Exponential growth of ponds in invasion percolation

J. Goodman

Department of Mathematics
University of British Columbia

2009 December 8, IHP Paris

Outline

1. Introduction
   - Definition and terminology
   - Relation to ordinary percolation
   - Ponds and outlets
2. Main results
   - Law of large numbers
   - Central limit theorem
   - Large deviations
   - Tail asymptotics
3. Outline of proofs
   - Markov structure of the outlet weights
   - From outlet weights to pond measurements
4. Comparison with other graphs
   - Results in 2 dimensions
   - Comparison to percolation with defects

Outline

Definition

Definition

Consider an infinite graph G with finite degrees and select a vertex o as the origin. To each edge of G associate an i.i.d. Unif[0, 1] edge weight. We may assume that the edge weights are all distinct. Inductively define a sequence of connected subgraphs (clusters) as follows: set C_0 to be the origin o, and C_i to be the subgraph obtained by adjoining to C_{i-1} the boundary edge having smallest weight. The invasion percolation cluster (IPC) is the union $C = \cup_{i=0}^{\infty} C_i$.

Illustration of invasion percolation

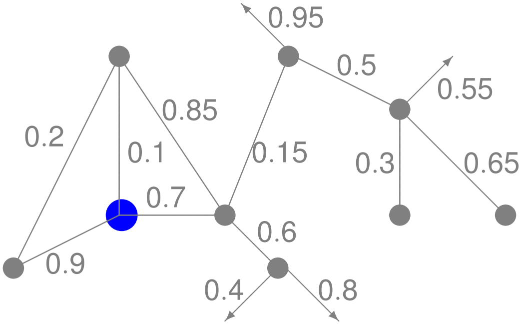

Figure: Invasion starting from the large dot. Blue: the current invasion cluster. Green: the boundary of the current cluster

Illustration of invasion percolation

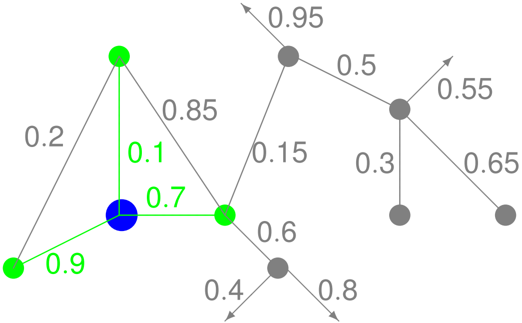

Figure: Invasion starting from the large dot. Blue: the current invasion cluster. Green: the boundary of the current cluster

Illustration of invasion percolation

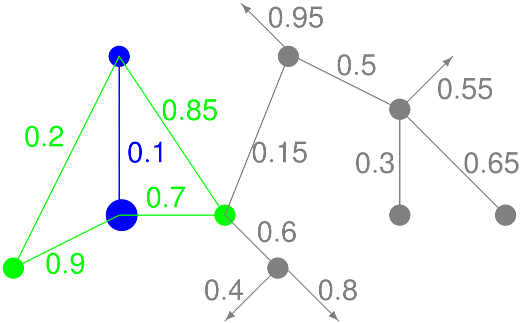

Figure: Invasion starting from the large dot. Blue: the current invasion cluster. Green: the boundary of the current cluster

Illustration of invasion percolation

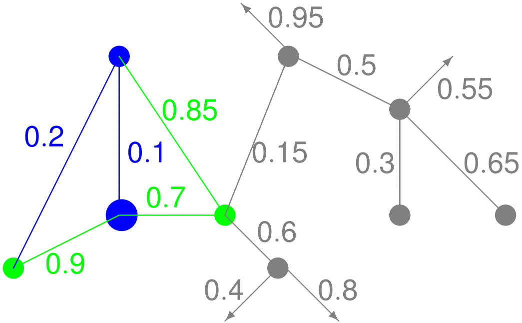

Figure: Invasion starting from the large dot. Blue: the current invasion cluster. Green: the boundary of the current cluster

Illustration of invasion percolation

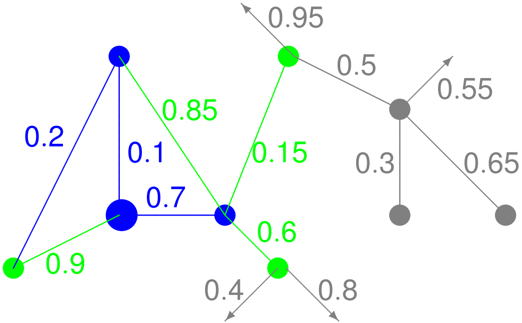

Figure: Invasion starting from the large dot. Blue: the current invasion cluster. Green: the boundary of the current cluster

