Lecture notes for minicourse

by Laura Eslava.

X EPPE, Mexico

August, 2022

Critical Percolation and the Emergence of the Giant Component
Rather: Some of their associated Branching Processes.

Part 1: Introduction

- Bond Percolation
- Critical Percolation ($\mathbb{Z}^d$)
- Branching Processes
- Exploration of clusters

Part 2: Erdös-Rényi Graphs: Emergence of the giant.

- Erdös-Rényi graph process.
- Subcritical, Upper bound proof
- Subcritical, lower bound proof
- Supercritical, key ideas.

Part 3: Hypercubic graphs: Heuristic for its critical probability

- Setup for generalized branching process
- (Unknown) properties and threshold

Wishlist:
- Proof of corollary for $p_c(d)$.
- Key ideas for the analysis of thms.

Other models: Their associated branching processes

Part 4:
- Erdös-Rényi: k-core emergence.
- d-processes: giant component emergence.

References:

Part 1:

* Grimmett, Percolation, Springer 1980
- Steif, A minicourse on percolation, Lecture Notes, 2009
- Hofstad, Chapter 3 of Random Graphs and Complex Networks Vol 1. Cambridge Univ. Press. 2016

Part 2:

* Svante, Tomasz, Andrzej, Random Graphs, Wiley 2000
- Hofstad, Chapter 4 of Random Graphs and Complex Networks Vol 1. Cambridge Univ. Press.

Part 3:

- Eslava, Penington, Skerman, Survival for a Galton-Watson tree with cousin mergers, Procedia Computer Science, 2021.
- Eslava, Penington, Skerman, A branching process with deletions and mergers that matches the threshold for hypercube perc. arxiv:2104.04407
* Heydenreich, Hofstad, Progress in high-dimensional percolation and random graphs. Springer 2017.

Part 1

General Assumptions: $G=(V,E)$ is connected and transitive

$E \subseteq \{\{u,v\} : u,v \in V\}$

$\{u,v\}$ also written $uv = e$

$e \in E$: $u \sim v$ neighbors,
e is incident to v
v is the endpoint of e

Examples.

Complete: $K_n = ([n], \{uv : u,v \in [n]\})$

$[n] = \{1,2,\dots n\}$

Hyperc. Lattice: $\mathbb{Z}^d = (\mathbb{Z}^d, \{uv : |u-v| = 1\})$

nearest neighbors
$\omega = (\omega_1, \dots, \omega_d)$
$|\omega| = \sum_{i=1}^{d} |\omega_i|$

Hypercube: $Q_d = (\{0,1\}^d, \{uv : |u-v| = 1\})$

Note: Transitive graphs have constant degree Ω.

Definition of connected component

$$C(v) = C_G(v) = \{w \in V : u \longleftrightarrow w \text{ in } G\}$$

$u \longleftrightarrow w$ if there is a path in G connecting u and w.

Transitive graphs

$\forall u,v \in V$ there is automorphism $\varphi : V \to V$ $\varphi(u) = v$ that maps edges into edges.

Bond Percolation process

generally fixed

For $G=(V,E)$, $p\in[0,1]$ let $G_p=(V,E_p)\subseteq G$ such that

underlying graph

$\downarrow$

may also be random

$e\in E$: $e\in E_p$ independently with prob. p

$\hookrightarrow$ e is open (otherwise closed)

What about the size of connected components in G_p ?

$C(v)=C_{G_p}(v)=\{w:$ there is a path of open edges connecting v to $w\}$

$\rightarrow$ A Property/Event $\mathcal{P}$ is increasing if for $H_1\subseteq H_2\subseteq G$

$\subseteq\{H: H\subseteq G\}$

$$H_1\in\mathcal{P}\Rightarrow H_2\in\mathcal{P}$$

Exercise 1:
- $\{H$: H contains a triangle$\}$ is increasing
- $\{H$: H has no cycles$\}$ is decreasing
- $\{H$: $|C_H(v)|=k\}$ is neither inc/decreas.
- $\{H$: $|C_H(v)|\geq k\}$ is increasing

If G is finite then the probability space may be

$$\left(\{0,1\}^{|E|},\ \mathcal{P}(\{0,1\}^{|E|}),\ \mathbb{P}\right) \qquad \mathbb{P}(\omega)=p^{\text{open}}(1-p)^{\text{closed}}$$

open $=\#\,\omega_i=1$
closed $=\#\,\omega_i=0$

For G infinite we may extend such probability spaces (of independent coin-flips) to an infinite one.

- Recall that $C(v)=C(v,\omega)$ where $\omega\in\Omega$ in the implicit state space of the probability space.

→ A Natural Coupling for increasing events

$(G_p)_{p\in[0,1]}$ is defined by $(U_e)_{e\in E}$ independent Unif(0,1) r.v.'s

letting $e \in E_p \iff U_e \le p$ → arrival time of e into the process

Exercise 2: If $\mathcal{P}$ is increasing, for $p_1 < p_2$

$$\mathbb{P}(G_{p_1} \in \mathcal{P}) \le \mathbb{P}(G_{p_2} \in \mathcal{P}).$$

→ Percolation probability: $\Theta(p) = \mathbb{P}(|C(v)| = \infty)$ $v \in V$ fixed.

→ Critical probability: $p_c = \sup\{p : \Theta(p) = 0\}$

Facts: $\Theta(0)=0$, $\Theta(1)=1$, $\Theta(p)$ is non-decreasing and right-continuous → Thm 2.5 in Steif's notes

---

$\Theta(p) = \lim_{k\to\infty} \mathbb{P}(\text{there is } k\text{-path starting at origin}) = \lim_{k\to\infty} g_k(p)$

$g_k(p)$ is a polynomial in p, $g_k(p) \downarrow \Theta(p)$

+ nondecreasing upper-semicont. function ⇒ is right continuous.

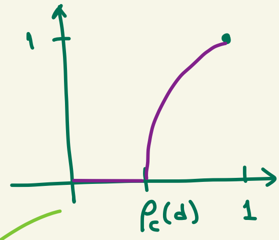
The Nearest-neighbors lattice $\mathbb{Z}^d$ Part 1-4

Write $\Theta_d(p)$ and $p_c(d)$; we use origin $\bar{0}$ as the fixed vertex.

