

Zero forcing with random sets

Bryan Curtis^{a,1}, Luyining Gan^{b,*}, Jamie Haddock^{c,2}, Rachel Lawrence^{d,3},
Sam Spiro^{e,4}

^a Department of Mathematics, Iowa State University, United States of America

^b School of Science, Beijing University of Posts and Telecommunications, China

^c Department of Mathematics, Harvey Mudd College, United States of America

^d Department of Electrical Engineering and Computer Science, University of California, Berkeley, United States of America

^e Department of Mathematics, University of California, San Diego, United States of America

ARTICLE INFO

Article history:

Received 5 August 2023

Received in revised form 16 February 2024

Accepted 17 February 2024

Available online xxxx

Keywords:

Random set zero forcing

Zero forcing polynomial

Random graphs

Graph threshold

ABSTRACT

Given a graph G and a real number $0 \leq p \leq 1$, we define the random set $B_p(G) \subseteq V(G)$ by including each vertex independently and with probability p . We investigate the probability that the random set $B_p(G)$ is a zero forcing set of G . In particular, we prove that for large n , this probability for trees is upper bounded by the corresponding probability for a path graph. Given a minimum degree condition, we also prove a conjecture of Boyer et al. regarding the number of zero forcing sets of a given size that a graph can have.

© 2024 Elsevier B.V. All rights reserved.

1. Introduction

Let G be a graph with vertex set $V(G)$, with each vertex initially colored either blue or white. If u is a blue vertex of G and the neighborhood $N_G(u)$ of u contains exactly one white vertex v , then we may change the color of v to blue. This iterated procedure for coloring a graph is called “zero forcing”. A *zero forcing set* B is a subset of vertices of G such that, if G initially has all of the vertices of B colored blue, then the zero forcing process may eventually color all of $V(G)$ blue. We let $ZFS(G)$ denote the set of all zero forcing sets of G . The *zero forcing number* $Z(G)$ is the minimum cardinality of a zero forcing set in G ; that is, $Z(G) = \min_{B \in ZFS(G)} |B|$. The zero forcing process was first introduced by Burgarth and Giovannetti [5], and the zero forcing number was introduced by an AIM research group [14] as a bound for the maximum nullity of a graph G . In addition to its independent mathematical interest, applications of zero forcing have included models of rumor spreading [11] and power grid domination [22].

In this paper we consider a randomized version of the zero forcing process. There are seemingly two natural ways to define such a random process: one can either use a deterministic set of blue vertices B together with forces that occur randomly, or one can use a random set of vertices B together with deterministic forces. The process known as *probabilistic*

* Corresponding author.

E-mail addresses: bcurtis1@iastate.edu (B. Curtis), elainegan@bupt.edu.cn (L. Gan), jhaddock@g.hmc.edu (J. Haddock), rlaw@berkeley.edu (R. Lawrence), sspiro@ucsd.edu (S. Spiro).

¹ This work is partially supported by NSF grant 1839918.

² This work is supported by NSF DMS award #2211318.

³ This work is supported by NSF grant CCF-2231095 and by NSF grant CCF-1815328.

⁴ This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1650112.

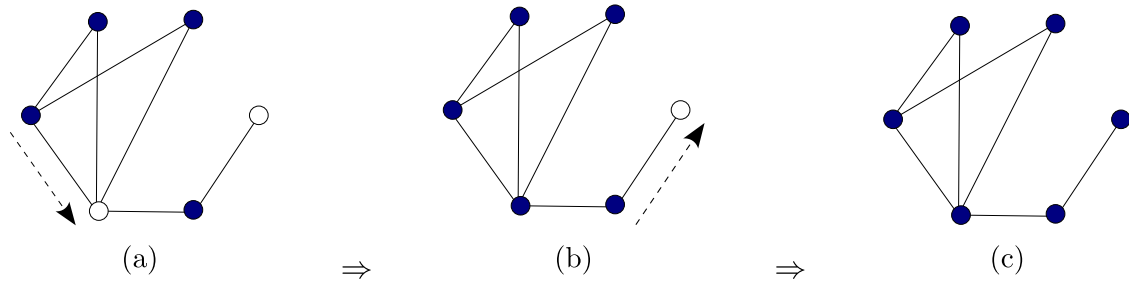


Fig. 1. An example of zero forcing on a graph. An edge connecting the blue vertex that forces the white vertex in the next frame is indicated with arrows in panel (1a) and (1b). Note that three blue vertices could force in panel (1a). (For interpretation of the color(s) in the figure(s), the reader is referred to the web version of this article.)

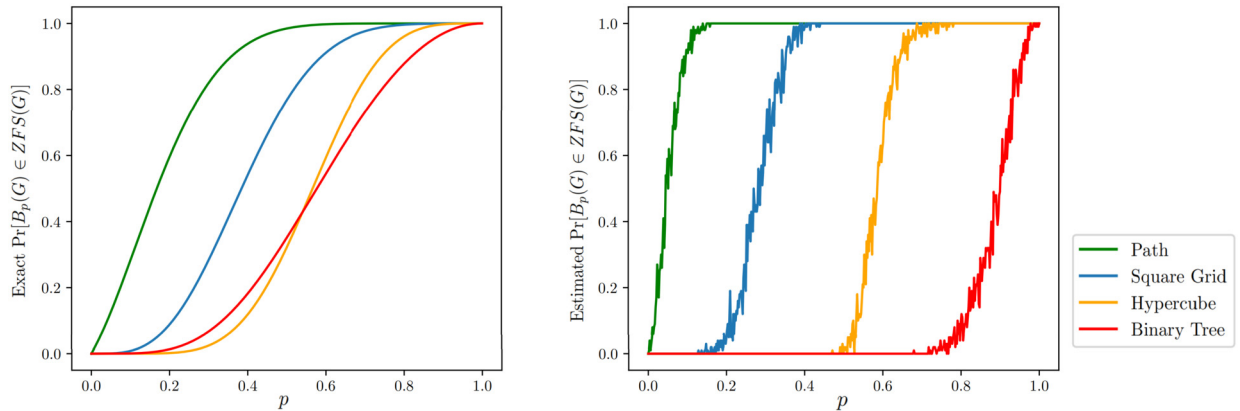


Fig. 2. Exact brute-force computation (left) and Monte Carlo estimates (right) of $\Pr[B_p(G) \in \text{ZFS}(G)]$ for the path, square grid, hypercube, and left-complete binary tree graphs on 16 vertices (left) and 256 vertices (right).

zero forcing, which was introduced by Kang and Yi [17], is of the former type and is by now well studied, see for example [10,13,19,11,16,7,3]. In this paper we introduce a process of the latter type which we call *random set zero forcing*. This version of zero forcing can be interpreted as investigating the distribution of zero forcing sets among all subsets of the vertices in a graph. For example, in the application of power grid domination, random set zero forcing analyzes the *random initialization* scenario: if phase measurement units (PMUs) are placed on n randomly chosen nodes of an electric power network, determine the probability that the entire graph is monitored.

1.1. Main results

Given a graph G and real number $0 \leq p \leq 1$, we define the random set $B_p(G) \subseteq V(G)$ by including each vertex of G independently and with probability p . For example, $B_1(G) = V(G)$, $B_0(G) = \emptyset$, and $B_{1/2}(G)$ is equally likely to be any subset of $V(G)$.

The central question we wish to ask is: Given G and p , what is (approximately) the probability that $B_p(G)$ is a zero forcing set of G ? For example, one general bound we can prove is the following.

Proposition 1.1. *Let G be an n -vertex graph with minimum degree at least $\delta \geq 1$. For all p , we have*

$$\Pr[B_p(G) \in \text{ZFS}(G)] \leq \delta np^\delta.$$

In fact, we prove a slightly stronger version of this theorem that holds for graphs with “few” vertices of degree less than δ ; see Proposition 2.3. Proposition 1.1 can be viewed as a probabilistic analog of the basic fact that $Z(G) \geq \delta$ if G is a graph with minimum degree δ .

For many graphs G , it will happen that there exists a p such that $B_{p'}(G)$ is very unlikely to be a zero forcing set if p' is much smaller than p , and that $B_{p'}(G)$ is very likely to be zero forcing if p' is much larger than p , see for example Fig. 2. This line of inquiry is motivated by the study of thresholds in random graphs, which is one of the fundamental topics in probabilistic combinatorics (see, for example, [12]). In fact, observing that whether an initial vertex set B is zero forcing is a nontrivial monotone property on subsets of V , it follows from the foundational work of [8] that a threshold function for

the zero forcing property in any given graph exists. With this in mind, we define the *threshold probability* $p(G)$ to be the unique p such that $\Pr[B_p(G) \in \text{ZFS}(G)] = \frac{1}{2}$, and it is not difficult to see that $p(G)$ is well defined.