Illustration of invasion percolation

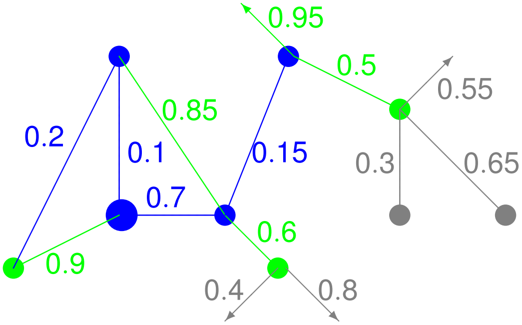

Figure: Invasion starting from the large dot. Blue: the current invasion cluster. Green: the boundary of the current cluster

Illustration of invasion percolation

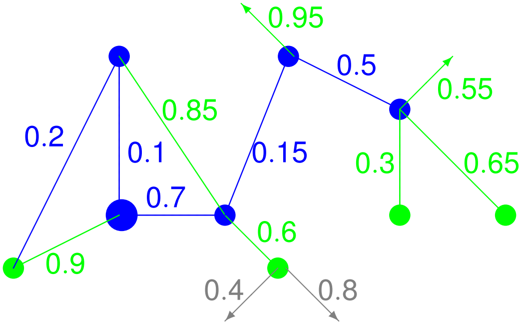

Figure: Invasion starting from the large dot. Blue: the current invasion cluster. Green: the boundary of the current cluster

Illustration of invasion percolation

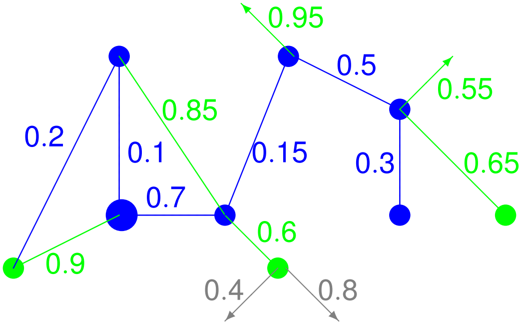

Figure: Invasion starting from the large dot. Blue: the current invasion cluster. Green: the boundary of the current cluster

Illustration of invasion percolation

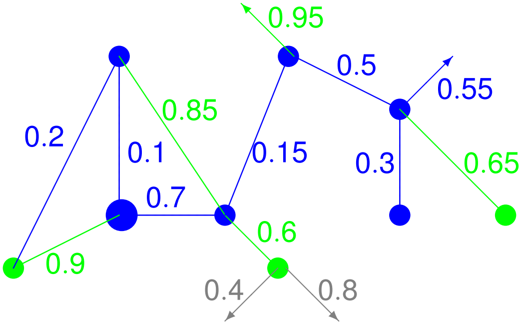

Figure: Invasion starting from the large dot. Blue: the current invasion cluster. Green: the boundary of the current cluster

Relation to ordinary percolation

- The edge weights that are less than p form a model of ordinary p-percolation. We say that a vertex v is connected to infinity, $v \overset{p}{\longleftrightarrow} \infty$, if there exists an infinite path of edges all having weight at most p.
- If the invasion encounters a p-cluster, no edges outside it will be accepted until the p-cluster has been completely invaded. In particular, once an infinite p-cluster ($p > p_c$) is encoutered, no other edges will ever be accepted.

Relation to ordinary percolation

- The edge weights that are less than p form a model of ordinary p-percolation. We say that a vertex v is connected to infinity, $v \xleftrightarrow{p} \infty$, if there exists an infinite path of edges all having weight at most p.
- If the invasion encounters a p-cluster, no edges outside it will be accepted until the p-cluster has been completely invaded. In particular, once an infinite p-cluster ($p > p_c$) is encoutered, no other edges will ever be accepted.

Relation to ordinary percolation

Definition

Consider a graph G with origin o and uniform i.i.d. edge weights. The *percolation probability at level* p, written $\theta(p)$, is the probability that o belongs to an infinite p-cluster (a connected subgraph of edges whose weights are all $< p$). The *critical percolation threshold* p_c is

$$p_c = \inf\{p : \theta(p) > 0\}.$$

Self-organized criticality

Let ξ_i denote the weight of the i^{th} invaded edge.

