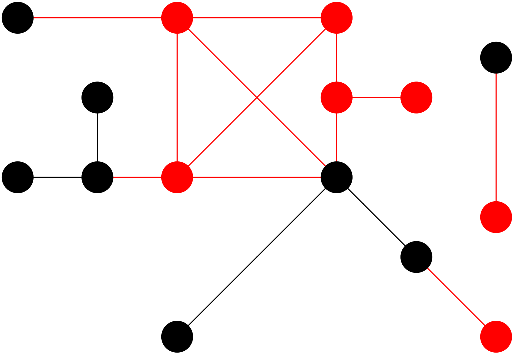
Bootstrap percolation



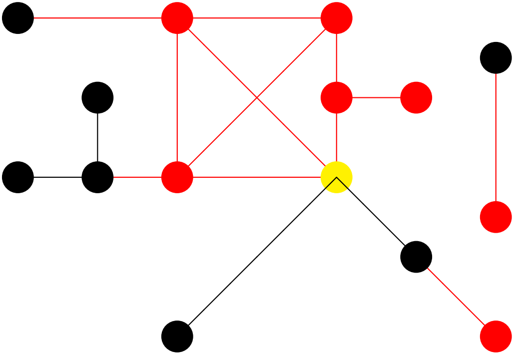
Bootstrap percolation



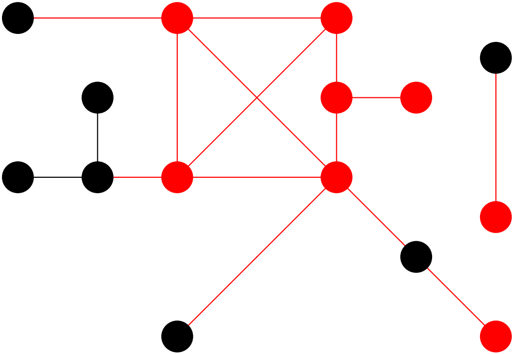
Bootstrap percolation



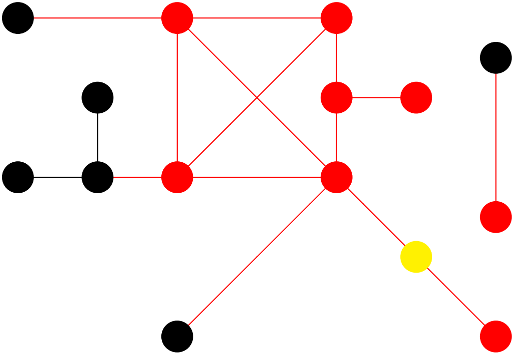
Bootstrap percolation



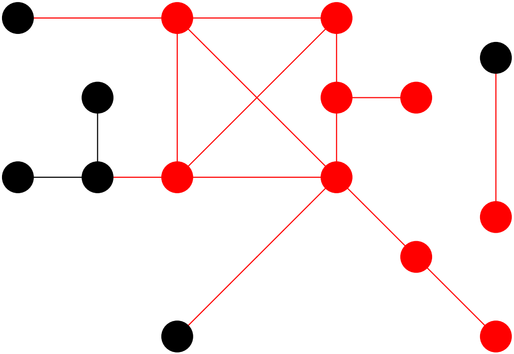
Bootstrap percolation



Bootstrap percolation



Bootstrap percolation



Applications?

Bootstrap percolation is not a good model for usual infectious diseases, but it has been studied a lot for other reasons. Some reasons, apart from the intrinsic mathematical interest:

- Epidemics with different degrees of severity (Scalia-Tomba (1985); Ball and Britton (2005))
- A model for the spread of rumors or beliefs.
- Neural networks (Amini (2010)).
- Spread of defaults in banking systems (Amini, Cont and Minca (2010+) with a more refined model).
- Relations to statistical mechanics and the Ising model.
- Cellular automata.

Extensions of the model include different thresholds for different vertices; weighted edges; directed edges.

Different types of problems

Several types of problems have been studied by a number of authors:

- The graph G can be *deterministic* (e.g. a grid in 2 or d dimensions, a torus, a hypercube, a regular infinite tree, ...) or *random* (e.g. an Erdős–Rényi graph $G(n,p)$, a random regular graph, ...).
- The initial set $\mathcal{A}_0$ can be *deterministic* (e.g. the minimal percolating set) or *random* (with a given number a active vertices chosen at random, or with each vertex initially active with probability q, independent of all others).
- One can ask for exact results for a fixed graph G, or for asymptotic results as the size of G grows.

Different types of problems

Several types of problems have been studied by a number of authors:

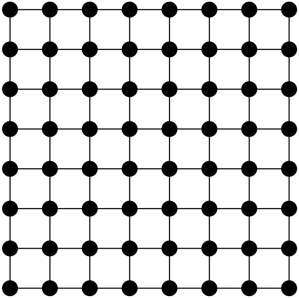
- The graph G can be *deterministic* (e.g. a grid in 2 or d dimensions, a torus, a hypercube, a regular infinite tree, ...) or *random* (e.g. an Erdős–Rényi graph $G(n, p)$, a random regular graph, ...).
- The initial set $\mathcal{A}_0$ can be *deterministic* (e.g. the minimal percolating set) or *random* (with a given number a active vertices chosen at random, or with each vertex initially active with probability q, independent of all others).
- One can ask for exact results for a fixed graph G, or for asymptotic results as the size of G grows.