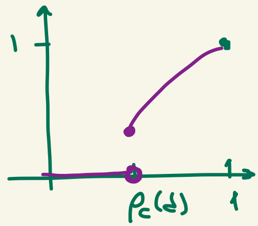
Thm 1. For $d \geq 2$, $\Theta_d(p)$ is continuous in $(p_c(d), 1)$

and $$\frac{1}{2d-1} \leq p_c(d) < 1 .$$

Exercise 3: $p_c(1) = 1$ and $p_c(d+1) \leq p_c(d)$



$p_c(d)$
?



Known for $d=2$, $d \geq 19$

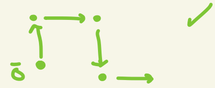
---

* so $\Theta(p_c(2)) = 0$

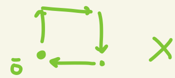
* Hara and Slade proved $\Theta(p_c(d)) = 0$ for $d \geq 19$
1994

Proof of lower bound

Let $\sigma(k) = \#$ self-avoiding paths in $\mathbb{Z}^d$ of length k starting at $\bar{o}$



$P_k = \#$ " " in $\mathbb{Z}^d_p$



Since $\{|C(\bar{o})| = \infty\} = \bigcap_{k=1}^{\infty} \{P_k \geq 1\}$, $\Theta_d(p) = \lim_{k\to\infty} \mathbb{P}(P_k \geq 1)$

We prove that if $p < \frac{1}{2d-1}$ then $\mathbb{P}(P_k \geq 1) \longrightarrow 0$ as $k \to \infty$.

$$\mathbb{P}(P_k \geq 1) \leq \mathbb{E}[P_k] = p^k \sigma(k) \leq p^k \cdot 2d\,(2d-1)^{k-1}$$

Markov's ineq

$$= \frac{2d}{2d-1}\left(p\,(2d-1)\right)^k = \left(p\,(2d-1) + o(1)\right)^k$$

as $k \to \infty$ $\square$

$$\Theta_d(p) = \lim_{k\to\infty} \mathbb{P}(P_k \geq 1) \leq \liminf_{k\to\infty} \left(p(2d-1) + o(1)\right)^k \quad \text{if } p(2d-1) < 1$$

$\mathbb{P}(P_k \geq 1) \searrow \Theta_p(p)$

Proof Sketch of upper bound

It suffices to prove $1 > p_c(2) \geq p_c(d)$ $\quad d \geq 3$

$$\Theta_2(p) \geq 1 - \sum_{k=1}^{\infty} \underbrace{(1-p)^k}_{\text{k closed edges}} \underbrace{k\,\sigma(k)}_{\text{bound on dual cycles length k}} \geq \frac{1}{2} \quad \text{for } p \text{ close to } 1.$$

$\mathbb{Z}^2$ duality argument:

$|C(0)| < \infty \iff \bar{0}$ surrounded by closed cycle

$\square$

- if $(1-p)(2d-1) < 1$ equiv. $p > 1 - \frac{1}{2d-1}$ then

$$\sum_{k=1}^{\infty} (1-p)^k\, k\, \sigma(k) \leq \sum_{k=1}^{\infty} k \Big((1-p)(2d-1) + o(1) \Big)^k < \infty$$

- Continuity of $\Theta_d(p)$ in Section 4 of Steif's notes.

Remarks

- $\Theta_d(p)$ continuity on $(p_c(d), 1)$ uses uniqueness of infinite cluster.
- Factor $(2d-1)$ may be replaced by $\lambda(d)^{-1}$

  where $\lambda(d) = \lim_{k\to\infty} \sigma(k)^{1/k}$

- For $\mathbb{Z}^2$ $\quad 1/3 \le p_c(2) \le 1/2$

  Harris-Kesten duality argument in 1960 equality 20 years later!

- Hofstad, Slade 2005 $\quad p_c(d) = \frac{1}{2d-1} + \frac{5}{2(2d-1)^3} + O\left((2d-1)^{-4}\right)$ as $d\to\infty$
- Kesten 1988 obtained first-order term: $\frac{1}{2d}$

Subject to time.

- Expansion for $d=2$ would give $p_c(2) \approx .42$
- $\lambda(d)$ is known as connective constant.

Critical percolation for spherically sym. trees

BONUS 1

Consider a tree T with root r and a_0 children each of which has a_1 children, and vertices in generation k have a_k children.

Thm* Let $A_k = \#$ vertices in generation k (this case $A_k = \prod_{i=0}^{k-1} a_i$)

Then $$p_c(T) = \frac{1}{\left(\liminf_{k\to\infty} A_k^{1/k}\right)}$$

Proof of lower bound: Essentially the same proof as for $\mathbb{Z}^d$

$$\mathbb{P}(|C(\rho)| = \infty) \leq \mathbb{E}[\# \text{ paths to gen } k \text{ from } \rho] = p^k A_k,$$

if $p < \left(\liminf A_k^{1/k}\right)^{-1}$, $\exists$ subsequence k_ℓ for which $A_{k_\ell} p^{k_\ell} \to 0$.

If $p < \left(\liminf A_k^{1/k}\right)^{-1}$ then $\liminf_{k\to\infty} (A_k^{1/k} p) < 1$

* Proof for general trees by Lyons in early 90's

Proof of upper bound via 2nd moment.

BONUS 2

Let $X_k = \#$ vertices in gen k connected to ρ $\qquad \mathbb{P}(X_k>0) \ge \dfrac{\mathbb{E}[X_k]^2}{\mathbb{E}[X_k^2]}$

so it suffices to obtain $C>0$ such that

$$\mathbb{E}[X_k]^2 \ge C\,\mathbb{E}[X_k^2] \qquad \text{for any } k\in\mathbb{N}$$

or suff. large

We just computed $\mathbb{E}[X_k] = p^k A_k$. Let $P_{u,w} = \mathbb{E}[\mathbb{1}_{\{u\leftrightarrow\rho,\, w\leftrightarrow\rho\}}]$

then

$$\mathbb{E}[X_k^2] = \sum_{\substack{u,w \text{ in} \\ \text{gen } k}} P_{u,w} = \sum_{\substack{u,w \text{ in} \\ \text{gen } k}} p^{2k-m_{u,w}}$$

where $m_{u,w}$ is the level at which u and w split.