Many other results from classical zero forcing also have probabilistic analogs for random set zero forcing. For example, it is straightforward to show that if G is an n -vertex graph, then $Z(G) \leq Z(\overline{K_n}) = n$ with equality if and only if $G = \overline{K_n}$. In the random setting, it is also easy to show the analogous result that for all n -vertex graphs G and $0 \leq p \leq 1$, we have

$$\Pr[B_p(G) \in \text{ZFS}(G)] \geq \Pr[B_p(\overline{K_n}) \in \text{ZFS}(\overline{K_n})] = p^n,$$

with equality holding if and only if either $p \in \{0, 1\}$ or $G = \overline{K_n}$. In fact, we can use Observation 1.2 stated below to give the exact result $p(\overline{K_n}) = 2^{-1/n}$. Moreover, Observation 1.2 allows us to reduce our focus to connected graphs.

Observation 1.2. Let G be the disjoint union of the graphs G_1 and G_2 . Then

$$\Pr[B_p(G) \in \text{ZFS}(G)] = \Pr[B_p(G_1) \in \text{ZFS}(G_1)] \cdot \Pr[B_p(G_2) \in \text{ZFS}(G_2)].$$

Let G be a graph on $n \geq 2$ vertices with no isolated vertices. It is well known that every subset of $V(G)$ of size $n - 1$ is a zero forcing set of G , and that $Z(G) = n - 1$ if and only if $G = K_n$, the complete graph on n vertices (see, for example, [15]). With these observations it is not difficult to prove the following proposition (see Appendix A).

Proposition 1.3. If G is a graph on n vertices with no isolated vertices, then $p(G) \leq p(K_n)$. Moreover, $p(K_n) = 1 - \Theta(n^{-1})$.

While it is straightforward to determine the graphs with the largest threshold probabilities, the analogous problem for smallest thresholds appears much harder. Intuitively, the path graph P_n is a natural candidate for the minimizer, since it is known that P_n is the unique n -vertex graph with zero forcing number 1. A proof of the following basic result can be found in Appendix A.

Proposition 1.4. The threshold probability of the path on n vertices satisfies

$$p(P_n) = \Theta(n^{-1/2}).$$

In the classical setting, it is well known that among n -vertex graphs, the path P_n is the unique graph with the smallest zero forcing number. We conjecture that an analog of this result holds in the random setting.

Conjecture 1.5. If G is an n -vertex graph and $0 \leq p \leq 1$, then

$$\Pr[B_p(G) \in \text{ZFS}(G)] \leq \Pr[B_p(P_n) \in \text{ZFS}(P_n)],$$

with equality holding if and only if either $p \in \{0, 1\}$ or $G = P_n$.

While we do not prove this conjecture in full, we provide some partial results; in particular, we prove the conjecture when restricted to trees and with n sufficiently large.

Theorem 1.6. If T is an n -vertex tree with n sufficiently large, then for all $0 \leq p \leq 1$,

$$\Pr[B_p(T) \in \text{ZFS}(T)] \leq \Pr[B_p(P_n) \in \text{ZFS}(P_n)],$$

with equality holding if and only if either $p \in \{0, 1\}$ or $T = P_n$.

We note that if Conjecture 1.5 were true, then in particular we would have $p(G) > p(P_n)$ for all n -vertex graphs $G \neq P_n$; we subsequently prove that this is true up to a constant factor. Here and throughout the paper we use standard asymptotic notation, which is recalled in Subsection 1.2.

Theorem 1.7. If G is an n -vertex graph, then

$$p(G) = \Omega(p(P_n)) = \Omega(n^{-1/2}).$$

In essence this result says that a random set of significantly less than $n^{1/2}$ vertices of any n -vertex graph G is very unlikely to be a zero forcing set for sufficiently large n .

Conjecture 1.5 can be viewed as a weakened version of a conjecture involving the number of zero forcing sets of a given size. To this end, we observe that if G is an n vertex graph and $z(G; k)$ is the number of zero forcing sets of G of size k , then

$$\Pr[B_p(G) \in \text{ZFS}(G)] = \sum_{k=1}^n z(G; k) p^k (1-p)^{n-k}. \quad (1)$$

The notation $z(G; k)$ follows that of Boyer et al. in [2] who introduced the study of zero forcing polynomials and found many explicit formulas for $z(G; k)$, including:

$$z(P_n; k) = \binom{n}{k} - \binom{n-k-1}{k}. \quad (2)$$

Observe that, by (1), Conjecture 1.5 is a weakened version of the following conjecture.

Conjecture 1.8 ([2]). *If G is an n -vertex graph, then for all k ,*

$$z(G; k) \leq z(P_n; k) = \binom{n}{k} - \binom{n-k-1}{k}.$$

It was shown in [2] that Conjecture 1.8 holds whenever G contains a Hamiltonian path, but other than this very little is known. By extending our proof of Proposition 1.1, we prove Conjecture 1.8 whenever k is sufficiently small, as a function of the minimum degree of G .

Proposition 1.9. *If G is an n -vertex graph with minimum degree $\delta \geq 3$, then for all $k \leq (2\delta)^{-1/\delta} n^{1-1/\delta}$ we have $z(G; k) \leq z(P_n; k)$.*

We additionally show that this implies Conjecture 1.8 whenever G has sufficiently large minimum degree.

Corollary 1.10. *If G is an n -vertex graph with minimum degree $\delta \geq \log_2(n) + 2 \log_2 \log_2(n)$, then $z(G; k) \leq z(P_n; k)$ for all k .*

1.2. Organization and notation

This paper is organized as follows. In Section 2, we provide a general bound on the probability that $B_p(G)$ is zero forcing given the minimum degree. In Section 3, we prove that the threshold probability for an n -vertex graph G is $\Omega(n^{-1/2})$. In Section 4, we prove that among trees on sufficiently many vertices, paths have the largest probability of $B_p(G)$ being a zero forcing set. We conclude with some remarks and open questions in Section 5.

Throughout we use standard asymptotic notation. Let $f(n)$ and $g(n)$ be functions from the non-negative integers to the reals. We write $f(n) = o(g(n))$ if $\lim_{n \rightarrow \infty} f(n)/g(n) = 0$, and $f(n) = O(g(n))$ if there exists a $C > 0$ such that $f(n) \leq Cg(n)$ for all sufficiently large n . We write $f(n) = \Omega(g(n))$ if $g(n) = O(f(n))$, and $f(n) = \Theta(g(n))$ if both $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$. We write, for example, $\Theta_k(g(n))$ if the implicit constants depend on k .

2. Bounds using degrees

In this warmup section, we give bounds on the probability that $B_p(G)$ is a zero forcing set in terms of the degree sequence of G . Our most general bound of this form is the following, where here and throughout $d(v)$ denotes the degree of v in the graph G .

Lemma 2.1. *Let G be an n -vertex graph with at least one edge and $p \in [0, 1]$. Then*

$$\Pr[B_p(G) \in \text{ZFS}(G)] \leq \sum_{v \in V(G)} d(v) p^{d(v)}.$$

Proof. Let A be the event that $B_p(G) = V(G)$. For any $v \in V(G)$, let F_v be the event that v and exactly $d(v) - 1$ of its neighbors are in $B_p(G)$. We claim that for $B_p(G)$ to be a zero forcing set, either A or F_v for some v must occur. Indeed, if $B_p(G) \neq V(G)$ and $B_p(G)$ is a zero forcing set, then there must be some blue vertex v in $B_p(G)$ that forces a white vertex to be blue. In particular, if v is the first vertex which performs such a force, then it and exactly $d(v) - 1$ of its neighbors must be in $B_p(G)$. This proves our claim. Thus by the union bound we have

$$\Pr[B_p(G) \in \text{ZFS}(G)] \leq \Pr[A \cup \bigcup_{v \in V(G)} F_v] \leq \Pr[A] + \sum_{v \in V(G)} \Pr[F_v].$$

By definition, the event F_v occurring means v is included in $B_p(G)$ and exactly $d(v) - 1$ of its $d(v)$ neighbors are included in $B_p(G)$. As each vertex is included in $B_p(G)$ independently and with probability p , we have

$$\Pr[F_v] = p \cdot \binom{d(v)}{d(v)-1} p^{d(v)-1} (1-p) = d(v) p^{d(v)} (1-p).$$

Plugging this into the bound above and using $\Pr[A] = p^n$ gives

$$\Pr[B_p(G) \in \text{ZFS}(G)] \leq p^n + \sum_{v \in V(G)} d(v)p^{d(v)}(1-p). \quad (3)$$

By assumption, G contains a vertex u with $d(u) \geq 1$. For this vertex we have

$$d(u)p^{d(u)}(1-p) = d(u)p^{d(u)} - d(u)p^{d(u)+1} \leq d(u)p^{d(u)} - p^n,$$

where this last step used $d(u) \geq 1$ and $d(u) + 1 \leq n$ (which always holds for n -vertex graphs). Plugging this bound into (3), and using the bound $d(v)p^{d(v)}(1-p) \leq d(v)p^{d(v)}$ for every other term of the sum gives the desired result. $\square$

We also make use of the following, which can be proven using calculus.