My main interest is in asymptotic results for a random initial set in a random graph.

Examples of other types

A classical deterministic folklore problem:

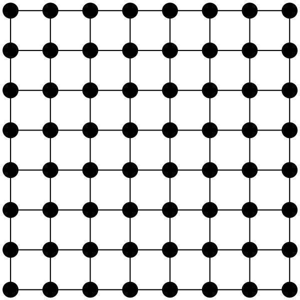
Find the smallest initial set that percolates on a chessboard (8×8 grid) with $r = 2$.



Examples of other types

A classical deterministic folklore problem:

Find the smallest initial set that percolates on a chessboard (8×8 grid) with $r = 2$.

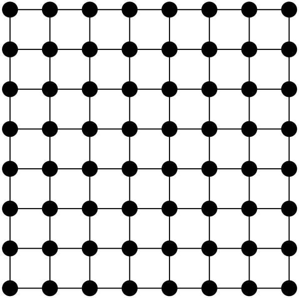


Example: A diagonal.

Examples of other types

A classical deterministic folklore problem:

Find the smallest initial set that percolates on a chessboard (8×8 grid) with $r = 2$.



Example: A diagonal.
This is optimal: At least 8 initially active are needed.

Deterministic graph, random initial set:

Consider a $n \times n$ grid and let each vertex be initially active with probability $q = c/\log n$ (independently of each other). Take $r = 2$. Let $n \to \infty$.

Theorem (Holroyd)

If $c < \pi^2/18$, then $\mathbb{P}(percolation) \to 0$ (w.h.p. no percolation)
If $c > \pi^2/18$, then $\mathbb{P}(percolation) \to 1$ (w.h.p. percolation).

Deterministic graph, random initial set:

Consider a $n \times n$ grid and let each vertex be initially active with probability $q = c/\log n$ (independently of each other). Take $r = 2$. Let $n \to \infty$.

Theorem (Holroyd)

If $c < \pi^2/18$, then $\mathbb{P}(percolation) \to 0$ (w.h.p. no percolation)
If $c > \pi^2/18$, then $\mathbb{P}(percolation) \to 1$ (w.h.p. percolation).

This has been generalized to any dimension d and $2 \leq r \leq d$ by Balogh, Bollobás, Duminil-Copin and Morris (2011+); the threshold is

$$\left(\frac{\lambda(d,r)}{\log \cdots \log n}\right)^{d-r+1}$$

with an $r-1$ iterated logarithm.

Some further references

Deterministic initial set (extremal problem): Balogh and Pete (1998) and Bollobás (2006) (grids)

Random initial set: Chalupa, Leath and Reich (1979) (regular infinite tree); Aizenman and Lebowitz (1988), Balogh and Pete (1998), Cerf and Manzo (2002) (grids); Balogh and Bollobás (2006) (hypercube); Balogh, Peres and Pete (2006), Fontes and Schonmann (2008) (infinite trees); Balogh and Pittel (2007), Janson (2009) (random regular graphs); Amini (2010) (random graphs with given vertex degrees).

The random graph G(n,p)

Consider now the case $G = G(n, p)$, the Erdős–Rényi random graph with n vertices and each possible edge appearing with probability p, independently of all other edges.

The random graph G(n,p)

Consider now the case $G = G(n, p)$, the Erdős–Rényi random graph with n vertices and each possible edge appearing with probability p, independently of all other edges.

(The random graph $G(n, m)$ with a fixed number m of edges, uniformly chosen among all such graphs on n labelled vertices, yields similar asymptotic results as $G(n, p)$ with $p = m/\binom{n}{2}$.)

Setup

We let $G = G(n, p)$ where $p = p(n)$ and start with a random set $\mathcal{A}_0$ with $a = a(n)$ elements; $r \geq 2$ is fixed.

We consider $p = p(n)$ as given, and ask how large a must be in order to give percolation.

Alternatively, and essentially equivalently, one might regard $a = a(n)$ as given and ask how large p must be.

We assume for simplicity

$$n^{-1} \ll p \ll n^{-1/r}.$$

(Boundary cases $p = c/n$ and $p = c/n^{1/r}$ are similar but different.)
We assume also $a \leq n/2$.