$\le \dfrac{A_k}{A_\ell}$

$$\le A_k\, p^{2k} \sum_{\ell=0}^{k} \sum_{\substack{w \text{ in gen } k \\ \text{with split} \\ \text{at } \ell}} p^{-\ell} \le A_k^2\, p^{2k} \sum_{\ell=0}^{k} \frac{1}{(p\,A_\ell^{-1/\ell})^\ell}$$

That is,

$$\mathbb{E}[X_k^2] \le \mathbb{E}[X_k]^2 \sum_{\ell=0}^{\infty} (p\,A_\ell^{-1/\ell})^{-\ell}$$

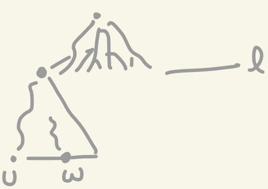
If $p > (\liminf_{k\to\infty} A_k^{-k})$ then the series converges (yields $C>0$). $\square$

exponential decay $\le \dfrac{1}{1-\delta}$

$$\mathbb{P}(|C(T)|=\infty) = \bigcap_{k=1}^{\infty} \mathbb{P}(X_k \ge 1) = \lim_{k\to\infty} \mathbb{P}(X_k\ge 1) \ge C > 0$$

crux

$\#$ w in gen k with split at ℓ from u $\le \dfrac{A_k}{A_\ell}$:



inequality since $w \neq u$.

If $p > (\liminf_{k\to\infty} A_k^{1/k})^{-1}$ then $\liminf_{k\to\infty} (p^k A_k)^{1/k} > 1+\delta$ so

$$\frac{1}{p^k A_k} \le (1+\delta)^{-k}$$

Branching Process $(Z_k)_{k\geq 1}$ with offspring dist ξ Part 1_8

Let $Z_0 = 1$, $Z_{k+1} = \sum_{l=1}^{Z_k} \xi_l^{(k)}$ where $(\xi_l^{(k)})_{k,l\geq 1}$ are iid $\sim \xi$

an initial ♀   # of ♀'s at generation k+1   all individuals reproduce identically and independently.

Do the lineage of the initial ♀ survives forever?

Extinction probability: $\eta = \mathbb{P}(\exists n\geq 1 : Z_n = 0)$

Exercise 4: η is a fixed point of $G_\xi(s) = \sum s^k P(\xi = k)$.

Thm 2. If $\mathbb{P}(\xi = 1) \neq 1$ then

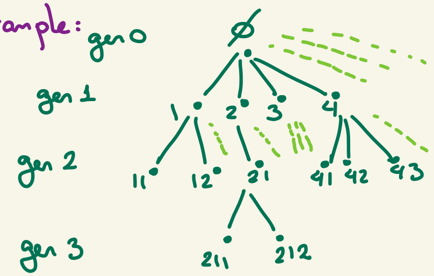
$$\mathbb{E}[\xi] \leq 1 \Rightarrow \eta = 1$$
$$\mathbb{E}[\xi] > 1 \Rightarrow \eta < 1.$$

$$\mathbb{P}(\text{extinction}) = \sum_{k=0}^{\infty} \mathbb{P}(Z_1 = k)\, \mathbb{P}(\text{extinction of } Z_1 \text{ independent subfamilies} \mid Z_1 = k)$$

Note that $G_\xi(1) = 1$ so conclusion in thm 2 follows from showing that η is smallest fixed point (there are exactly two if $\mathbb{E}[\xi] > 1$ and exactly one if $\mathbb{E}[\xi] \leq 1$ but $\mathbb{P}(\xi = 1) \neq 1$; otherwise $G(s) = s$).

The Genealogy tree T (Embedded in Ulam-Harris tree) Part 1_9

An example:



e.g. Individual $\overset{u}{42}$ lives in gen $\overset{|u|}{2}$

it has 4 older relatives in generation 2 and has $\xi_5^{(2)}$ children named $421, \dots, 42\xi$

(or no children if $\xi = 0$)

Obs. From T we recover $(Z_k)_{k\geq 1}$; $|T| = \sum_{k=0}^{\infty} Z_k$.

The indexing $\xi_\ell^{(k)}$ suggests a construction of T through a Breadth-first-search process

---

Algorithm 1: Construction of T $(A_m, U_m)_{m\geq 0}$

Sequentially sample the number of children of each $\mathfrak{f}$ from $(\xi_m)_{m\geq 1}$

In-queue vertices: A_m $\qquad A_0 = \{\emptyset\}$ $\qquad S_m = |A_m|$

Used/explored vertices: $\mathcal{U}_m$ $\qquad \mathcal{U}_0 = \emptyset$ $\qquad |U_m| = m$

→ At step m: Select $v_m \in A_{m-1}$ Create $v_m 1, \dots, v_m \xi_m$ children for v_m

$$A_m = A_{m-1} \cup \{v_m 1, \dots, v_m \xi_m\} \setminus \{v_m\} \qquad \mathcal{U}_m = \mathcal{U}_{m-1} \cup \{v_m\}$$

→ Stop when $A_m = \emptyset$.

Selection: Depth-First S. if v_m is lexicographically smallest.
Breadth-First S. " " length, then lexic. smallest.

Since $(\xi_m)_{m\geq 1}$ are iid. the choice of v_m does not affect the law of $(A_m, \mathcal{U}_m)_{m\geq 0}$ but it does affect how we recover T from $(S_m)_{m\geq 0}$.

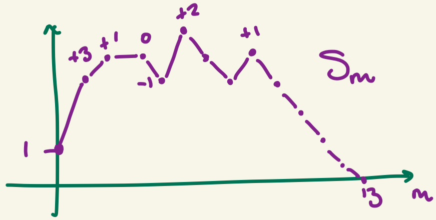
Exploration's Random Walk $(S_m)_{m\geq 1}$

This is defined by $S_0 = 1$ $\quad S_m - S_{m-1} = \underbrace{\xi_m - 1}$

steps of R.Walk. have ≥ -1

$$|T| = \inf\{m : A_m = \emptyset\} = \inf\{m : S_m = 0\}$$

If selection rule is explicit then we recover T from $(S_m)_{m\geq 0}$



Exercise 5:

$$\mathbb{P}(|T| = n) = \frac{1}{n}\mathbb{P}\left(\sum_{m=1}^{n}\xi_m = n-1\right)$$