Theorem

(Chayes, Chayes and Newman; Häggström, Peres and Schonmann) Suppose G is a (quasi-)transitive graph. Then with probability 1,

$$\limsup_{i\to\infty} \xi_i = p_c$$

Notice that invasion percolation finds p_c despite not having a parameter in its definition. This makes it an example of *self-organized criticality*.

Ponds and outlets

Suppose e_1 is the highest-weight edge ever invaded, with weight Q_1. We call e_1 the *first outlet*. The edges invaded up to e_1 we call the *first pond*.

Similarly, the invaded edge after the first outlet having highest weight is the second outlet, and the edges between them are called the second pond, and so on.

- $Q_n > p_c$
- $Q_{n+1} < Q_n$
- $\lim_{n\to\infty} Q_n = p_c$
- $\mathbb{P}(Q_1 < p) = \theta(p)$

Ponds and outlets

Suppose e_1 is the highest-weight edge ever invaded, with weight Q_1. We call e_1 the *first outlet*. The edges invaded up to e_1 we call the *first pond*.
Similarly, the invaded edge after the first outlet having highest weight is the second outlet, and the edges between them are called the second pond, and so on.

- $Q_n > p_c$
- $Q_{n+1} < Q_n$
- $\lim_{n\to\infty} Q_n = p_c$
- $\mathbb{P}(Q_1 < p) = \theta(p)$

Ponds and outlets

Suppose e_1 is the highest-weight edge ever invaded, with weight Q_1. We call e_1 the *first outlet*. The edges invaded up to e_1 we call the *first pond*.
Similarly, the invaded edge after the first outlet having highest weight is the second outlet, and the edges between them are called the second pond, and so on.

- $Q_n > p_c$
- $Q_{n+1} < Q_n$
- $\lim_{n\to\infty} Q_n = p_c$
- $\mathbb{P}(Q_1 < p) = \theta(p)$

Ponds and outlets

Suppose e_1 is the highest-weight edge ever invaded, with weight Q_1. We call e_1 the *first outlet*. The edges invaded up to e_1 we call the *first pond*.
Similarly, the invaded edge after the first outlet having highest weight is the second outlet, and the edges between them are called the second pond, and so on.

- $Q_n > p_c$
- $Q_{n+1} < Q_n$
- $\lim_{n\to\infty} Q_n = p_c$
- $\mathbb{P}(Q_1 < p) = \theta(p)$

Ponds and outlets

Suppose e_1 is the highest-weight edge ever invaded, with weight Q_1. We call e_1 the *first outlet*. The edges invaded up to e_1 we call the *first pond*.
Similarly, the invaded edge after the first outlet having highest weight is the second outlet, and the edges between them are called the second pond, and so on.

- $Q_n > p_c$
- $Q_{n+1} < Q_n$
- $\lim_{n\to\infty} Q_n = p_c$
- $\mathbb{P}(Q_1 < p) = \theta(p)$

Ponds and outlets

Suppose e_1 is the highest-weight edge ever invaded, with weight Q_1. We call e_1 the *first outlet*. The edges invaded up to e_1 we call the *first pond*.

Similarly, the invaded edge after the first outlet having highest weight is the second outlet, and the edges between them are called the second pond, and so on.

- $Q_n > p_c$
- $Q_{n+1} < Q_n$
- $\lim_{n\to\infty} Q_n = p_c$
- $\mathbb{P}(Q_1 < p) = \theta(p)$

Ponds and outlets

Interpret the edge weights as heights of barriers between vertices. Pour water into the origin. When enough water has accumulated, water will flow over the lowest adjacent edge into a new vertex. When more water has accumulated, water will again pour over the lowest boundary edge, etc. Run this process forever.

Ponds and outlets

Suppose an edge e, with weight p, is first examined while invading the n^{th} pond. (That is, i is the first step at which e is on the boundary of $\mathcal{C}_{i-1}$.) Then we have the following dichotomy: either

- e will be invaded as part of the n^{th} pond (if $p \leq Q_n$); or
- e will never be invaded (if $p > Q_n$)

This implies that the ponds are connected subgraphs and touch each other only at the outlets. Moreover, the outlets are pivotal in the sense that any infinite non-intersecting path in $\mathcal{C}$ starting at o must pass through every outlet. Consequently $\mathcal{C}$ is decomposed as an infinite chain of ponds, connected at the outlets.