Define (the special case $r = 2$ in blue)

$$t_c := \left(\frac{(r-1)!}{np^r}\right)^{1/(r-1)} \qquad \frac{1}{np^2}$$

$$a_c := \left(1 - \frac{1}{r}\right) t_c \qquad \frac{1}{2np^2}$$

$$b_c := n\frac{(pn)^{r-1}}{(r-1)!}e^{-pn} \qquad n^2 pe^{-pn}.$$

Then

$$\begin{array}{ccc} t_c \to \infty, & pt_c \to 0, & t_c/n \to 0, \\ a_c \to \infty, & pa_c \to 0, & a_c/n \to 0, \\ & pb_c \to 0, & b_c/n \to 0. \end{array}$$

Theorem

(i). *If* $a/a_c \to \alpha < 1$, *then* $|\mathcal{A}^*| = (\varphi(\alpha) + o_p(1))t_c$, *where* $\varphi(\alpha)$ *is the unique root in* $[0,1]$ *of*

$$r\varphi(\alpha) - \varphi(\alpha)^r = (r-1)\alpha.$$

(For $r = 2$, $\varphi(\alpha) = 1 - \sqrt{1-\alpha}$*.) Further,* $|\mathcal{A}^*|/a \xrightarrow{p} \varphi_1(\alpha) := \frac{r}{r-1}\varphi(\alpha)/\alpha$, *with* $\varphi_1(0) := 1$*. In particular,* $|\mathcal{A}^*| < 2a$ *w.h.p.*

(ii). *If* $a/a_c \geq 1 + \delta$, *for some* $\delta > 0$, *then* $|\mathcal{A}^*| = n - o_p(n)$*; in other words, we have w.h.p. almost percolation. More precisely,* $|\mathcal{A}^*| = n - O_p(b_c)$*.*

(iii). *In case (iii) we further have complete percolation, i.e.* $|\mathcal{A}^*| = n$ *w.h.p., if and only if* $b_c \to 0$*, if and only if* $np - (\log n + (r-1)\log\log n) \to \infty$*.*

The number of vertices with degree $\leq r-1$ is about b_c. These vertices are never activated unless they happen to be among the initially active a vertices.

Part (iii) of the theorem says that these vertices are the main obstacle to complete percolation, and it is more interesting to study almost percolation.

Typical behaviour (when $a \approx a_c$):

1. First a slow growth; the number of activated vertices in each generation decreases.
2. There is a bottleneck when the total size of the active set is $\approx t_c$. The process may die out at this stage. If it does not, it will then grow rapidly (doubly exponentially) until almost all vertices are active. (There are many vertices with $r-1$ active neighbours; these have a large chance to become active in the next round.)
3. If p is sufficiently large, phase 2 ends with all vertices active (percolation). If p is small, there will be some vertices of degree $< r$ which will never be activated (perhaps together with some other vertices). In this case there may be a final phase of slow growth when the last vertices are activated.

A dynamical version of the threshold

We can study the treshold for a by considering a dynamical version, where we start with all vertices inactive and then activate them (from the outside) one by one, in random order. The bootstrap percolation mechanism works (instantaneously) after each external activation. Let A_0 be the number of externally activated vertices when the active set $\mathcal{A}$ becomes big, say $0.5n$ vertices (or $0.99n$ vertices, or in the case of complete percolation, all vertices).

Theorem

$$A_0/a_c \xrightarrow{\mathrm{p}} 1.$$

More precise threshold

Let

$$\tilde{\pi}(t) := \mathbb{P}\big(\mathrm{Po}(tp) \geq r\big) = \sum_{j=r}^{\infty} \frac{(pt)^j}{j!} e^{-pt}.$$

and

$$a_c^* := -\min_{t \leq 3t_c} \frac{n\tilde{\pi}(t) - t}{1 - \tilde{\pi}(t)},$$

Then $a_c^* \sim a_c$.

The precise threshold for a is $a_c^* \pm O(\sqrt{a_c})$,
with a width of the threshold of the order $\sqrt{a_c} \sim \sqrt{a_c^*}$.

More precisely, we have a Gaussian limit.

Theorem

$$A_0 \sim \mathrm{AsN}\big(a_{\mathrm{c}}^*, a_{\mathrm{c}}/(r-1)\big).$$

In other words,

$$\frac{A_0 - a_{\mathrm{c}}^*}{\sqrt{a_{\mathrm{c}}/(r-1)}} \xrightarrow{\mathrm{d}} N(0,1).$$

The number of generations

Theorem

Suppose that $a - a_c^* \gg \sqrt{a_c}$ *(so that* $\mathcal{A}(0)$ *w.h.p. almost percolates) and* $a = o(n)$*.*
Then the number of generations is w.h.p.

$$\sim \frac{\pi\sqrt{2}}{\sqrt{r-1}}\left(\frac{t_c}{a - a_c^*}\right)^{1/2} + \frac{1}{\log r}\left(\log\log(np) - \log_+ \log\frac{a}{a_c}\right) + \frac{\log n}{np} + O_p(1)$$

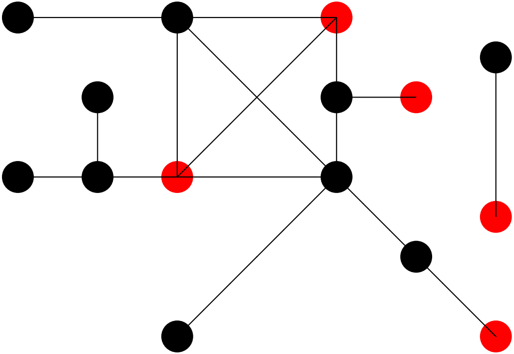
The three terms (excepting the error term) correspond to the three phases above.
Each of the three terms may be the dominating one.