Algorithm 2 Exploration of $C(v)$ in $G=(V,E)$

Sequentially explore the number of 'children' of each vertex

In-queue vertices : A_m $\qquad A_0 = \{v\}$

Used vertices : $\mathcal{U}_m$ $\qquad \mathcal{U}_0 = \emptyset$

→ At step m : Select $v_m \in A_{m-1}$, let Γ_m be its neighbors in G

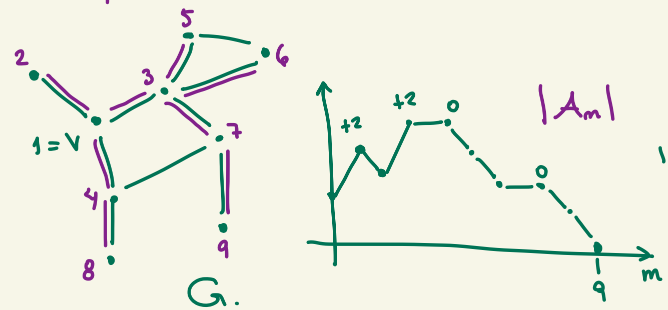
$$A_m = A_{m-1} \cup \left(\Gamma_m \setminus (A_{m-1} \cup \mathcal{U}_{m-1})\right) \setminus \{v_m\}$$

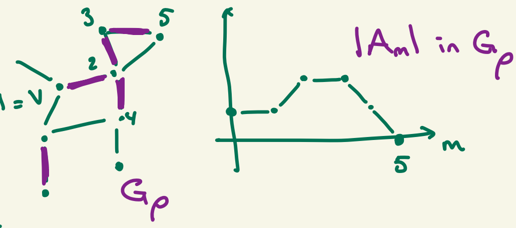
$$\mathcal{U}_m = \mathcal{U}_{m-1} \cup \{v_m\}$$

→ Stop when $A_m = \emptyset$

Then $(A_m, \mathcal{U}_m)_{m \geq 0}$ recovers a spanning tree of $C(v)$ and $|C(v)|$

Example: * Vertices labeled in order of exploration





Exploration in G_ρ : Replace $\Gamma_m \setminus (A_{m-1} \cup U_{m-1})$ with $\Gamma_m^{open} \setminus (A_{m-1} \cup U_{m-1})$

conditional on (A_{m-1}, U_{m-1}) $|\Gamma_m^{open} \setminus U_{m-1}| \overset{d}{=} \mathrm{Bin}(|\cdot|, \rho)$

- Edges that close cycles are not relevant to counting the number of vertices in the current explored component.

- In G_ρ we can 'sample' the edges as we explore $C(v)$. This means that we don't sample / generated beyond $C(v)$ and its boundary edges.

A Branching-Process proof for $\frac{1}{2d-1} \leq p_c(d)$

When exploring $C(\bar{o})$ with Algorithm 2, $|\Gamma_m| = 2d$ and $m \geq 2$

$$|\Gamma_m^{open} \setminus \mathcal{U}_{m-1}| \preceq_{st} \text{Bin}(2d-1, p)$$

needs more 'coin-flips'

$\mathbb{P}(X \geq a) \leq \mathbb{P}(Y \geq a)\ \forall a \in \mathbb{R}$

$X \leq_{st} Y$

then $|C(\bar{o})| \preceq_{st} |T| + 1$

where T is the genealogy tree of a BP with offspring $\xi \overset{d}{=} \text{Bin}(2d-1, p)$; if $p < \frac{1}{2d-1}$ then $|T| < \infty$ a.s.

Part 2

Part 2 _ 1

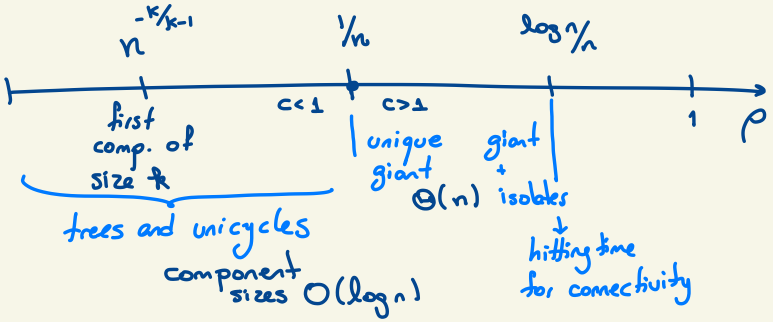
Erdös-Rényi Graph Process. $(G(n,p))_{p\in[0,1]}$

$G(n,p) = ([n], E_p)$ such that $e \in E_p \iff U_e \le p$ — iid unif(0,1)

New parameter: $n = |V|$

A property $\mathcal{P}$ holds a.a.s. (whp) if $\mathbb{P}(G(n,p) \in \mathcal{P}) \to 1$, $n \to \infty$

A brief story of thresholds: (a.a.s.)



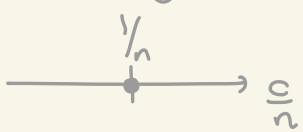
* If $p = \frac{1}{n}$ the components are $\Theta(n^{2/3})$ — not this week

---

= What does it mean to be 'infinite'/giant? =

* Hamiltonicity threshold: Path $\leftrightarrow$ mindeg 1
  Cycle $\leftrightarrow$ mindeg 2

*



the critical window is invisible in this scale $\frac{c}{n}$.

The critical window has width $\Theta(n^{-4/3})$, larger than n^{-2}.

Critical point $\frac{1}{n}$ is equivalent to $\frac{1}{n-1}$:

if $p = \frac{1+\varepsilon}{n}$ then $p = \frac{1+\varepsilon'}{n-1}$ $\quad \varepsilon' = \varepsilon + O(n^{-1})$
$= \varepsilon + o(n^{-1/3})$

$1+\varepsilon' = (1+\varepsilon)(1 - 1/n) = 1 + \varepsilon - \frac{1}{n}(1+\varepsilon)$

GIANT = VISIBLE on the scale of $|V| = n$. Part 2 _ 2

Scaling $p = \frac{c}{n}$ makes # neighbors of 1 $\overset{d}{=} \text{Bin}(n-1, \frac{c}{n}) \underset{n\to\infty}{\approx} \text{Poi}(c)$,

$\mathbb{E}[\text{\# neighbors of } 1] \approx c$

Branching Process' heuristic suggests the threshold lies at $c=1$.

Thm 3: For $c<1$ let $I_c = c - 1 - \log c > 0$ then, in $G(n, \frac{c}{n})$, all components are $O(\log n)$ a.a.s and

largest component

Large Dev. Rate Function. for Poi(c) r.v.

$$\frac{|\mathcal{C}_{max}|}{\log n} \xrightarrow{\mathbb{P}} I_c^{-1} \quad \text{as } n\to\infty$$

---

* $C_{max} = \max_{v\in[n]} |C(v)|$ is well defined.

whereas $\mathcal{C}_{max}$ = largest component containing smallest labelled vertex.

needs to break ties (if any).