A technical requirement

Is the maximum edge weight actually attained? In other words, are the outlets actually well defined? The answer is *Yes*, provided that

- $\limsup_{i\to\infty} \xi_i = p_c$ (for example if G is transitive); and
- $\theta(p_c) = 0$, so that $Q_n > p_c$ a.s.

For example, the one-dimensional case $G = \mathbb{Z}$ does *not* have a pond and outlet structure.

A technical requirement

Is the maximum edge weight actually attained? In other words, are the outlets actually well defined? The answer is *Yes*, provided that

- $\limsup_{i\to\infty} \xi_i = p_c$ (for example if G is transitive); and
- $\theta(p_c) = 0$, so that $Q_n > p_c$ a.s.

For example, the one-dimensional case $G = \mathbb{Z}$ does *not* have a pond and outlet structure.

A technical requirement

Is the maximum edge weight actually attained? In other words, are the outlets actually well defined? The answer is *Yes*, provided that

- $\limsup_{i\to\infty} \xi_i = p_c$ (for example if G is transitive); and
- $\theta(p_c) = 0$, so that $Q_n > p_c$ a.s.

For example, the one-dimensional case $G = \mathbb{Z}$ does *not* have a pond and outlet structure.

A technical requirement

Is the maximum edge weight actually attained? In other words, are the outlets actually well defined? The answer is *Yes*, provided that

- $\limsup_{i \to \infty} \xi_i = p_c$ (for example if G is transitive); and
- $\theta(p_c) = 0$, so that $Q_n > p_c$ a.s.

For example, the one-dimensional case $G = \mathbb{Z}$ does *not* have a pond and outlet structure.

Percolation on a regular rooted tree

Henceforth we will work exclusively on a rooted regular tree. For a rooted binary tree, we have

$$p_c = \frac{1}{2},$$

$$\theta(p) = \begin{cases} \frac{2p-1}{p^2} & \text{if } p \geq p_c \\ 0 & \text{if } p \leq p_c \end{cases}$$

Note particularly that

$$\theta(p) \approx p - p_c$$

for p slightly above p_c.
Since the graph is a tree, there is now a unique path from the root through all of the outlets, which we call the *backbone*.

Outline

Notation

Write

- L_n for the distance from the $(n-1)^{\text{st}}$ outlet to the n^{th} – i.e., the length of backbone in the n^{th} pond;
- R_n for the length of the longest upward path in the n^{th} pond;
- R'_n for the length of the longest upward path in union of the first n ponds;
- V_n for the volume (number of edges) in the n^{th} pond;
- $\hat{L}_n = \sum_{j=1}^{n} L_j$, $\hat{V}_n = \sum_{j=1}^{n} V_j$, the total backbone length and volume for the first n ponds.

Pond measurements in terms of the outlet weights

Earlier work (Angel, G., den Hollander, Slade) gave descriptions of the laws of L_n, R_n and V_n, conditional on the values of Q_n. For instance, conditional on Q_n, the L_n are conditionally independent geometric random variables with mean $\approx (Q_n - p_c)^{-1}$. In particular, their fluctuations arise from fluctuations of Q_n, together with additional randomness. The results for Q_n are extended to results for L_n, R_n and V_n using bounds on this additional randomness.

Law of large numbers

Theorem

Define the 7-tuples

$$\vec{Z}_n = (\log(Q_n - p_c)^{-1}, \log L_n, \log \hat{L}_n, \log R_n, \log R'_n, \tfrac{1}{2}\log V_n, \tfrac{1}{2}\log \hat{V}_n)$$

and

$$\vec{1} = (1,1,1,1,1,1,1)$$

Then w.p. 1,

$$\lim_{n\to\infty} \frac{1}{n} Z_n = \vec{1}$$

Central limit theorem

Theorem

Let B_t denote a standard Brownian motion. Then, as processes,

$$\left(\frac{Z_{\lfloor Nt \rfloor} - Nt \cdot \vec{1}}{\sqrt{N}} \right)_{t \geq 0} \Rightarrow (B_t \cdot \vec{1})_{t \geq 0}$$

when $N \to \infty$.