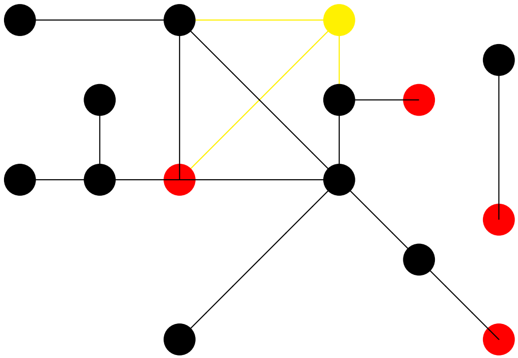
Proofs

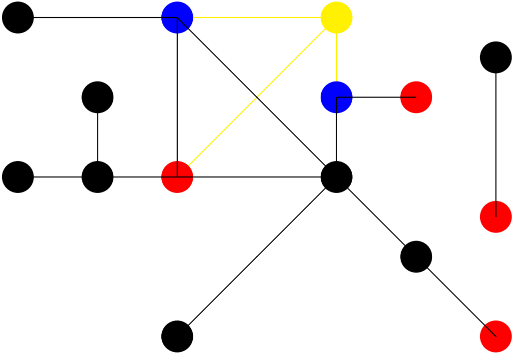
We first change the time scale; we forget the generations and consider at each time step the infections from one vertex only. Choose $u_1 \in \mathcal{A}(0) = \mathcal{A}_0$ and give each of its neighbours a *mark*; we then say that u_1 is *used*, and let $\mathcal{Z}(1) := \{u_1\}$ be the set of used vertices at time 1.
We continue recursively: At time t, choose a vertex $u_t \in \mathcal{A}(t-1) \setminus \mathcal{Z}(t-1)$. We give each neighbour of u_t a new mark. Let $\Delta\mathcal{A}(t)$ be the set of inactive vertices with r marks; these now become active and we let $\mathcal{A}(t) = \mathcal{A}(t-1) \cup \Delta\mathcal{A}(t)$ be the set of active vertices at time t. We finally set $\mathcal{Z}(t) = \mathcal{Z}(t-1) \cup \{u_t\} = \{u_s : s \leq t\}$, the set of used vertices.
The process stops when $\mathcal{A}(t) \setminus \mathcal{Z}(t) = \emptyset$, i.e., when all active vertices are used. We denote this time by T;

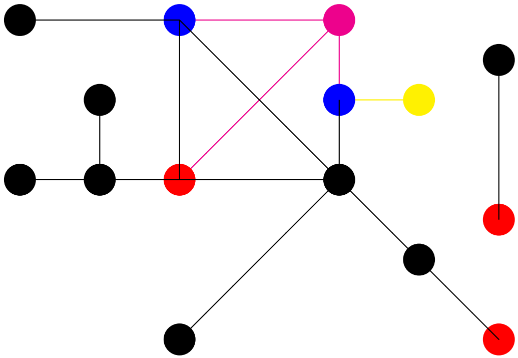
$$T := \min\{t \geq 0 : \mathcal{A}(t) \setminus \mathcal{Z}(t) = \emptyset\}.$$

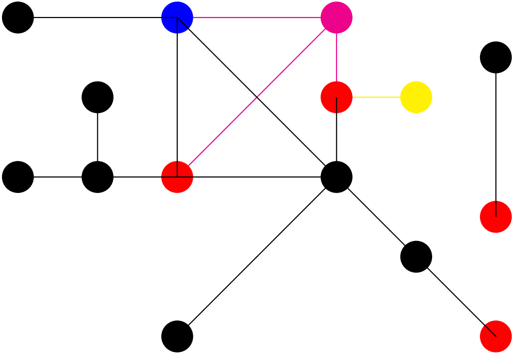
The final active set is $\mathcal{A}(T)$.
(Cf. Scalia-Tomba (1985) and Sellke (1983).)