* Thm 3 $\forall \varepsilon > 0\ \exists\ \delta = \delta(\varepsilon, c)$ s.t

$$\mathbb{P}\left(\left||C_{max}| - I_c^{-1}\log n\right| > \varepsilon \log n\right) \le O(n^{-\delta})$$

Exercise 1: If $Z_k = \sum_{v \in [n]} \mathbb{1}_{\{|C(v)| = k\}}$ then # components of size k $= \frac{Z_k}{k}$

$$\{C_{max} \geq k\} = \{Z_{\geq k} \geq k\} \quad \text{where} \quad Z_{\geq k} = \sum_{v \in [n]} \mathbb{1}_{\{|C(v)| \geq k\}}$$

Exercise 2: Verify that it suffices to prove that if $k = \lfloor a \log n \rfloor$

Upper bound: $a > I_c^{-1}$ then $Z_{\geq k} = 0$ a.a.s.

Lower bound: $a < I_c^{-1}$ then $Z_{\geq k} \geq 1$ a.a.s.

We will use the first and second moment method.

Proof of $Z_{\geq k} = 0$ a.a.s. for $k > I_c^{-1} \log n$

Let T be a $\mathrm{Bin}(n, \frac{c}{n})$ branching process. We will show that

$$\mathbb{P}(|C(1)| > k) \overset{i)}{\leq} \mathbb{P}(|T| > k) \overset{ii)}{\leq} e^{-k I_c}$$

Then $\mathbb{P}(Z_{\geq k} \geq 1) \leq \mathbb{E}[Z_{\geq k}] = n\,\mathbb{P}(|C(1)| \geq k)$

$$\leq n^{1 - a I_c} \longrightarrow 0,$$

as $n \to \infty$ if $a > I_c^{-1}$

Couplings to RW's Recall the random walk exploration of $C(1)$: $S'_0 = 1$ for $m \geq 1$

$$S'_m - S'_{m-1} \stackrel{d}{=} \mathrm{Bin}(n-1-x_m, \tfrac{c}{n}) - 1 \preceq_{st} \mathrm{Bin}(n, \tfrac{c}{n})$$

So, if T is a $\mathrm{Bin}(n, \frac{c}{n})$ branching Process then

$$|C(1)| \preceq_{st} |T| \longrightarrow \text{implies i).}$$

Exercise 3: If $k \in \mathbb{N}$ and T' is a $\mathrm{Bin}(n-k, \frac{c}{n})$ b.p. then

$$\mathbb{P}(T' \geq k) \leq \mathbb{P}(|C(1)| \geq k) \longrightarrow$$ if $k = o(n)$ then $|T|$ and $|T'|$ are close.

- We may couple T and T' a $\mathrm{Poi}(c)$ Branching P. so that $\mathbb{P}(|T| > k) = \mathbb{P}(|T'| > k) + e_{n,c}(k)$

with $|e_{n,c}(k)| \leq \frac{c^2}{n} \sum_{s=1}^{k-1} \mathbb{P}(T' \geq s)$ or $|e_{n,c}(k)| \leq \frac{kc^2}{n}$

see Thm 3.20 RGCN1 Ch. 3.7.

Technique uses for lower bound on $\mathbb{E}[Z_{\geq k}]$ $k > I_c^{-1} \log n$

- **Coupling for exercise:** Up to verifying that $|C(1)| \geq k$ ($|A_{m-1} \cup U_{m-1}| < k$) all explored vertices have at least $|\Gamma_m \setminus (A_{m-1} \cup U_{m-1})| \geq n-k$ edges to be tested

so we use these first $n-k$ coin-flips for reproduction in T' and the remaining coin-flips makes $S_m \preceq_{st} S'_m$ for as long as $|T'| < k$.

Proof of ii) For the construction of T,

$$S_0 = 1 \qquad S_m = S_{m-1} + \xi_m - 1 = \sum_{\ell=1}^{m} \xi_\ell - (m-1)$$

$(\xi_\ell)_{\ell \geq 1}$ iid $\sim \mathrm{Bin}(n, \frac{c}{n})$

$$|T| = \inf \{ m : S_m = 0 \}$$

then $\{ |T| > k \} \subseteq \{ S_k > 0 \} = \{ \sum_{\ell=1}^{k} \xi_\ell \geq k \}$ $\quad \sum_{\ell=1}^{k} \xi_\ell \sim \mathrm{Bin}(nk, \frac{c}{n})$

use the large deviations rate: since $c < 1$,

$$\mathbb{P}(\mathrm{Bin}(nk, \tfrac{c}{n}) \geq k) \leq e^{-kI_c}$$

$\mathbb{P}(T' \geq k) \geq e^{-kI_c(1+o(1))}$

$\square$

---

* $\mathbb{P}(\mathrm{Bin}(m, p) \geq ma) \leq e^{-m I_p(a)}$ if $p < a \leq 1$

$$I_p(a) = p - a - a \log\left(\frac{p}{a}\right)$$

take $m = nk$, $p = c/n$, $a = 1/n$

then $m I_p(a) = k I_c$

- Also $\mathbb{P}(\mathrm{Poi}(ck) > k) \leq e^{-kI_c}$ for $c < 1$.