Large deviations

Theorem

$\frac{1}{n}\log(Q_n - p_c)^{-1}$ satisfies a large deviation principle with rate function

$$\phi(u) = u - \log u - 1.$$

$\frac{1}{n}\log L_n$, $\frac{1}{n}\log R_n$ and $\frac{1}{2n}\log V_n$ satisfy large deviation principles with rate function

$$\psi(u) = \begin{cases} \phi(u) & \text{if } u \geq 1/2 \\ \phi(1/2) + (1/2 - u) & \text{if } u \leq 1/2 \end{cases}$$

$\psi(u)$ arises as the solution of the variational problem

$$\psi(u) = \inf_{v > u} [\phi(v) + (v - u)]$$

Tail asymptotics

Theorem

For n fixed and $\epsilon \to 0^+$, $k \to \infty$, $\sigma =$ *the degree,*

$$\mathbb{P}\left(Q_n < p_c(1+\epsilon)\right) \sim \frac{2\sigma}{\sigma-1} \frac{\epsilon \left(\log \epsilon^{-1}\right)^{n-1}}{(n-1)!}$$

and

$$\mathbb{P}\left(L_n > k\right) \sim \mathbb{P}\left(\hat{L}_n > k\right) \sim \frac{2\sigma}{\sigma-1} \frac{(\log k)^{n-1}}{k(n-1)!}$$

$$\mathbb{P}\left(R_n > k\right) \asymp \mathbb{P}\left(R_n' > k\right) \asymp \frac{(\log k)^{n-1}}{k}$$

$$\mathbb{P}\left(V_n > k\right) \asymp \mathbb{P}\left(\hat{V}_n > k\right) \asymp \frac{(\log k)^{n-1}}{\sqrt{k}}$$

Tail asymptotics reformulated

These asymptotics can be expressed as

$$\mathbb{P}\left(Q_n < p_c(1+\epsilon)\right) \sim \frac{\left(\log \epsilon^{-1}\right)^{n-1}}{(n-1)!}\theta(p_c(1+\epsilon))$$

$$\mathbb{P}\left(L_n > k\right) \sim \mathbb{P}\left(\hat{L}_n > k\right) \sim \frac{(\log k)^{n-1}}{(n-1)!}\mathbb{P}_{p_c}(o \leftrightarrow \partial B(k))$$

$$\mathbb{P}\left(R_n > k\right) \asymp \mathbb{P}\left(R'_n > k\right) \asymp (\log k)^{n-1}\mathbb{P}_{p_c}(o \leftrightarrow \partial B(k))$$

$$\mathbb{P}\left(V_n > k\right) \asymp \mathbb{P}\left(\hat{V}_n > k\right) \asymp (\log k)^{n-1}\mathbb{P}_{p_c}(|C(o)| > k)$$

Outline

Markov structure of the outlet weights

In any graph G, conditional on the union of the first n ponds C and the n^{th} outlet weight Q_n, Q_{n+1} is chosen from the distribution function $\theta_{G \setminus C}$, conditioned to be smaller than Q_n. While this representation is complicated in general, on the tree the modified graph $G \setminus C$ is equivalent to G itself.

Hence the Markov structure of $(Q_n)_{n=1}^{\infty}$ is: choose Q_1 according to the distribution function $\theta(p)$; then choose subsequent Q_n's from the *same* distribution, conditioned to be smaller than the previous one.

Markov structure of the outlet weights

In any graph G, conditional on the union of the first n ponds C and the n^{th} outlet weight Q_n, Q_{n+1} is chosen from the distribution function $\theta_{G\setminus C}$, conditioned to be smaller than Q_n. While this representation is complicated in general, on the tree the modified graph $G\setminus C$ is equivalent to G itself.
Hence the Markov structure of $(Q_n)_{n=1}^{\infty}$ is: choose Q_1 according to the distribution function $\theta(p)$; then choose subsequent Q_n's from the *same* distribution, conditioned to be smaller than the previous one.

Reinterpreting the Markov structure

It is convenient to apply θ: then, using the smoothness of θ,

- $\theta(V_1)$ is Uniform$[0, 1]$
- conditional on $\theta(Q_n)$, $\theta(Q_{n+1})$ is Uniform$[0, \theta(Q_n)]$.