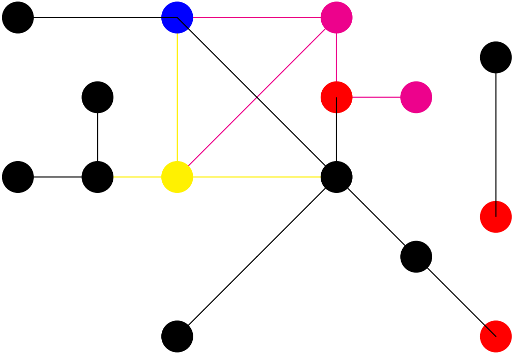


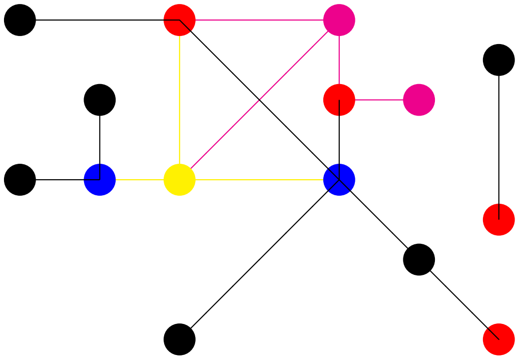


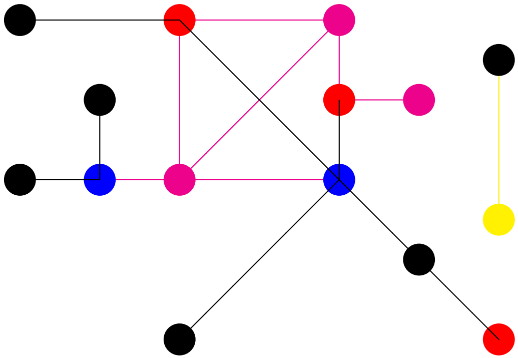


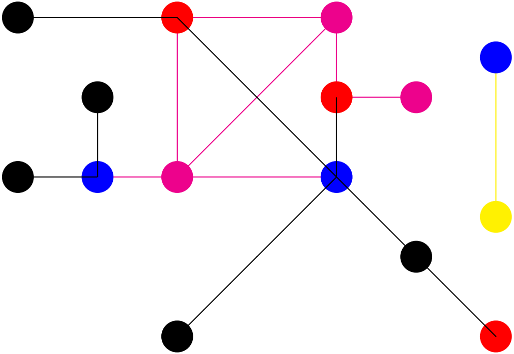


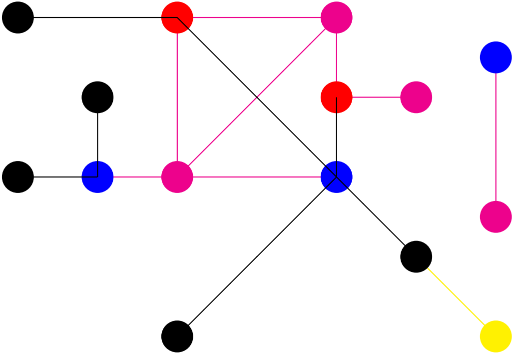


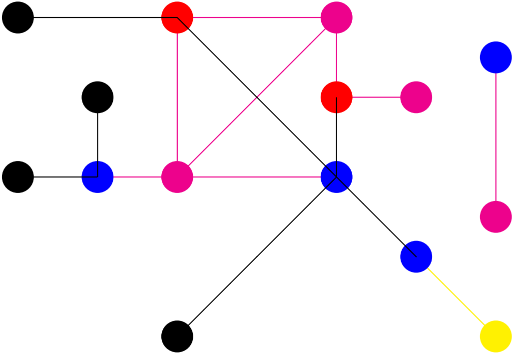


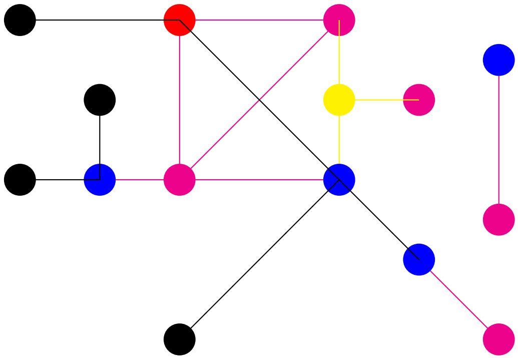


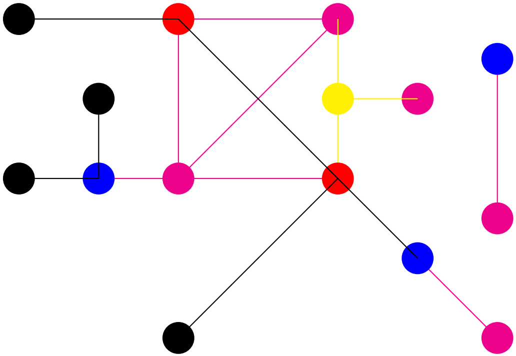


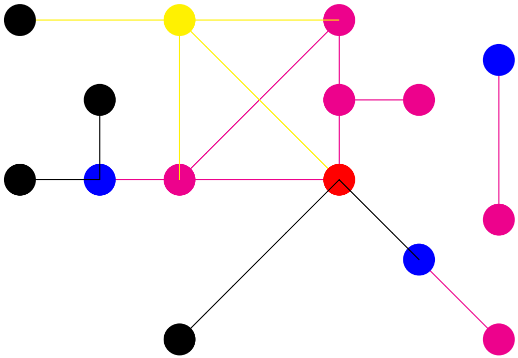


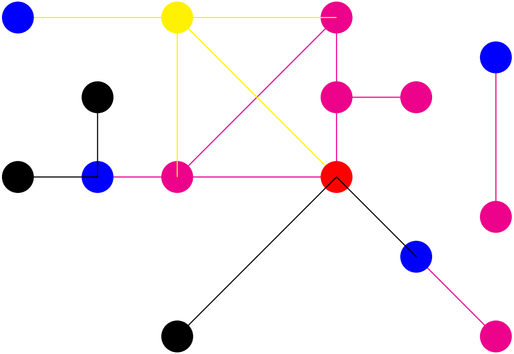


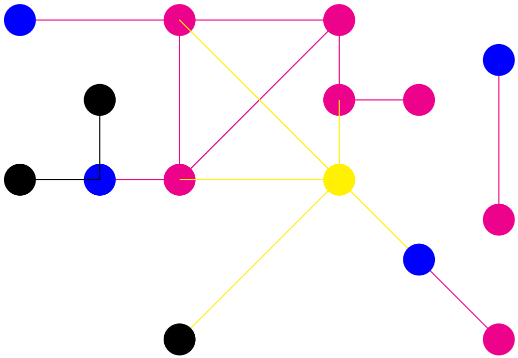


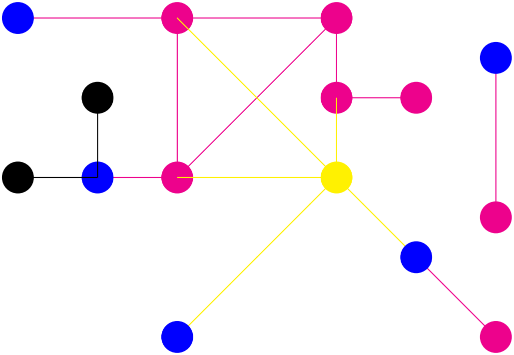


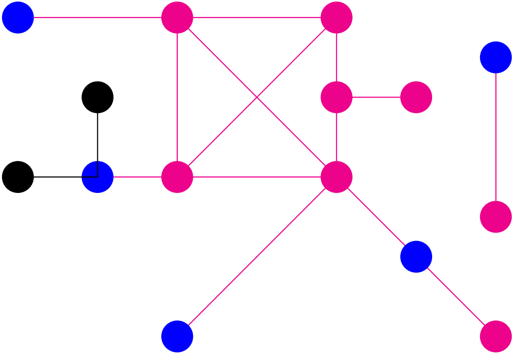