Proof Sketch for $Z_{\geq k} \geq 1$ a.a.s. $k < I_c^{-1} \log n$

By Chebyshev's inequality: $\mathbb{P}(Z_{\geq k} = 0) \leq \dfrac{\mathrm{Var}(Z_{\geq k})}{\mathbb{E}[Z_{\geq k}]^2}$

Goal: Upper bound for $\mathrm{Var}(Z_{\geq k})$ of same order as

$$\mathbb{E}[Z_{\geq k}] = n\,\mathbb{P}(|C(1)| \geq k) \approx n^{1-aI_c}$$

Exercise 4: If $X \geq 0$ is integer valued + $\mathbb{P}(X \geq s) \leq e^{-sI_c}$ ← exponential decay.

$$\mathbb{E}[X \mathbb{1}_{\{X \geq k\}}] = k\,\mathbb{P}(X \geq k) + \sum_{s > k} \mathbb{P}(X \geq s)$$

$X = |C(v)|$ $\qquad \leq (k+A)\, e^{-kI_c}$ $\qquad$ some constant A

By Chebyshev's inequality

$$\mathbb{P}(Z_{\geq k} = 0) \leq \mathbb{P}(|Z_{\geq k} - \mathbb{E}[Z_{\geq k}]| \geq \mathbb{E}[Z_{\geq k}]) \leq \frac{\mathrm{Var}(Z_{\geq k})}{\mathbb{E}[Z_{\geq k}]^2}$$

Also: $\quad \mathbb{P}(X > 0) \geq \dfrac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}$

Ex 4: $\mathbb{E}[X \mathbb{1}_{\{X \geq k\}}] = \sum_{s=k}^{\infty} \sum_{l=1}^{s} \mathbb{P}(X = s) = \sum_{l=1}^{\infty} \sum_{s = l \vee k}^{\infty} \mathbb{P}(X = s)$

$$= k\,\mathbb{P}(X \geq k) + \sum_{l > k} \mathbb{P}(X \geq l)$$

$$\leq k\,\kappa^{-k} + \sum_{l > k} \kappa^{-l} = k\,\kappa^{-k} + \frac{\kappa^{-k+1}}{1-\kappa}$$

$$= \kappa^{-k}\left(k + \frac{\kappa}{1-\kappa}\right)$$

$\to A$, $\kappa = e^{I_c}$.

With foresight ↘

Exercise 5. Use a coupling of $(G(n,p))_{n\geq 1}$ to show

$$\mathbb{P}\left(\begin{matrix}|C(1)|\geq k\\ |C(2)|\geq k\end{matrix}, C(1)\neq C(2)\right) \leq \mathbb{P}(|C(1)|\geq k)\,\mathbb{P}(|C(2)|\geq k)$$

In what follows, change $\mathbb{P}(A\cap B)$ to $\mathbb{E}[\mathbb{1}_{\{A\}}\mathbb{1}_{\{B\}}]$.

Recall that $Z_{\geq k} = \sum_{v\in[n]} \mathbb{1}_{\{|C(v)|\geq k\}}$

and $\mathrm{Var}(Z_{\geq k}) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

There is nuance in the claim since T' has offspring $\mathrm{Bin}(n-k, \frac{c}{n})$ which may be approximated to T^{Poi} with offspring $\mathrm{Poi}(c(1-\frac{k}{n}))$ and $c(1-\frac{k}{n}) \approx c$

$$\mathrm{Var}(Z_{\geq k}) = \sum_{v\in[n]}\sum_{w\in[n]} \left(\mathbb{E}\left[\mathbb{1}_{\{|C(v)|\geq k\}}\cdot \mathbb{1}_{\{|C(w)|\geq k\}}\right] - \mathbb{E}\left[\mathbb{1}_{\{|C(v)|> k\}}\right]\mathbb{E}\left[\mathbb{1}_{\{|C(w)|> k\}}\right] \right)$$

$$\leq n \sum_{w\in[n]} \mathbb{E}\left[\mathbb{1}_{\{|C(1)|\geq k\}}\mathbb{1}_{\{w\in C(1)\}}\right]$$

$$= n\, \mathbb{E}\left[|C(1)|\,\mathbb{1}_{\{|C(1)|\geq k\}}\right]$$

Truncated Suceptibility

$$\leq (a\log n + A)\, n^{1-aI_c} \qquad \text{for some } A.$$

$\square$

Supercritical Phase: Statement and Key ideas

Thm 4. For $c > 1$, let ζ_c (zeta) satisfy $1 - \zeta_c = e^{-c\zeta_c}$

survival prob. of a Poi(c) branching proc.

then $$\frac{|C_{max}|}{n} \xrightarrow{\mathbb{P}} \zeta_c$$

A B.P. heuristic: It is likely that $|C(1)|$ is large

with prob ζ_c $\qquad$ $\approx$ infinite

then, in $G(n, \frac{c}{n})$ $\quad \mathbb{E}[\#\text{vertices in 'large' components}] \approx n\zeta_c$

* Uniqueness of C_{max} follows after 'large' is precised.

---

* Full statement: $\forall \gamma \in (1/2, 1)\ \exists\ \delta = \delta(c, \gamma)$

$$\mathbb{P}\left(\big|\, |C_{max}| - \zeta_c n \,\big| \geq n^{\gamma} \right) = O(n^{-\delta})$$

* In addition $\quad |C_i| = O(\log n)$ a.a.s $\quad \forall i \geq 2$

Duality: if d is the dual parameter of the dual distribution of Poi(c) (satisfies $de^{-d} = ce^{-c}$) $\quad d<1$

then $\quad G(n, \frac{c}{n}) \setminus C_{max} \cong G(m, \frac{d}{n})$ with $m \simeq (1-\zeta_c)n$

* Dual distribution $\quad p'_k = \dfrac{e^{-c}}{\eta} \dfrac{(\eta c)^k}{k!} = e^{-\eta c} \cdot \dfrac{(\eta c)^k}{k!}$

then $\quad d = \eta c = c \cdot e^{-c(1-\eta)} = c\, e^{-c + c\eta} \iff de^{-d} = ce^{-c}$

$0 \leq f(x) = x e^{-x}$ $\qquad f'(x) = e^{-x}(1-x)$ $\qquad$ maximum at 1.

$\qquad f''(x) = e^{-x}(x-2)$

(Naive) proof strategy: Suppose $k = k(n)$ is so large that ① $\mathbb{P}(|C(1)| \geq k) \approx \varrho_c$ large/close to infinite.

and ② $|C_{max}| \cong Z_{\geq k}$ if $\exists$ only one 'large' component.

③ $\mathbb{P}(|Z_{\geq k} - \mathbb{E}[Z_{\geq k}]| > \varepsilon n) \to 0$ as $n \to \infty$

Concentration needs upper bounds: $\mathrm{Var}(Z_{\geq k})$
lower bounds: $\mathbb{E}[Z_{\geq k}]$

then $\mathbb{P}(||C_{max}| - n\varrho_c| > \varepsilon n) \to 0$ as $n \to \infty$ ☺

Recall. Focus on ①, ②, ③ first then add other details.