Observation 2.2. If d is a positive integer and $p \leq e^{-1/d}$, then

$$\max_{x \geq d} xp^x = dp^d.$$

This result quickly gives Proposition 1.1, which we restate below.

Proposition 1.1. Let G be an n -vertex graph with minimum degree at least $\delta \geq 1$. For all p , we have

$$\Pr[B_p(G) \in \text{ZFS}(G)] \leq \delta np^\delta.$$

Proof. When $G = K_2$ the theorem is equivalent to $2p(1-p) + p^2 \leq 2p$, i.e. that $-p^2 \leq 0$, so the result holds. From now on we assume G has at least 3 vertices. For all p and $n \geq 3$, we have

$$\Pr[B_p(G) \in \text{ZFS}(G)] \leq e^{-1} \delta n$$

since $e^{-1} \delta n \geq 1$. This implies the result when $p \geq e^{-1/\delta}$.

Observation 2.2 together with Lemma 2.1 and the fact that $d(v) \geq \delta$ for all v gives

$$\Pr[B_p(G) \in \text{ZFS}(G)] \leq \sum_{v \in V(G)} d(v)p^{d(v)} \leq \delta np^\delta. \quad \square$$

We next prove a slightly stronger version of this result which holds for graphs with “few” vertices of degree less than a given degree d .

Proposition 2.3. Let G be an n -vertex graph without isolated vertices. Suppose that there exist integers $1 \leq d \leq n$ and $N \geq 0$ such that G contains at most N^k vertices of degree k for all $1 \leq k < d$. Then for all $p \leq e^{-1/d}$, we have

$$\Pr[B_p(G) \in \text{ZFS}(G)] \leq 4pN + dnp^d.$$

Proof. The result is trivial if $pN > \frac{1}{2}$, so we can assume $pN \leq \frac{1}{2}$. Using Observation 2.2 together with Lemma 2.1 and the assumptions on G , we find

$$\begin{aligned} \Pr[B_p(G) \in \text{ZFS}(G)] &\leq \sum_{v \in V(G)} d(v)p^{d(v)} \leq \sum_{\substack{v \in V(G) \\ d(v) < d}} d(v)p^{d(v)} + \sum_{\substack{v \in V(G) \\ d(v) \geq d}} dp^d \\ &\leq \sum_{k=1}^{d-1} k(pN)^k + dnp^d \leq \sum_{k=1}^{\infty} k(pN)^k + dnp^d. \end{aligned}$$

Note that in general we have $\sum_{k=1}^{\infty} kc^k = \frac{c}{(1-c)^2}$ provided $|c| < 1$. Applying this with $c = pN \leq \frac{1}{2}$ gives the desired result. $\square$

Recall that $z(G; k)$ is the number of zero forcing sets of G of size k . We now prove analogs of these results for $z(G; k)$.

Lemma 2.4. Let G be an n -vertex graph with at least one edge. Then for all non-negative integers k ,

$$z(G; k) \leq \sum_{v \in V(G)} d(v) \binom{n-d(v)}{k-d(v)}.$$

Proof. The result is trivial if $k = n$. For $k < n$, every zero forcing set S must contain some vertex v of positive degree and exactly $d(v) - 1$ of its neighbors in order to have a vertex force. Thus every zero forcing set of size k can be constructed by first including a vertex v , then including exactly $d(v) - 1$ of its neighbors, then arbitrarily including $k - d(v)$ additional vertices. In total the number of ways to construct such a set is

$$\sum_{v \in V(G)} \binom{d(v)}{d(v)-1} \binom{n-d(v)}{k-d(v)} = \sum_{v \in V(G)} d(v) \binom{n-d(v)}{k-d(v)},$$

so G has at most this many zero forcing sets of size k . $\square$

We next need the following lower bound on $z(P_n; k)$.

Lemma 2.5. *For all non-negative integers k we have*

$$z(P_n; k) \geq \frac{k^2}{n+k^2} \binom{n}{k}.$$

Proof. Recall from (2) that $z(P_n; k) = \binom{n}{k} - \binom{n-k-1}{k}$ for all k . Observe that

$$\begin{aligned} \frac{\binom{n-k-1}{k}}{\binom{n}{k}} &= \frac{(n-k-1)(n-k-2) \cdots (n-2k)}{n(n-1) \cdots (n-k+1)} = \prod_{i=0}^{k-1} \left(1 - \frac{k+1}{n-i}\right) \\ &\leq \left(1 - \frac{k}{n}\right)^k \leq \frac{1}{1+k^2/n} = \frac{n}{n+k^2}, \end{aligned}$$

where the last inequality follows from the Bernoulli inequality; see e.g., [18, Equation (r'_5)]. This implies

$$z(P_n; k) = \binom{n}{k} - \binom{n-k-1}{k} \geq \left(1 - \frac{n}{n+k^2}\right) \binom{n}{k} = \frac{k^2}{n+k^2} \binom{n}{k}. \quad \square$$

We now prove Proposition 1.9, which we restate below.

Proposition 1.9. *Let G be an n -vertex graph with minimum degree $\delta \geq 3$ and $k \leq (2\delta)^{-1/\delta} n^{1-1/\delta}$. Then $z(G; k) \leq z(P_n; k)$.*

Proof. By (2), for $k \geq n/2$ we have $z(P_n; k) = \binom{n}{k}$. Thus we may assume throughout that $k \leq n/2$.

Observe that for all t ,

$$\binom{n-t}{k-t} / \binom{n}{k} = \frac{k(k-1) \cdots (k-t+1)}{n(n-1) \cdots (n-t+1)} \leq (k/n)^t,$$

with this last step using the fact that $(k-i)/(n-i) \leq k/n$ for $i \geq 1$ if and only if $k \leq n$. Using this and Lemma 2.4 gives

$$z(G; k) \leq \sum_v d(v) (k/n)^{d(v)} \binom{n}{k}.$$

Because $\delta \geq 3$, we have $k \leq e^{-1/\delta} n$, so by Observation 2.2 we have

$$z(G; k) \leq \delta n (k/n)^\delta \binom{n}{k}. \quad (4)$$

First consider the case $k \leq \sqrt{n}$. By (4) and Lemma 2.5, to prove $z(G; k) \leq z(P_n; k)$, it suffices to have $\delta n (k/n)^\delta \leq k^2/2n$, or equivalently $n/k \geq (2\delta)^{1/(\delta-2)}$. Since $k \leq \sqrt{n}$, it suffices to prove $n \geq (2\delta)^{2/(\delta-2)}$, and this is true for $3 \leq \delta \leq n$ and $n \geq 5$. Thus we may assume $k \geq \sqrt{n}$. In this case (4) and Lemma 2.5 imply $z(G; k) \leq z(P_n; k)$ provided

$$\delta n (k/n)^\delta \leq \frac{1}{2},$$

and this holds precisely when $k \leq (2\delta)^{-1/\delta} n^{1-1/\delta}$. $\square$

With Proposition 1.9 we can prove Corollary 1.10, which we restate below.

Corollary 1.10. *Let G be an n -vertex graph with minimum degree $\delta \geq \log_2(n) + 2 \log_2 \log_2(n)$. Then $z(G; k) \leq z(P_n; k)$ for all k .*

Proof. The result trivially holds for $k \geq n/2$, so it suffices to prove the result for $k \leq n/2$. By Proposition 1.1, it suffices to show

$$n/2 \leq (2\delta)^{-1/\delta} n^{1-1/\delta},$$

or equivalently $n \leq 2^\delta (2\delta)^{-1}$. And indeed, for $n \geq 9$ the minimum degree condition implies

$$2^\delta (2\delta)^{-1} \geq n \frac{(\log_2(n))^2}{2(\log_2(n) + 2 \log_2 \log_2(n))} \geq n.$$

For $n \leq 8$ one can check that $\lceil \log_2(n) + 2 \log_2 \log_2(n) \rceil \geq n - 1$, so our hypothesis on δ implies G is complete and the result is immediate. In either case we conclude the result. $\square$

3. Bounds on threshold probabilities

In this section we prove that for any n -vertex graph G , the threshold probability $p(G)$ is asymptotically at least that of P_n , i.e. $p(G) = \Omega(n^{-1/2})$. At a high level, our proof revolves around finding a graph $\tilde{G}$ which has minimum degree 2 and $p(\tilde{G}) \approx p(G)$. Because $\tilde{G}$ has minimum degree 2, Proposition 1.1 implies $p(G) \approx p(\tilde{G}) = \Omega(n^{-1/2})$. We begin with a preliminary result regarding graphs containing pendant paths.