Let $A(t) := |\mathcal{A}(t)|$, the number of active vertices at time t. Since $|\mathcal{Z}(t)| = t$ and $\mathcal{Z}(t) \subseteq \mathcal{A}(t)$ for $t = 0, \ldots, T$, we also have

$$T = \min\{t \geq 0 : A(t) = t\} = \min\{t \geq 0 : A(t) \leq t\}.$$

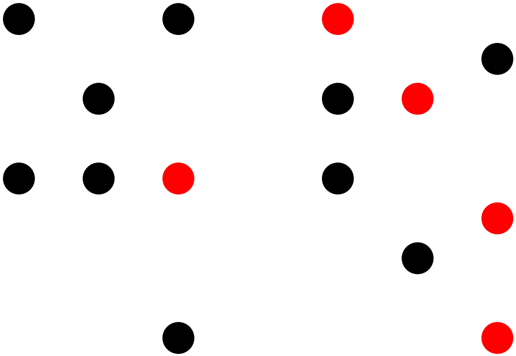
Moreover, since the final active set is $\mathcal{A}(T) = \mathcal{Z}(T)$, its size $|\mathcal{A}^*|$ is

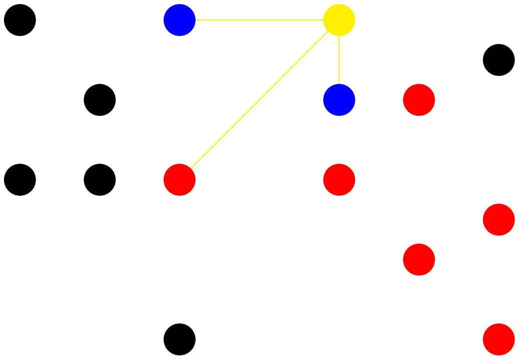
$$|\mathcal{A}^*| := A(T) = |\mathcal{A}(T)| = |\mathcal{Z}(T)| = T.$$

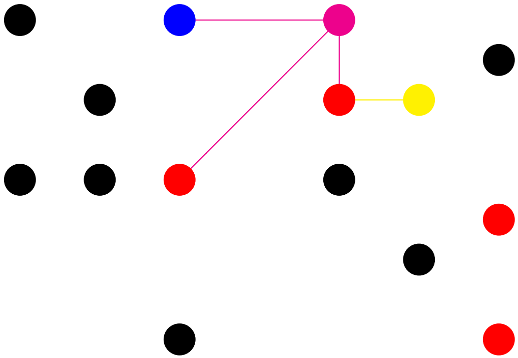
Hence, the set $\mathcal{A}_0$ percolates if and only if $T = n$, and $\mathcal{A}_0$ almost percolates if and only if $T = n - o(n)$.