Duality (Part 2)

$$\mathbb{P}(Z_1 = k \mid \text{extinction}) = \frac{1}{\eta}\mathbb{P}(Z_1 = k, \text{extinction})$$

$$= \frac{p_k}{\eta}\mathbb{P}(\text{extinction})^k = \eta^{k-1} p_k$$

For an edge uv in $G(n,p)$ conditional on $m = n - |C_{max}|$ and $u, v \notin C_{max}$

its edge probability is

$$\frac{c}{n} = \frac{c}{n} \cdot \frac{m}{m} = \frac{d}{m} \cdot \frac{cm}{dn}$$

conditional on 'knowing'

≈ 1 since $\frac{m}{n} \approx (1 - \varrho_c)$ and $c(1 - \varrho_c) = d = c\eta$

Key Estimates in the proof

We actually choose $k = K \log n$ for K suitably large!

① $\mathbb{P}(|C(1)| \geq K \log n) \overset{\text{recall}}{\approx} \mathbb{P}(|T| \geq K \log n)$
$= \mathbb{P}(|T| = \infty) + o(1/n)$

② Follows for $a < \zeta_c$ since $\mathbb{E}[Z_{\geq an} - Z_{\geq k}] \xrightarrow{n \to \infty} 0$ no middle ground!

and a.a.s. $|Z_{\geq k} - \mathbb{E}[Z_{\geq k}]| \leq n^{\varepsilon}$ $\quad \varepsilon < 1/2$

③ For concentration: $\mathrm{Var}(Z_{\geq k}) \leq (ck+1)n\, \mathbb{E}[|C(1)| \mathbb{1}_{\{|C(1)| < k\}}]$

---

③* Compare upper bounds for $\mathrm{Var}(Z_{\geq k})$ in the supercritical phase $\mathbb{1}_{\{|C(1)| \geq k\}}$ is very likely, replace with $\mathbb{1}_{\{|C(1)| < k\}}$ and logarithmic term

②* Once we know there is no middle ground then if there were more than one giant then $\mathbb{E}[Z_{\geq k}] \neq n \cdot \zeta_c$ (it would be more, say twice, as likely to be in giant-type components).

①* Important that error probability is $o(1/n)$ to be overall negligible in the next bound $\mathbb{E}[Z_{\geq k}] = \zeta_c n + o(1)$.

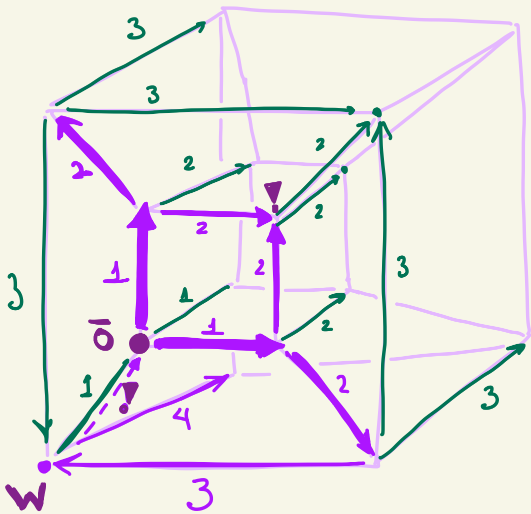
Now: All other components are of logarithmic size!!!

Part 3

PART3_1

Hipercubic graphs: Towards their geometry

Consider the following example of exploration of $Q_{d,p}$:



- Exploration starts at $\bar{0}$
- Edges are numbered according to their exploration time.

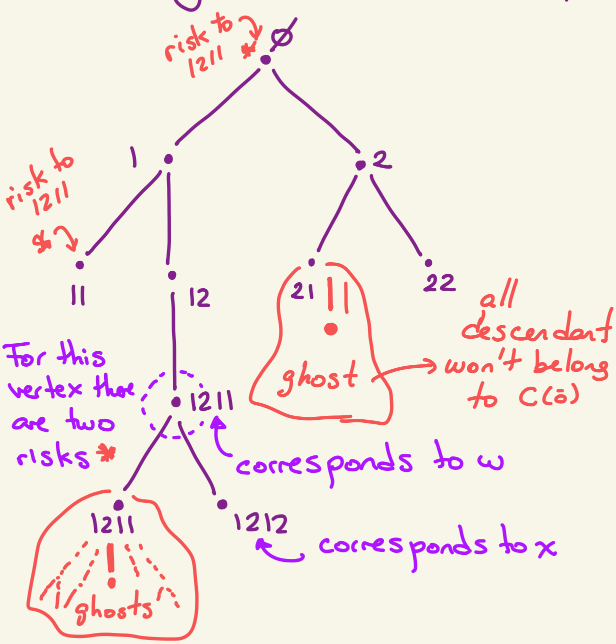
→ open edges

→ closed (and tested) edges

Obs. If an exploration could record all geometry we could avoid 'clashes' !

(there are two in example)

'Standard' coupling to branching process; Bin$(d-1,p)$



Moral:

1. Some 'cousins' should merge
2. Some children weren't born

On the lookout for a proper description of a 'good' process:

$$\rho(d-1) = 1 + p \qquad p > 0$$

so that branching has chance to survive

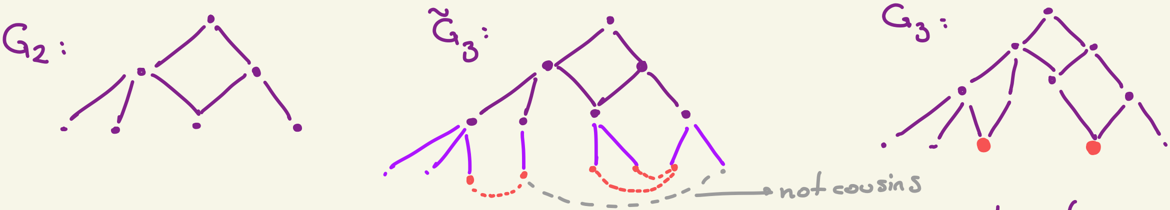
and mergers? and deletions?

could differ, in principle.

An open-problem for a workshop

To the $Poi(1+p)$ branching process incorporate merges of any pair of cousins, independently with prob q.

Create each generation with 2 steps: ① generate children ② identify individuals



Goal: Give sufficient conditions for a.s. extinction.

Thm 5 (E., P., S.) For fixed p (small): $q > \frac{1}{2}p + Cp^2$ implies

---

Good news: Relation between p and q is linear which is nice.