We say that a path $v_1 \cdots v_k$ in a graph G is a *pendant path*⁵ provided $k \geq 2$, $d_G(v_1) = 1$, $d_G(v_i) = 2$ for $1 < i < k$, and $d_G(v_k) > 2$. We refer to the vertex v_1 , the vertex of degree one, as the *pendant vertex*, and to v_k , the vertex of degree at least 3, as the *anchor vertex*. Observe that the only tree that does not contain a pendant path is the path graph.

Lemma 3.1. *Let G be an n -vertex graph. If there exists a vertex $w \in V(G)$ that is the anchor of two distinct pendant paths in G , then $p(G) = \Omega(n^{-1/2})$.*

Proof. Let $w \in V(G)$ and assume that w is the anchor of two distinct pendant paths in G , i.e., there exist distinct pendant paths $u_1 \cdots u_k w$ and $u_{k+1} \cdots u_\ell w$ in G . Let

$$I = \{j \in \mathbb{Z} : 1 < j < k \text{ or } k+1 < j < \ell\}$$

and for each $i \in I$, let A_i be the event that $u_i, u_{i+1} \in B_p(G)$. Let A' be the event that $B_p(G) \cap \{u_1, u_k, u_{k+1}, u_\ell\} \neq \emptyset$. Observe that if $B_p(G) \in \text{ZFS}(G)$, then A' or some A_i event occurs. Thus,

$$\Pr[B_p(G) \in \text{ZFS}(G)] \leq \Pr\left[\bigcup_{i \in I} A_i \cup A'\right] \leq \Pr[A'] + \sum_{i \in I} \Pr[A_i] \leq 4p + np^2.$$

Thus to have $\Pr[B_p(G) \in \text{ZFS}(G)] = \frac{1}{2}$, we must have $np^2 + 4p \geq \frac{1}{2}$, which implies $p = \Omega(n^{-1/2})$. $\square$

With this lemma, we see that when proving Theorem 1.7 we may assume each vertex is the anchor of at most one pendant path. The next lemma allows us to assume that none of these paths are too long (unless G consists of a single path). In order to prove the next lemma, we recall various definitions and notation related to forcing chains which can be found, for example, in [4].

Let G be a graph and $B \subseteq V(G)$. Using B as the initial set of blue vertices, apply the color change rule and record the forces. If a vertex v forces u we write $v \rightarrow u$. The *chronological list of forces* $\mathcal{F}$ is the ordered list of forces, written in the order they were performed, that produces the final coloring generated by B in G . We shall sometimes use $\mathcal{F}$ to denote the *set of forces* that produces the final coloring generated by B in G . A *forcing chain* of $\mathcal{F}$ is a sequence of vertices $(v_1, \dots, v_k)$ such that $v_i \rightarrow v_{i+1}$ for $1 \leq i \leq k-1$. A *maximal forcing chain* of $\mathcal{F}$ is a forcing chain that is not a proper subsequence of another forcing chain of $\mathcal{F}$. The *reversal* of B for $\mathcal{F}$ is the set of all vertices that do not perform a force, i.e., the set of all vertices that are the last element in a maximal forcing chain of $\mathcal{F}$.

Lemma 3.2. *Let G be an n -vertex graph and let $\{v_1, \dots, v_M\}$ denote a set of vertices of degree 1 in G . Let $\tilde{G}$ be the graph obtained from G by adding a clique on $\{v_1, \dots, v_M\}$. Then*

$$\Pr[B_p(G) \in \text{ZFS}(G)] \leq \Pr[B_p(\tilde{G}) \in \text{ZFS}(\tilde{G})] + pM.$$

Proof. We begin by showing that if $B \in \text{ZFS}(G)$ and $B \notin \text{ZFS}(\tilde{G})$, then $v_i \in B$ for some $i = 1, \dots, M$. Let $B \subseteq V(G)$ and suppose that $v_i \notin B$ for all i . Assume that $B \in \text{ZFS}(G)$ and let $\mathcal{F}$ be the set of forces for B in G . Since each v_i is a pendant vertex in G and each $v_i \notin B$, the set $\{v_1, \dots, v_M\}$ is contained in the reversal of B for $\mathcal{F}$. This, and the fact that the

⁵ Most authors do not impose any conditions on the degree of v_k in the definition of a pendant path, but this formulation will be more useful to us.

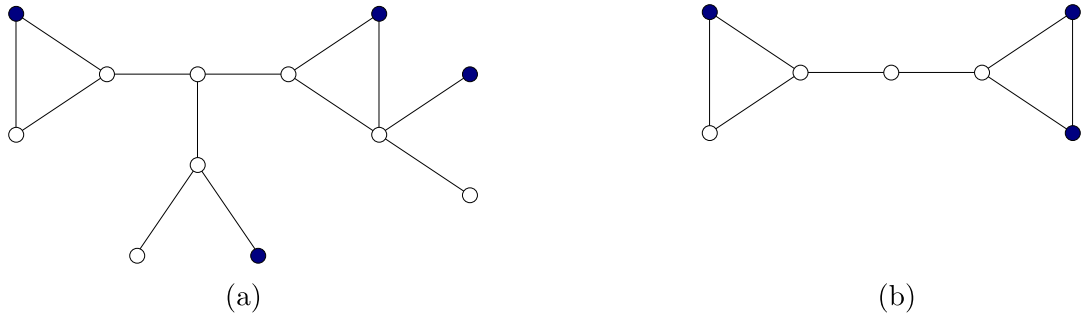


Fig. 3. (a) Graph G with zero forcing set B colored blue. (b) Graph $C_2(G)$ with zero forcing set $C_2(B)$ colored blue.

neighborhood of each vertex in $V(G) \setminus \{v_1, \dots, v_M\}$ is unchanged by adding a clique to $\{v_1, \dots, v_M\}$, implies $\mathcal{F}$ is a set of forces for B in $\tilde{G}$. Thus, $B \in \text{ZFS}(\tilde{G})$ and hence we have shown that if $B \in \text{ZFS}(G)$ and $B \notin \text{ZFS}(\tilde{G})$, then $v_i \in B$ for some $i = 1, \dots, M$.

Let A_i be the event that $v_i \in B_p(G)$. By the preceding argument and the union bound,

$$\Pr[B_p(G) \in \text{ZFS}(G) \wedge B_p(G) \notin \text{ZFS}(\tilde{G})] \leq \Pr\left[\bigcup_{i=1}^M A_i\right] \leq pM.$$

We also have

$$\begin{aligned} & \Pr[B_p(G) \in \text{ZFS}(G) \wedge B_p(G) \notin \text{ZFS}(\tilde{G})] \\ &= \Pr[B_p(G) \in \text{ZFS}(G)] - \Pr[B_p(G) \in \text{ZFS}(G) \wedge B_p(G) \in \text{ZFS}(\tilde{G})] \\ &\geq \Pr[B_p(G) \in \text{ZFS}(G)] - \Pr[B_p(G) \in \text{ZFS}(\tilde{G})]. \end{aligned}$$

Combining both inequalities gives the result. $\square$

The 2-core of a graph G , denoted $C_2(G)$, is obtained from G by repeatedly removing all isolated vertices and all vertices of degree 1 from G until no further removals are possible. See [6] for basic facts about 2-cores. We say that T is a *pendant tree* of a graph G if T is a maximal induced subgraph of G such that T is a tree, and if there exists a unique vertex $w \in V(T)$ contained in $C_2(G)$. The vertex w is called the *anchor vertex* of T . It is known that a vertex v is in $C_2(G)$ if and only if v is contained in a cycle or a path between cycles. Thus, the 2-core of G can be obtained by removing all non-anchor vertices of each pendant tree and all components of G that are trees.

Let G be a graph and $B \subseteq V(G)$. We define $C_2(B, G)$ to be the set of vertices that are either contained in $B \cap C_2(G)$ or are anchor vertices of a pendant tree T such that $B \cap V(T)$ is a zero forcing set of T . When G is clear from context we simply write $C_2(B)$. These definitions are illustrated in Fig. 3. The motivation for these definitions is found in Lemma 3.4.

Before proving our next lemma, we note the following observation about zero forcing on graphs with a cut vertex. Observation 3.3 follows from some of the concepts introduced in [21]. We write $G[S]$ to denote the induced subgraph of the graph G on $S \subseteq V(G)$.

Observation 3.3. Let G be a graph with cut vertex w , and let $W_1 \cup W_2 \cup \{w\}$ be a partition of $V(G)$ such that W_1 and W_2 are the disjoint union of connected components of $G - w$. Let $G_i = G[V(W_i) \cup \{w\}]$ for $i = 1, 2$. Then $B \cap V(G_i) \in \text{ZFS}(G_i)$ for some index i , and $B \cap V(G_i) \cup \{w\}$ is a zero forcing set of G_i for each $i = 1, 2$.