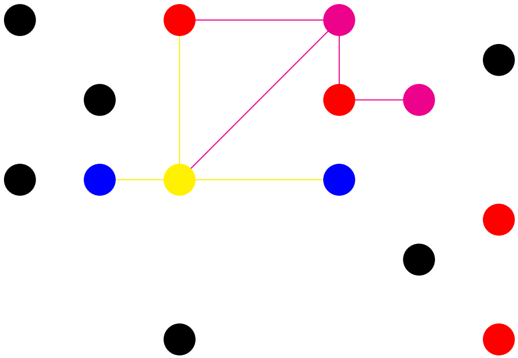
Analysis

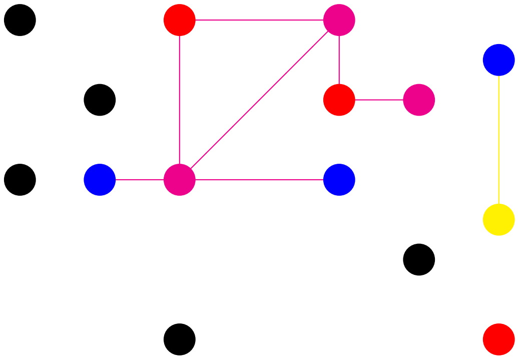
We use the standard method of revealing the edges of the graph $G(n, p)$ only on a need-to-know basis:
We begin by choosing u_1 as above and then reveal its neighbours; we then find u_2 and reveal its neighbours, and so on.

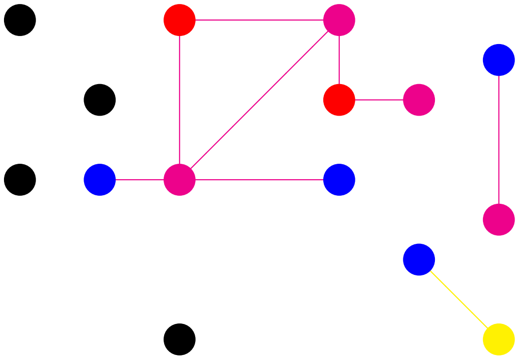


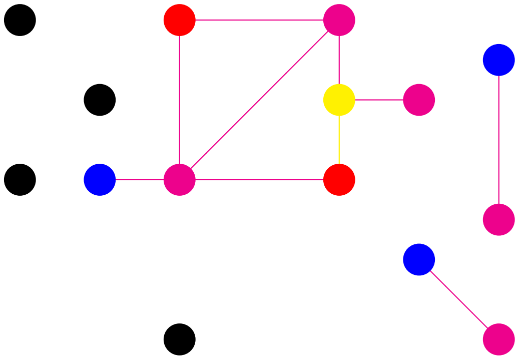


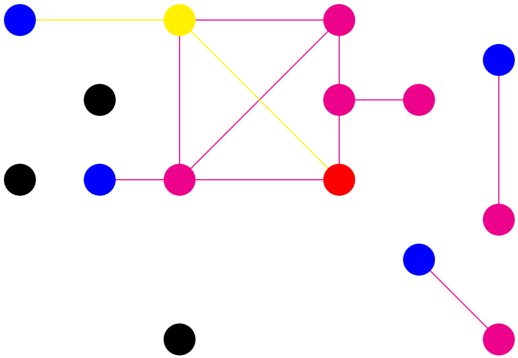


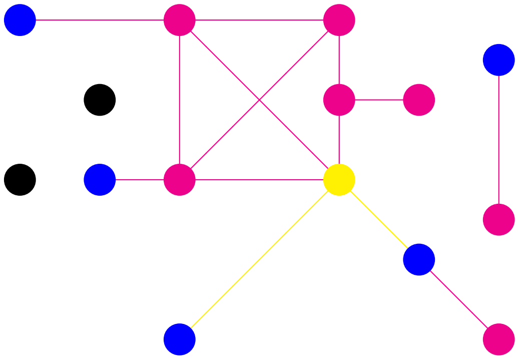


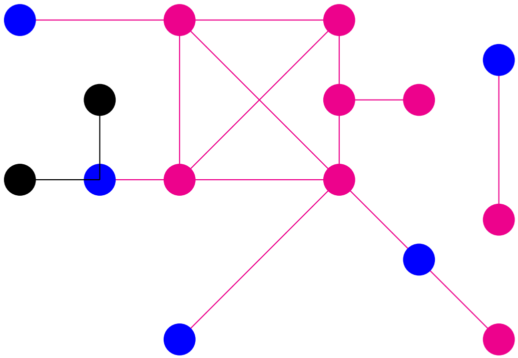












Analysis (cont.)

Let, for $i \notin \mathcal{Z}(s)$, $I_i(s)$ be the indicator that there is an edge between the vertices u_s and i. This is also the indicator that i gets a mark at time s, so if $M_i(t)$ is the number of marks i has at time t, then

$$M_i(t) = \sum_{s=1}^{t} I_i(s),$$

at least until i is activated (and what happens later does not matter).

- If $i \notin \mathcal{A}(0)$, then, for every $t \leq T$, $i \in \mathcal{A}(t)$ if and only if $M_i(t) \geq r$.
- The random indicators $I_i(s)$ are i.i.d. $\mathrm{Be}(p)$.