Bad news: This threshold does not coincide with critical percolation for $\mathbb{Z}^d$ nor Q_d (recall they coincide in at least 3 terms)

Non-backtracking walks

A random walk on a graph G is a sequence of edges e_0, e_1, e_2, e_3 $\quad e_0 = v_0 v_1,\ e_1 = v_1 v_2,\ e_2 = v_2 v_3 \dots$ such that $\mathbb{P}(v_{j+1} = v \mid v_j, \dots v_0) = \frac{1}{\deg(v_j)} \mathbb{1}\{v \sim v_j\}$

Non-backtracking if:

$$\mathbb{P}(v_{j+1} = v \mid v_j, v_{j-1}, \dots v_0) = \frac{1}{\deg(v_j) - 1} \mathbb{1}_{\{v \sim v_j,\ v \neq v_{j-1}\}}$$

Exercise 1: If $v_0, v_1, \dots v_5$ form a non-backtracking walk

then $\mathbb{P}\left(\begin{matrix}\text{walk forms}\\ \text{a 4-cycle}\end{matrix}\right) = \begin{cases} \frac{1}{(d-1)^2} & \text{if } G = Q_d \\ \frac{1}{(2d-1)^2} - \frac{1}{(2d-1)^3} & G = \mathbb{Z}^d \end{cases}$

This is a good 'analogue' for g

---

* A non-backtracking walk in Q_d boils down, at each step on selecting one of $d-1$ coordinates and 'flip' it from 0 to 1 or viceversa.

Without loss of generality $\quad (0,0,\dots\dots 0) \to (1,0,\dots\dots)$ change one coordinate

$\uparrow ? \qquad \downarrow$

$\leftarrow (1,1,\dots\dots)$ change a different one

the 3rd and 4th steps are forced to choose precisely one coordinate to close the cycle.

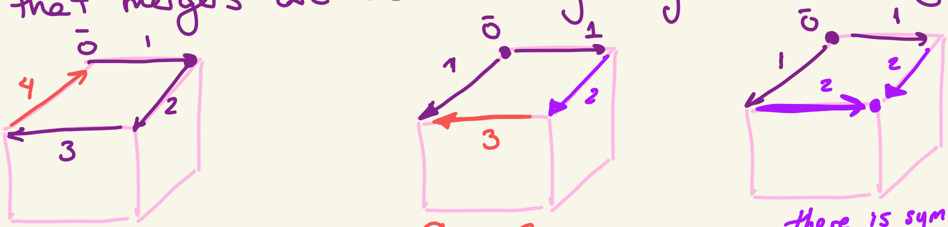
* Same argument works for $\mathbb{Z}^d$ where argument fails if the first two step were $(0,0,\dots 0) \to (1,0,\dots 0)$

$\uparrow \qquad \downarrow$

as there is no way to close back: $\times \longleftarrow (2,0,\dots 0)$

Talking to others about difficulties

Recall that mergers are not the only way of clashing!



Case 1 Case 2 Case 2 + 1/2

there is symetry!

Problem: Clashing risk depends on genealogy!

If $v \in G_n \setminus G_{n-1}$ (v lives in generation n) then

$$k_v = \# \{ w \in G_{n-1} : \text{minimal path btw } w \text{ and } v \text{ has three edges} \}$$

Q: Why exactly the non-backtracking walk help us encode our deletions?

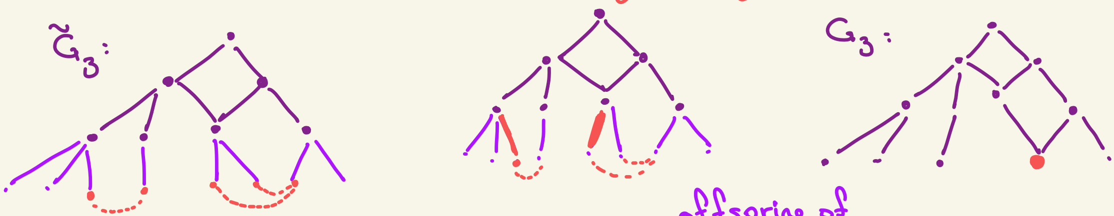
Open Question: Why doesn't the non-backtracking give heuristic for site percolation?

* In site percolation vertices (and not edges) are tested to be open/closed so edges incident in to a common vertex are not independently open/closed.

Generalized Process Remodelled (Alg 3 $B(p,q)$) PART3_6

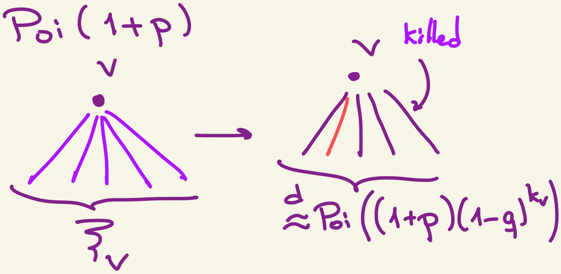
Add an intermediate step for deletions

① generate children ② Delete inhomogeneously ③ identify individuals



Close up: Given k_v and $\xi_v \overset{d}{=} \mathrm{Poi}(1+p)$ ← offspring of v

Try and kill each children k_v times, each indep. with prob q



---

* The distribution of surviving children is explicit due to the thinning property of Poisson r.v.'s

OPEN PROBLEM:

Can you find a coupling of $(B(p,q))_{p,q}$ such that $B(p,q)$ is monotone in p or q?

in terms of: generation sizes
or number of mergers
or ¿?

Finally $\ddot{O}$

Thm 6 (E., P., S.) There is $C>0$ and $p_0 \in (0,1)$ such that for $0<p<p_0$:

- $q < \frac{2}{5}p(1-Cp)$ then $B(p,q)$ survives with positive prob.
- $q > \frac{2}{5}p(1+Cp)$ then $B(p,q)$ dies out a.s.

Corollary (Heuristic for $\mathbb{Q}_{\Omega,\rho}$ or $\mathbb{Z}^{\frac{\Omega}{2}}_{\rho}$) If Ω is large enough and ρ is 'good'* then letting

$$1+p(\rho) = (\Omega-1)\rho \quad \text{and} \quad q = (\Omega-1)^{-2}$$

then $\hat{\rho}_c = \frac{1}{\Omega-1} + \frac{5}{2}\frac{1}{(\Omega-1)^3}$ is an (approx.) threshold for extinction/survival of $B(p,q)$

---

* Many more details, at several math-levels, on slides accesible from Laura Eslava's website.