Lemma 3.4. Let G be a graph and $B \subseteq V(G)$. If B is a zero forcing set of G , then $C_2(B)$ is a zero forcing set of $C_2(G)$.

Proof. Assume for contradiction that there exists a pair (B, G) such that B is a zero forcing set of G and $C_2(B)$ is not a zero forcing set of $C_2(G)$. Moreover, choose a minimal counterexample (B, G) such that G has as few vertices as possible.

We begin with a few observations to simplify the proof. The 2-core of G is the disjoint union of the 2-cores of each connected component of G . If $C_2(G)$ is the null graph, then it is vacuously true that $C_2(B)$ is a zero forcing set of $C_2(G)$. If $C_2(G) = G$, then $C_2(B) = B$ and hence $C_2(B)$ is a zero forcing set of $C_2(G)$. We may therefore assume that G is connected and that G contains a pendant tree.

Let T be a pendant tree of G with anchor vertex w , and let G_T be the induced subgraph of G on $(V(G) \setminus V(T)) \cup \{w\}$. Let

$$B_T = \begin{cases} B \cap V(G_T) & \text{if } B \cap V(T) \notin \text{ZFS}(T) \\ (B \cap V(G_T)) \cup \{w\} & \text{if } B \cap V(T) \in \text{ZFS}(T). \end{cases}$$

Since w is a cut vertex, Observation 3.3 implies that B_T is a zero forcing set of G_T . By our assumption of (B, G) being a vertex minimal counterexample, we have that $C_2(B_T, G_T)$ is a zero forcing set of $C_2(G_T)$ since G_T has strictly fewer vertices than G . It is not difficult to check that $C_2(G_T) = C_2(G)$ and $C_2(B_T, G_T) = C_2(B, G)$, so $C_2(B, G)$ is a zero forcing set of $C_2(G)$. This gives a contradiction, proving the result. $\square$

We can now prove the main result of this section, which we restate below.

Theorem 1.7. *If G is an n -vertex graph, then*

$$p(G) = \Omega(p(P_n)) = \Omega(n^{-1/2}).$$

Proof. Observe that if H is a connected component of G , then $p(G) \geq p(H)$. Also, by Lemma 3.1, if G is a tree that is not a path, or if G has a pendant tree that is not a pendant path, then $p(G) = \Omega(n^{-1/2})$. We may therefore assume that G is connected, G contains a cycle, and every pendant tree of G is a pendant path. Note that the definition of pendant trees implies every vertex v is the anchor of at most one pendant path in G , and that every anchor vertex is contained in $C_2(G)$. Let $p = cn^{-1/2}$, where $c < 1$ is a positive constant which we specify later.

Let $\{P_1, \dots, P_M\}$ be the set of all pendant paths in G on at least $100n^{1/2} + 1$ vertices, and let $v_i \in P_i$ denote the vertex of P_i of degree 1. Let $\tilde{G}$ be the graph obtained from G by adding a clique on $\{v_1, \dots, v_M\}$. Observe that $M(100n^{1/2} + 1) \leq n$, and hence $M < n^{1/2}/100$. Thus by Lemma 3.2,

$$\Pr[B_p(G) \in \text{ZFS}(G)] < \Pr[B_p(\tilde{G}) \in \text{ZFS}(\tilde{G})] + .01.$$

Observe that $C_2(\tilde{G})$ is nonempty since G contains a cycle, and by Lemma 3.4,

$$\Pr[B_p(\tilde{G}) \in \text{ZFS}(\tilde{G})] \leq \Pr[C_2(B_p(\tilde{G})) \in \text{ZFS}(C_2(\tilde{G}))].$$

Thus it suffices to prove $\Pr[C_2(B_p(\tilde{G})) \in \text{ZFS}(C_2(\tilde{G}))] < 0.49$ for n sufficiently large.

For $v \in V(C_2(\tilde{G}))$, let A_v denote the event that $v \in C_2(B_p(\tilde{G}))$. We claim that

$$\Pr[A_v] \leq 2p + (100n^{1/2})p^2 = (2c + 100c^2)n^{-1/2} := q. \quad (5)$$

Indeed, $\Pr[A_v] = p$ if v is not the anchor of some pendant path. If v is the anchor of the pendant path P , then for A_v to occur, either an endpoint of P or two consecutive vertices of P must be in $B_p(\tilde{G})$, and a union bound gives the result since P is assumed to have length at most $100n^{1/2} + 1$.

Because the A_v events are independent of each other, our bound above implies that

$$\Pr[C_2(B_p(\tilde{G})) \in \text{ZFS}(C_2(\tilde{G}))] \leq \Pr[B_q(\tilde{G}) \in \text{ZFS}(C_2(\tilde{G}))].$$

Since $C_2(\tilde{G})$ has at most n vertices and minimum degree at least 2, Proposition 1.1 implies

$$\Pr[B_q(\tilde{G}) \in \text{ZFS}(C_2(\tilde{G}))] \leq 2nq^2.$$

Taking $c = 1/17$ and recalling the definition of q in (5) gives the desired result. $\square$

4. Bounds for trees

In this section we prove $\Pr[B_p(T) \in \text{ZFS}(T)] \leq \Pr[B_p(P_n) \in \text{ZFS}(P_n)]$ whenever T is an n -vertex tree with n sufficiently large. We will break our proof into two cases, namely when $p = \Omega(n^{-1})$ and $p = O(n^{-1})$. The intuition for this choice is that when $p \ll n^{-1}$, the probability that $B_p(P_n)$ is zero forcing is roughly the probability of choosing an endpoint, while for $p \gg n^{-1}$ it is roughly the probability of choosing two consecutive vertices. As such, we will need two different arguments for these two regimes.

4.1. Large p

The following provides a concrete statement agreeing with the intuition outlined above.

Lemma 4.1. *Let $v_1, \dots, v_n$ be vertices of a graph G . If $\frac{8}{n} < p < 1$ and $n \geq 16$, then*

$$\Pr[v_1 \in B_p(G) \vee v_n \in B_p(G)] < \Pr[v_i, v_{i+1} \in B_p(G) \text{ for some } i].$$

Proof. Define

$$p_e := \Pr[v_1 \in B_p(G) \vee v_n \in B_p(G)] = 1 - (1 - p)^2,$$

$$p_m := \Pr[v_i, v_{i+1} \in B_p(G) \text{ for some } i] \geq 1 - (1 - p^2)^{\lfloor \frac{n-1}{2} \rfloor},$$

where this last inequality holds because p_m is strictly more than the probability of having at least one of the pairs (v_i, v_{i+1}) with i odd in $B_p(G)$; see the proof of Proposition 1.4 for a more formal argument.

We will show that when $n \geq \frac{4}{p-p^2} + 2$, we have $p_m > p_e$. Indeed, $n \geq \frac{4}{p-p^2} + 2$ is equivalent to $\frac{-2p}{(1-p)} \geq \frac{-p^2(n-2)}{2}$. Using the standard bounds, $\frac{x}{1+x} < \ln(1+x) < x$ for $|x| < 1$, so we have

$$\frac{-2p}{(1-p)} < 2\ln(1-p) \text{ and } \ln(1-p^2) \cdot \left\lfloor \frac{n-1}{2} \right\rfloor < -p^2 \left\lfloor \frac{n-1}{2} \right\rfloor < \frac{-p^2(n-2)}{2}.$$

Thus for this range of n we have $2\ln(1-p) > \left\lfloor \frac{n-1}{2} \right\rfloor \ln(1-p^2)$, or equivalently,

$$1 - (1-p)^2 < 1 - (1-p^2)^{\lfloor \frac{n-1}{2} \rfloor}. \quad (6)$$

Thus $p_m > p_e$ when $n \geq \frac{4}{p-p^2} + 2$. This bound on n holds for all $p \in (\frac{8}{n-2}, 1 - \frac{8}{n-2})$. Equation (6) also holds for $p \in [1 - \frac{8}{n-2}, 1)$ whenever $n \geq 18$, completing the proof. $\square$

Analogous to Proposition 3.1, we can show that graphs with a vertex at the end of two short pendant paths are harder to zero force than paths.