But this is equivalent to taking a *product of independent uniform random variables*:

Law of the sequence $\theta(Q_n)$

Theorem

The sequence $(\theta(Q_n))_{n=1}^{\infty}$ *has the same law as the sequence*

$$Y_n = \prod_{i=1}^{n} U_i$$

where the U_i *are independent Uniform*$[0,1]$ *random variables. Taking logarithms, the sequence* $(\log \theta(Q_n)^{-1})_{n=1}^{\infty}$ *has the same law as the sequence*

$$Z_n = \sum_{i=1}^{n} E_i$$

where the E_i *are independent Exponential(1) random variables.*

Proofs for the outlet weights

The LLN, CLT and LDP results for Q_n follow at once from the representation of $\log \theta(Q_n)^{-1}$ as a sum of independent random variables. The asymptotics for Q_n can be computed using the fact that $\log \theta(Q_n)^{-1}$ is Gamma$(n, 1)$:

$$
\begin{aligned}
\mathbb{P}(\theta(Q_n) < \epsilon) &= \mathbb{P}\left(\log\left(\theta(Q_n)^{-1}\right) > \log \epsilon^{-1}\right) \\
&= \int_{\log \epsilon^{-1}}^{\infty} \frac{x^{n-1}}{(n-1)!} e^{-x} dx \\
&= \frac{\epsilon \left(\log \epsilon^{-1}\right)^n}{(n-1)!} \int_0^{\infty} (1+u)^{n-1} e^{-u \log \epsilon^{-1}} \\
&\sim \frac{\epsilon \left(\log \epsilon^{-1}\right)^{n-1}}{(n-1)!}
\end{aligned}
$$

Conditional tail bounds

To prove the LLN and CLT results, it suffices to prove bounds of the form:

$$\mathbb{P}((Q_n - p_c)^a X_n > S) \leq O(S^{-\beta}), \qquad (S \to \infty)$$

$$\mathbb{P}((Q_n - p_c)^a X_n < s) \leq O(s^{\beta}), \qquad (s \to 0)$$

and for the tail asymptotics it suffices to prove

$$\mathbb{P}((Q_n - p_c)^a X_n > S | Q_n) \leq O(S^{-\beta}), \qquad (S \to \infty)$$

$$\mathbb{P}((Q_n - p_c)^a X_n < s_0 | Q_n) \geq p_0 \qquad (\text{for some } s_0, p_0)$$

for $a\beta > 1$.
These bounds are rather modest, and indeed the stronger bounds that follow imply the bounds above.

Conditional tail bounds

The strongest bounds are needed to prove the LDP:

$$\mathbb{P}((Q_n - p_c)^a X_n > S | Q_n) \le O(\exp(-cS^\beta)), \qquad (S \to \infty)$$

and

$$\mathbb{P}((Q_n - p_c)^a X_n < s | Q_n) \asymp s^{1/a}, \qquad (s \to 0)$$

and L_n, R_n and V_n all satisfy these bounds.

Outline

Results in 2 dimensions

In Z^2, it is known from results of van den Berg, Damron, Járai, Sapozhnikov and Vágvölgyi that

$$a^{-n} \leq Q_n - p_c \leq b^{-n},$$

$$c^n \leq R_n \leq d^n$$

as $n \to \infty$, and

$$\mathbb{P}(R_n > k) \asymp (\log k)^{n-1} \mathbb{P}_{p_c}(o \leftrightarrow \partial B(k))$$

just as on the tree. The fact that these same results hold, with the same correction factor of $(\log k)^{n-1}$, suggests that a more general phenomenon is at play.

2 dimensions compared to the tree

The analysis for the tree suggests that the outlet weight Q_n explain the common behaviour in these two very different graphs. On the tree, the representation of $\log \theta(Q_n)^{-1}$ as a Gamma$(n, 1)$ variable explained immediately both the exponential growth results and the somewhat mysterious correction factor $(\log k)^{n-1}$. While the exact representation for Q_n cannot be expected to hold in other graphs, it seems a reasonable heuristic even for more complicated graphs.

Comparison to percolation with defects

Since invasion percolation mostly accepts sub-critical edges, occasionally interspersed with super-critical edges, it is tempting to compare it to *percolation with defects*. For an ordinary percolation model, say that $o \leftrightarrow_n S$ if o is connected to S by a path having at most n closed edges. In 2 dimensions, for each n,

$$\mathbb{P}(\text{the } n^{\text{th}} \text{ pond intersects } \partial B(0,k)) \asymp \mathbb{P}_{p_c}(o \leftrightarrow_{n-1} \partial B(0,k))$$

By contrast, on the tree

$$\mathbb{P}(o \leftrightarrow_n \partial B(0,k)) \asymp k^{-2^{-n}}$$

Thank you.