We have defined $I_i(s)$ only for $s \le T$ and $i \notin \mathcal{Z}(s)$, but we add further (redundant) variables so that $I_i(s)$ are defined, and i.i.d. $\mathrm{Be}(p)$, for all $i \in V_n$ and all $s \ge 1$.

Then $M_i(t) = \sum_1^t I_i(s)$ is defined for all $t \ge 0$, and has a binomial distribution $\mathrm{Bin}(t, p)$.

Define also, for $i \in V_n \setminus \mathcal{A}(0)$,

$$Y_i := \min\{t : M_i(t) \ge r\}.$$

If $Y_i \le T$, then Y_i is the time vertex i becomes active, but if $Y_i > T$, then Y_i never becomes active. Thus, for $t \le T$,

$$\mathcal{A}(t) = \mathcal{A}(0) \cup \{i \notin \mathcal{A}(0) : Y_i \le t\}.$$

Further, each Y_i has a negative binomial distribution $\mathrm{NegBin}(r, p)$:

$$\mathbb{P}(Y_i = k) = \mathbb{P}\big(M_i(k-1) = r-1,\, I_i(k) = 1\big) = \binom{k-1}{r-1} p^r (1-p)^{k-r};$$

moreover, these random variables Y_i are i.i.d.

We let, for $t = 0, 1, 2, \dots,$

$$S(t) := |\{i \notin \mathcal{A}(0) : Y_i \leq t\}| = \sum_{i \notin \mathcal{A}(0)} \mathbf{1}[Y_i \leq t],$$

so

$$A(t) = A(0) + S(t) = S(t) + a.$$

and

$$T = \min\{t \geq 0 : S(t) + a \leq t\}$$

It thus suffices to study the stochastic process $S(t)$.

Note that $S(t)$ is a sum of $n-a$ i.i.d. processes $\mathbf{1}[t \geq Y_i]$, each of which is 0/1-valued and jumps from 0 to 1 at time Y_i, where Y_i has the distribution $\mathrm{NegBin}(r,p)$.
(In other words, $S(t)/(n-a)$ is the empirical distribution function of $\{Y_i\}$.)

The fact that $S(t)$, and thus $A(t)$, is a sum of i.i.d. processes makes the analysis easy; in particular, for any given t,

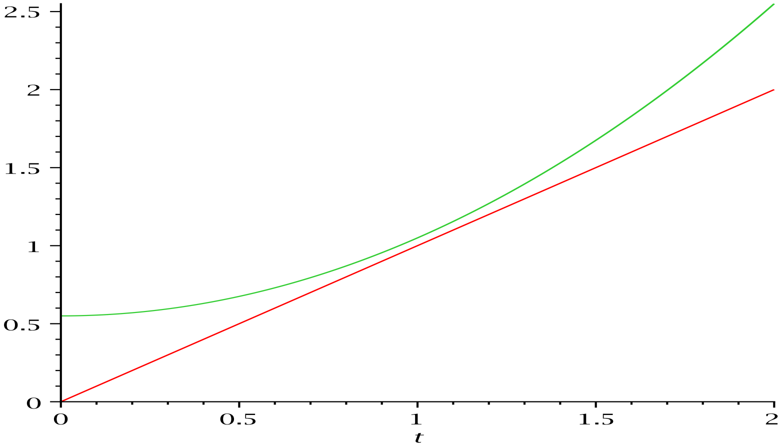
$$S(t) \sim \mathrm{Bin}\big(n-a, \pi(t)\big),$$

where

$$\pi(t) := \mathbb{P}(Y_1 \leq t) = \mathbb{P}(M_1(t) \geq r) = \mathbb{P}\big(\mathrm{Bin}(t,p) \geq r\big).$$

In particular, we have

$$\mathbb{E}\, S(t) = (n-a)\pi(t).$$



$\mathbb{E}\,S(t)$ and t (in units t_{c})

Random regular graphs

Suppose that G is a regular graph, where each vertex has degree d. The set of inactive vertices is obtained by first deleting the set $\mathcal{A}_0$ from the vertex set, and then sucessively eliminating every vertex that does not have at least $k = d - r + 1$ surviving neighbours. The final result is the *k-core* of $G \setminus \mathcal{A}_0$, the largest subgraph of $G \setminus \mathcal{A}_0$ where each vertex has degree $\geq k$.
In particular, $\mathcal{A}_0$ percolates if and only if the k-core of $G \setminus \mathcal{A}_0$ is empty.

Let G be a random d-regular graph with n vertices and let each vertex be initially infected with probability q (independently). Let $n \to \infty$. Assume $2 \le r \le d-2$.

Theorem (Balogh and Pittel)

Let

$$q_c := 1 - \inf_{0<p\le 1} \frac{p}{\mathbb{P}\big(\mathrm{Bi}(d-1, 1-p) \le r-1\big)}.$$

(i). *If $q > q_c$, then w.h.p. all vertices become activated.*

(ii). *If $q < q_c$, then w.h.p. a positive fraction of the vertices remain inactive: $(n - |\mathcal{A}^*|)/n \xrightarrow{\mathrm{p}} c > 0$.*

(Consistent with branching process approximation, but easier proved by similar methods as for $G(n,p)$.)