Lemma 4.2. *Let G be an n -vertex graph that has a vertex w which is the endpoint of two pendant paths $u_1 \cdots u_s w$ and $v_1 \cdots v_t w$. If $p \geq 8/(n-s-t)$ and $n-s-t \geq 14$, then*

$$\Pr[B_p(G) \in \text{ZFS}(G)] < \Pr[B_p(P_n) \in \text{ZFS}(P_n)].$$

Proof. Let $w_1, \dots, w_r$ be an arbitrary ordering of $V(G) \setminus \{u_1, \dots, u_s, v_1, \dots, v_t\}$. Relabel the vertices of P_n so that its vertices along the path are $u_1 \cdots u_s w_1 \cdots w_r v_t \cdots v_1$. Because $V(G) = V(P_n)$, we can couple our random variables so that $B_p := B_p(G) = B_p(P_n)$. Let $F = B_p \cap \{u_1, \dots, u_{s-1}, v_1, \dots, v_{t-1}\}$. It suffices to show for all $S \subseteq \{u_1, \dots, u_{s-1}, v_1, \dots, v_{t-1}\}$ that

$$\Pr[B_p(G) \in \text{ZFS}(G) | F = S] \leq \Pr[B_p(P_n) \in \text{ZFS}(P_n) | F = S],$$

with strict inequality for at least one such set. If S contains two consecutive vertices u_i and u_{i+1} , two consecutive vertices v_j and v_{j+1} , or u_1 or v_1 , then $B_p \in \text{ZFS}(P_n)$ so the result holds trivially. Thus from now on we can assume this is not the case.

With this assumption, the vertex u_s in G can only be colored blue by B_p if at least one of u_s or v_t is in B_p (this is because u_s is adjacent to u_{s-1} , which does not enact any forces by assumption on S , and to w , which can only enact a force if at least one of u_s, v_t are colored blue at some point). On the other hand, B_p will be a zero forcing set for P_n provided B_p contains two consecutive vertices from $\{u_s, w_1, \dots, w_r, v_t\}$.

By applying Lemma 4.1 to the vertex set $\{u_s, w_1, \dots, w_r, v_t\}$, we see that if $p > \frac{8}{n-s-t+2}$ and $n-s-t+2 \geq 16$, then

$$\Pr[B_p \text{ contains } u_s \text{ or } v_t] < \Pr[B_p \text{ contains a consecutive pair from } \{u_s, w_1, \dots, w_r, v_t\}].$$

As these two events are independent given the random set F , we conclude that for $p \geq 8/(n-s-t) > 8/(n-s-t+2)$ and $n-s-t+2 \geq 16$,

$$\Pr[B_p(G) \in \text{ZFS}(G) | F = S] < \Pr[B_p(P_n) \in \text{ZFS}(P_n) | F = S],$$

and from this we conclude the result. $\square$

With this we can prove Theorem 1.6 for the case $p = \Omega(n^{-1})$.

Proposition 4.3. *If $T \neq P_n$ is an n -vertex tree with $n \geq 42$, and if $\frac{24}{n} < p < 1$, then*

$$\Pr[B_p(T) \in \text{ZFS}(T)] < \Pr[B_p(P_n) \in \text{ZFS}(P_n)].$$

Proof. Let u, v be any two leaves of T which are at a shortest distance from each other. Observe that the path between u, v consists of two pendant paths, say $uu_2, \dots, u_s w$ and $vv_2, \dots, v_t w$. Because T is not a path, there either exists exactly one leaf $\ell \neq u, v$, or at least two leaves $i, j \neq u, v$.

Suppose for contradiction that $s + t > \frac{2n}{3}$. If T has exactly three leaves, then

$$d(\ell, u) = d(\ell, w) + s < n/3 + s,$$

where this inequality used the fact that none of the internal vertices along the path from ℓ to w use any of the vertices along the path from u to v , of which there are more than $2n/3 + 1$. By a symmetric argument we have $d(\ell, v) < n/3 + t$. In particular, we must have

$$d(\ell, u) + d(\ell, v) < 2n/3 + s + t < 2(s + t) = 2d(u, v),$$

and hence at least one of $d(\ell, u), d(\ell, v)$ is smaller than $d(u, v)$, a contradiction to our choice of u, v . Similarly if T has at least four leaves, then $d(i, j) < \frac{n}{3}$, which again gives a contradiction. Thus we can assume $s + t \leq 2n/3$. With this, our hypothesis implies $p \geq \frac{8}{n-s-t}$ and $n - s - t + 2 \geq 16$, so we can apply Lemma 4.2 to give the desired result. $\square$

4.2. Small p

We will prove our result for small p by upper bounding $z(T; k)$, which we recall is the number of zero forcing sets of T of size k .

Lemma 4.4. *If $T \neq P_n$ is an n -vertex tree, then*

$$z(T; k) \leq \frac{13k^4}{n^2} \binom{n}{k}.$$

Proof. Let Δ denote the maximum degree of T and ℓ the number of leaves of T . Observe that the zero forcing number of a graph is always at least the minimum number of paths needed to cover the vertices of the graph. In particular, every zero forcing set for the tree T has size at least $\ell/2$. It is also known (see for example [20]) that every zero forcing set for a tree T has size at least $\Delta - 1$. Thus for $k < \max\{\Delta - 1, \ell/2\}$, we have $z(T; k) = 0$ and the bound trivially holds. From now on we assume $\max\{\Delta - 1, \ell/2\} \leq k$. The bound is also trivial when $k = n$, so we may assume $k < n$. Lastly, we may also assume $\Delta \geq 3$ since T is not a path.

We first count the number of zero forcing sets S of size k which have two pairs of vertices u, v and x, y with $u \sim v$, $x \sim y$ and $\{u, v\} \cap \{x, y\} = \emptyset$. In this case the number of sets S is at most

$$(n-1)^2 \binom{n-4}{k-4},$$

since one can choose each pair (which is just an edge in T) in at most $n-1$ ways.

We next count the number of zero forcing sets S of size k which contain three vertices u, v, w with $u \sim v \sim w$. The number of such S is at most

$$(n-1)(2\Delta-2) \binom{n-3}{k-3},$$

since one can first choose two adjacent vertices in $n-1$ ways, then a third vertex which is adjacent to at least one of these in at most $2\Delta-2$ ways, and then the remaining vertices in $\binom{n-3}{k-3}$ ways.

We next count the number of S that contain no two adjacent vertices. Because $k < n$ and S is a zero forcing set, at least one vertex of S must be able to force. Because S contains no adjacent vertices, this is only possible if S contains a leaf. Choose such a leaf u_1 to include in S , which can be done in ℓ ways. Let $u_1 u_2 \dots u_s$ be the unique path in T with $\deg(u_i) = 2$ for $1 < i < s$ and $\deg(u_s) \neq 2$.

Claim 4.5. *The set S either contains a leaf $v \neq u_1$, or a neighbor of u_s other than u_{s-1} .*

Proof. Assume this were not the case. Because S contains no two adjacent vertices, no additional leaves, and no other neighbor of u_s , it is not difficult to see that the only vertices that will be colored blue by S are $S \cup \{u_2, \dots, u_s\}$. Because S is a zero forcing set, we must have $V(T) = S \cup \{u_2, \dots, u_s\}$. However, by assumption the only leaves that could be in $S \cup \{u_2, \dots, u_s\}$ are u_1 and u_s , but $T \neq P_n$ contains at least three leaves, so $V(T) \neq S \cup \{u_2, \dots, u_s\}$, giving the desired contradiction. $\square$

In total then, we see that the number of choices for such a set S is at most

$$\ell(\ell + \Delta - 1) \binom{n-2}{k-2},$$

where the terms in the expression above count the number of choices for u_1 , followed by the number of choices for some additional leaf or neighbor of u_s , followed by the number of arbitrary sets of $k-2$ vertices.

It remains to count S that have exactly one pair of adjacent vertices. One can first choose the adjacent pair $u_1, v_1 \in S$ in at most $n-1$ ways. If $\deg(u_1) = 2$, then let $u_1 \cdots u_s$ be the unique path from u_1 with u_2 the neighbor of u_1 not equal to v_1 , and with $\deg(u_i) = 2$ for all $i < s$ and $\deg(u_s) \neq 2$. If $\deg(u_1) \neq 2$, then we simply consider the 1 vertex path u_1 . Analogously define the path $v_1 \cdots v_t$. As in the previous case, because S contains no other pair of adjacent vertices, it must contain at least one leaf or one neighbor of either u_s or v_t that is not u_{s-1} or v_{t-1} . In total then the number of choices for such an S is at most

$$(n-1)(\ell + 2\Delta - 2) \binom{n-3}{k-3}.$$

In total, $z(T; k)$ is at most

$$\begin{aligned} & (n-1)^2 \binom{n-4}{k-4} + (n-1)(2\Delta - 2) \binom{n-3}{k-3} + \ell(\ell + \Delta - 1) \binom{n-2}{k-2} + (n-1)(\ell + 2\Delta - 2) \binom{n-3}{k-3} \\ & \leq n^2 \cdot \frac{k^4}{n^4} \binom{n}{k} + n(2\Delta - 2) \cdot \frac{k^3}{n^3} \binom{n}{k} + \ell(\ell + \Delta - 1) \cdot \frac{k^2}{n^2} \binom{n}{k} + n(\ell + 2\Delta - 2) \cdot \frac{k^3}{n^3} \binom{n}{k}, \end{aligned}$$

where this last inequality used $n-1 \leq n$ and that $\binom{n-t}{k-t} \leq (k/n)^t \binom{n}{k}$ for all integers $t \geq 0$. Using our assumptions $\Delta - 1 \leq k$ and $\ell \leq 2k$, we find that the above expression is at most $(1 + 2 + 6 + 4) \frac{k^4}{n^2} \binom{n}{k}$ as desired. $\square$

With this we can prove the following.

Proposition 4.6. *For every $C > 0$, there exists an integer n_0 such that for all $n \geq n_0$, if $T \neq P_n$ is an n -vertex tree and $0 < p \leq \frac{C}{n}$, then*

$$\Pr[B_p(T) \in \text{ZFS}(T)] < \Pr[B_p(P_n) \in \text{ZFS}(P_n)].$$

Proof. By the previous lemma and the trivial bound $\binom{n}{k} \leq n^k/k!$, we have

$$\Pr[B_p(T) \in \text{ZFS}(T)] \leq \sum_k z(T; k) p^k \leq \sum_k \frac{13k^4 n^{k-2}}{k!} p^k \leq p \cdot 13Cn^{-1} \sum_k \frac{k^4 C^{k-2}}{k!}.$$

The above sum is convergent, so for n sufficiently large we find

$$\Pr[B_p(T) \in \text{ZFS}(T)] \leq \frac{1}{2}p \leq \Pr[B_p(P_n) \in \text{ZFS}(P_n)],$$

where this latter inequality is strict provided $p > 0$. $\square$

With Propositions 4.3 and 4.6 we can prove Theorem 1.6, which we restate below.

Theorem 1.6. *If T is an n -vertex tree with n sufficiently large, then for all $0 \leq p \leq 1$,*

$$\Pr[B_p(T) \in \text{ZFS}(T)] \leq \Pr[B_p(P_n) \in \text{ZFS}(P_n)],$$

with equality holding if and only if either $p \in \{0, 1\}$ or $T = P_n$.

Proof. The equality of the result trivially holds for either $p \in \{0, 1\}$ or $T = P_n$. If $T \neq P_n$ with n sufficiently large, by Proposition 4.3, the result holds for $24/n < p < 1$, and by Proposition 4.6 the result holds for $0 < p \leq 24/n$. $\square$

5. Concluding remarks

We have many open problems regarding the threshold probability $p(G)$, which we recall is the unique $p \in [0, 1]$ such that $\Pr[B_p(G) \in \text{ZFS}(G)] = \frac{1}{2}$. For example, we conjecture the following refinement of Theorem 1.7.

Conjecture 5.1. *If G is an n -vertex graph which contains a clique of size k , then*

$$p(G) = \Omega(\sqrt{k/n}).$$

Table 1
Thresholds for graph families.

Family	Description	Threshold Probability
K_n	Complete graph on n vertices	$1 - \Theta(n^{-1})$
$\overline{K}_n$	Graph on n isolated vertices	$2^{-1/n}$
$K_{n_1, \dots, n_k}$	Complete multipartite graph	$1 - \Theta_k(\min_i \{n_i^{-1}\})$
P_n	Path on n vertices	$\Theta(n^{-1/2})$
C_n	Cycle on n vertices	$\Theta(n^{-1/2})$
W_n	Wheel on n vertices: $C_{n-1} + K_1$	$\Theta(n^{-1/3})$

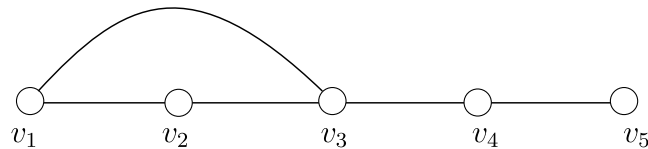


Fig. 4. The triangle with pendant path on five vertices, R_5 .

This conjecture can be viewed as a probabilistic analog to the classical result that $Z(G) \geq k$ if G has a clique of size k , which was proved by Butler and Young [9]. The motivation for the bound $\Omega(\sqrt{k/n})$ comes from considering a graph G which consists of a clique on k vertices, with each of these vertices attached to a path of length roughly n/k . For this graph, a given vertex of the clique will be forced by the path it is connected to with probability roughly $1 - e^{-p^2 n/k}$, so if p is much smaller than $\sqrt{k/n}$, then almost none of the clique vertices in G will be colored blue. Thus $p(G) = \Omega(\sqrt{k/n})$ in this case.⁶

Another natural problem is to compute $p(G)$ for various families of graphs. For example, Table 1 summarizes the order of magnitude of $p(G)$ for many such families. However, one case for which we do not understand $p(G)$ is when G is the n -dimensional hypercube Q_n .

Problem 5.2. Does there exist a constant c such that $p(Q_n) \sim c$? If so, what is this constant?

Because $Z(Q_n) = 2^{n-1}$, we must have $c \geq .5$ if it exists, but beyond this we know nothing about c . The empirical plot in Fig. 2 of the probability that $B_p(Q_8)$ is zero forcing suggests that c might be at least .58. Another family of graphs which we do not understand are grid graphs.

Problem 5.3. Determine the order of magnitude of $p(P_m \square P_n)$, where $P_m \square P_n$ denotes the $m \times n$ grid graph.

Assuming $2 \leq m \leq n$, we can apply Proposition 2.3 with $d = 4$ and $N \approx n^{1/3}$ to show $p(P_m \square P_n) = \Omega(\min\{n^{-1/3}, (mn)^{-1/4}\})$. The best general upper bound we have is $p(P_m \square P_n) = O(n^{-1/2m})$, since at this point it is fairly likely that $B_p(P_m \square P_n)$ contains two consecutive P_m paths, which forces the entire graph. For small m we suspect that our upper bound is closer to the truth than our lower bound, but for large m the situation is unclear.

Another reasonable problem is to study how graph operations affect the threshold probability of a graph. It is perhaps intuitive that $p(C_n) \approx p(P_n)$ since C_n can be formed from P_n by adding a single edge. In fact, it is easily verified that $p(C_n) = \Theta(n^{-1/2})$ by using an argument similar to the proofs in Appendix A. However, there are examples where this intuition fails. Indeed, let $v_1, \dots, v_n$ be the vertices of P_n , and let R_n denote the graph obtained from P_n by adding the edge $v_1 v_3$ (see Fig. 4).

Observe that any zero forcing set of R_n must contain either v_1 or v_2 . Thus,

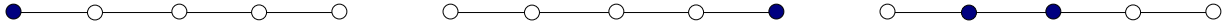
$$\Pr[B_p(R_n) \in \text{ZFS}(R_n)] \leq \Pr[v_1 \in B_p(R_n) \vee v_2 \in B_p(R_n)] \leq 2p.$$

This implies $p(R_n) \in [1/4, 1]$, and hence $p(R_n) = \Theta(1)$.

In the deterministic setting, it is well known that the zero forcing number $Z(G)$ of a graph G changes by at most one if a single edge or vertex is removed from G . This is far from true for $p(G)$. Indeed, recall that $p(P_n) = \Theta(n^{-1/2})$. Let P'_n be obtained by deleting the edge $v_1 v_2$. Since P'_n has K_1 as a connected component, by Observation 1.2 we have $p(P'_n) \geq p(K_1) = \frac{1}{2}$. A similar result holds if one deletes v_2 .

Similarly, deleting edges or vertices can dramatically decrease $p(G)$. Consider the triangle with pendant path R_n , which has $p(R_n) = \Theta(1)$. If one deletes v_1 , then the resulting graph is P_{n-1} , which has $p(P_{n-1}) = \Theta(n^{-1/2})$. If one deletes the edge $v_1 v_3$, then the resulting graph is P_n , which has $p(P_n) = \Theta(n^{-1/2})$.

⁶ In fact, a sharper analysis shows that $p(G) = \Omega(\sqrt{k \log(k)/n})$ for k not too large in terms of n . We suspect that Conjecture 5.1 can be strengthened to include this $\log(k)$ term, but for ease of presentation we have written the conjecture as is.

Fig. 5. Three zero forcing sets for P_5 .

Lastly, one could consider randomized versions of variants of the classical zero forcing number. For example, under skew zero forcing (which was originally introduced in [1]), one can easily generalize Proposition 1.1 to give an upper bound of roughly $\delta np^{\delta-1}$, generalizing the classical result $Z_-(G) \geq \delta - 1$ if G has minimum degree δ . It would also be interesting to consider probabilistic zero forcing with a random set of vertices initially colored blue.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

The authors would like to express their sincere gratitude to the organizers of the Mathematics Research Community on Finding Needles in Haystacks: Approaches to Inverse Problems Using Combinatorics and Linear Algebra, and the AMS for funding the program through NSF grant 1916439.

Appendix A. Threshold probability calculations

In this Appendix we provide proofs of Propositions 1.3 and 1.4. We make use of the following inequalities. Recall that

$$1 - x \leq e^{-x} \quad (7)$$

for all real values x , and

$$\left(1 - \frac{c}{n}\right)^n \geq 1 - c \quad (8)$$

for $|c| \leq n$ and $n \geq 1$.

Proposition 1.3. *If G is a graph on n vertices with no isolated vertices, then $p(G) \leq p(K_n)$. Moreover, $p(K_n) = 1 - \Theta(n^{-1})$.*

Proof. The result is immediate for $n = 1$, so assume $n \geq 2$. Define

$$f(p) = n(1 - p)p^{n-1} + p^n,$$

which is the probability that $B_p(G)$ contains at least $n - 1$ vertices. Since every subset of $V(G)$ of size $n - 1$ is a zero forcing set, we have for $p \in [0, 1]$,

$$\Pr[B_p(G) \in \text{ZFS}(G)] \geq f(p) = \Pr[B_p(K_n) \in \text{ZFS}(K_n)],$$

where this equality used that a set S is a zero forcing set of K_n if and only if $|S| \geq n - 1$. Since $\Pr[B_p(G) \in \text{ZFS}(G)]$ and $\Pr[B_p(K_n) \in \text{ZFS}(K_n)]$ are increasing functions of p , we conclude that $p(G) \leq p(K_n)$.

We now prove the asymptotic result. Let $p = 1 - c/n$, where $c \leq n$ is positive. By (8),

$$f(p) \geq p^n \geq 1 - c,$$

which implies $f(p) > 1/2$ if $p > 1 - \frac{1}{2n}$. Similarly, by (7),

$$f(p) = c(1 - c/n)^{n-1} + (1 - c/n)^n \leq ce^{-c+c/n} + e^{-c} \leq ce^{-c/2} + e^{-c},$$

where the last inequality holds since $n \geq 2$. Thus $f(p) < 1/2$ for $n \geq 5$ and $p < 1 - \frac{5}{n}$. We conclude $p(K_n) = 1 - \Theta(n^{-1})$. $\square$

By convention, we assume the vertices $v_1, \dots, v_n$ of P_n are in *path order*, i.e., the edges of P_n are $v_i v_{i+1}$ for $1 \leq i \leq n - 1$. Note that $S \subseteq V(P_n)$ is a zero forcing set if and only if S contains an endpoint or S contains two consecutive vertices (see Fig. 5).

Proposition 1.4. *The threshold probability of the path on n vertices satisfies*

$$p(P_n) = \Theta(n^{-1/2}).$$

Proof. Let $v_1, \dots, v_n$ denote the vertices of P_n (in path order). Define the random variable X to be the number of indices $i \in \{1, 2, \dots, n-1\}$ such that $v_i, v_{i+1} \in B_p(P_n)$. Markov's inequality yields

$$\Pr[X \geq 1] \leq \mathbb{E}[X] = (n-1)p^2.$$

Since $B_p(P_n) \in \text{ZFS}(P_n)$ if and only if either $X \geq 1$ or at least one of $v_1, v_n \in B_p(P_n)$, a union bound now implies

$$\Pr[B_p(P_n) \in \text{ZFS}(P_n)] \leq (n-1)p^2 + 2p.$$

This quantity is less than $1/2$ provided $p = cn^{-1/2}$ for any $c < 1/4$. Thus $p(P_n) = \Omega(n^{-1/2})$.

Next, for $i \in \{1, 2, \dots, n-1\}$, let A_i be the event that $v_i, v_{i+1} \in B_p(P_n)$, and define $A = \bigcup_{i \text{ odd}} A_i$. Then,

$$\begin{aligned} \Pr[B_p(P_n) \in \text{ZFS}(P_n)] &\geq \Pr[A] = 1 - \prod_{i \text{ odd}} (1 - \Pr[A_i]) \\ &= 1 - (1 - p^2)^{\lfloor (n-1)/2 \rfloor} \geq 1 - e^{-p^2 \lfloor (n-1)/2 \rfloor}, \end{aligned}$$

where the first equality follows from the fact that these events are independent, and the last step uses (7). This probability will be greater than $1/2$ for $p = Cn^{-1/2}$ with C sufficiently large. We conclude that $p(P_n) = \Theta(n^{-1/2})$. $\square$

References

- [1] Mary Allison, Elizabeth Bodine, Luz Maria DeAlba, Joyati Debnath, Laura DeLoss, Colin Garnett, Jason Grout, Leslie Hogben, Bokhee Im, Hana Kim, Reshmi Nair, Olga Pryporova, Kendrick Savage, Bryan Shader, Amy Wangsness Wehe, Minimum rank of skew-symmetric matrices described by a graph, *Linear Algebra Appl.* 432 (10) (2010) 2457–2472, IMA-ISU research group on minimum rank.
- [2] Kirk Boyer, Boris Brimkov, Sean English, Daniela Ferrero, Ariel Keller, Rachel Kirsch, Michael Phillips, Carolyn Reinhart, The zero forcing polynomial of a graph, *Discrete Appl. Math.* 258 (2019) 35–48.
- [3] Deepak Bal, Patrick Bennett, Sean English, Calum MacRury, Paweł Prałat, Zero-forcing in random regular graphs, *J. Comb.* 12 (1) (2021) 85–116.
- [4] Francesco Barioli, Wayne Barrett, Shaun M. Fallat, H. Tracy Hall, Leslie Hogben, Bryan Shader, P. van den Driessche, Hein van der Holst, Zero forcing parameters and minimum rank problems, *Linear Algebra Appl.* 433 (2) (2010) 401–411.
- [5] Daniel Burgarth, Vittorio Giovannetti, Full control by locally induced relaxation, *Phys. Rev. Lett.* 99 (Sep 2007) 100501.
- [6] Allan Bickle, The k -Cores of a Graph, PhD thesis, Western Michigan University, Kalamazoo, Michigan, 2010.
- [7] Natalie C. Behague, Trent G. Marbach, Paweł Prałat, Tight bounds on probabilistic zero forcing on hypercubes and grids, *Electron. J. Comb.* 29 (1) (2022) 1.3.
- [8] Béla Bollobás, Arthur G. Thomason, Threshold functions, *Combinatorica* 7 (1) (1987) 35–38.
- [9] Steve Butler, Michael Young, Throttling zero forcing propagation speed on graphs, *Australas. J. Comb.* 57 (2013) 65–71.
- [10] Yu Chan, Emelie Curl, Jesse Geneson, Leslie Hogben, Kevin Liu, Issac Odegard, Michael Ross, Using Markov chains to determine expected propagation time for probabilistic zero forcing, *Electron. J. Linear Algebra* 36 (2020) 318–333.
- [11] Sean English, Calum MacRury, Paweł Prałat, Probabilistic zero forcing on random graphs, *Eur. J. Comb.* 91 (2021) 103207.
- [12] Alan Frieze, Michał Karoński, Introduction to Random Graphs, Cambridge University Press, Cambridge, 2016.
- [13] Jesse Geneson, Leslie Hogben, Expected propagation time for probabilistic zero forcing, *Aust. J. Comb.* 83 (3) (2022) 397–417.
- [14] A.I.M. Minimum, Rank-Special Graphs Work Group, Zero forcing sets and the minimum rank of graphs, *Linear Algebra Appl.* 428 (7) (2008) 1628–1648.
- [15] Leslie Hogben, Jephian C-H. Lin, Bryan L. Shader, Inverse Problems and Zero Forcing for Graphs, vol. 270, American Mathematical Society, 2022.
- [16] David Hu, Alec Sun, Probabilistic zero forcing on grid, regular, and hypercube graphs, arXiv:2010.12343, 2020.
- [17] Cong X. Kang, Eunjeong Yi, Probabilistic zero forcing in graphs, *Bull. Inst. Comb. Appl.* 67 (2013) 9–16.
- [18] Yuan-Chuan Li, Cheh-Chih Yeh, Some equivalent forms of Bernoulli's inequality: a survey, *Springer Proc. Math. Stat.* 4 (07) (2013) 1070.
- [19] Shyam Narayanan, Alec Sun, Bounds on expected propagation time of probabilistic zero forcing, *Eur. J. Comb.* 98 (2021) 103405.
- [20] Mohammad Reza Oboudi, On the zero forcing number of trees, *Iran. J. Sci. Technol., Trans. A, Sci.* 45 (3) (2021) 1065–1070.
- [21] Darren D. Row, A technique for computing the zero forcing number of a graph with a cut-vertex, *Linear Algebra Appl.* 436 (12) (2012) 4423–4432.
- [22] Min Zhao, Liying Kang, Gerard J. Chang, Power domination in graphs, *Discrete Math.* 306 (15) (2006) 1812–1816.