

FROG MODEL WAKEUP TIME ON THE COMPLETE GRAPH

NIKKI CARTERN, BRITTANY DYGERT, MATTHEW JUNGE, STEPHEN LACINA,
COLLIN LITTERELL, AUSTIN STROMME, AND ANDREW YOU

ABSTRACT. The frog model is a system of random walks where active particles set sleeping particles in motion. On the complete graph with n vertices it is equivalent to a well-understood rumor spreading model. We give an alternate and elementary proof that the wake-up time, i.e. the expected time for every particle to be activated, is $\Theta(\log n)$. Additionally, we give an explicit distributional equation for the wakeup time as a weighted sum of geometric random variables. This project was part of the University of Washington Research Experience for Undergraduates program.

1. INTRODUCTION

The *frog model* starts with an awake frog at the root of a graph and one sleeping frog at each other vertex. In discrete time, awake frogs perform nearest neighbor simple random walks and wake any sleeping frogs they encounter. When first introduced by K. Ravishankar about twenty years ago, the model was known as the egg-model (see [TW99]); R. Durrett is credited with the zoomorphism of viewing particles as frogs. This likely comes from the chaotic way particles wake up. We study the model on the complete graph with n vertices, K_n . In particular we deduce that the *wakeup time* T_n , the time for all frogs to wake up, has expected value on the order of $\log n$.

It was brought to our attention in the final stages of this project that the wakeup time for the frog model on K_n is equivalent to a rumor spreading model introduced in [FG85]. This model starts with a town of n people where one knows a rumor. At each time step those who know the rumor call a uniformly random resident and inform them. The frog model can be naturally coupled with the spread of the rumor so that the number of awake frogs is the same as the number of informed residents. Hence the wakeup time is equivalent to all n residents knowing the rumor. We remark that the locations of the frogs are an extra bit of randomness not accounted for in the rumor spreading model. For this reason, the coupling only works on K_n .

In [FG85] they show that $T_n / \log_2 n \xrightarrow{\mathbf{P}} 1 + \log n + O(1)$ which implies our theorem. The idea of the proof in [FG85] is to break up the spread of the rumor into five stages (see the appendix for a list of the stages). For example, phase one is the time to wake up N frogs with N some large fixed constant. A finer analysis in [Pit87] shows that $T_n = \log_2 n + \log n + O(1)$ in probability. They show that waking is closely approximated by a deterministic equation (see the appendix). A tight analysis for $\mathbf{E}T_n$ is in [DK], where they use three phases and some sophisticated estimates to show that $\mathbf{E}T_n = c \log n + b + o(1)$ with $c = 1 + (1/\log(2)) \approx 2.44$ and $b = 2.765$. Our result is less precise, but the proof is more elementary. We use two phases and only rely on couplings and Markov's inequality.

Although it is equivalent to the spread of a rumor on K_n , the wakeup time for the frog model has otherwise not been studied. The recent article [HJJ15, Open Question 5] introduces the problem on finite d -ary trees. A survey article [Pop03] asks a similar question

for a variant of the frog model where frogs perish after taking t steps. They propose a study of the minimal t that guarantees at least half the frogs on a given graph will be activated with probability greater than $1/2$. It is claimed (without proof) that this value on the complete graph is $O(\log n)$. Though not equivalent, this is closely related to our result for the frog model cover time.

On finite graphs the frog model is a model for epidemics, or the spread of a rumor. The article [CF09] describes many related variants. It also appears in physics literature as a model for combustion [RS04] known as $A + B \mapsto 2A$, where we replace awake and sleeping frogs with flames and fuel, respectively. The combustion process is studied on \mathbb{Z}^d . Noteworthy theorems include the fact that the origin is visited infinitely often for all $d \geq 1$ [TW99] and a shape theorem that says, when properly scaled, the set of activated vertices converges to a convex region in the unit simplex ([AMP02, AMPR01]).

A model that is in a loose sense the reverse of the frog model is coalescing random walk. Introduced in [EN74], this is the process that starts with a particle at each site and when particles collide they coalesce into one. Like the frog model, this is typically studied in \mathbb{Z}^d . For instance, coalescing random walk is recurrent for all $d \geq 1$. This was first shown in [BG80], and refined further in [Gri78, vdBK00, Arr83, Arr81]. Computer science literature studies coalescing random walk on finite graphs. Of particular interest is the *coalescence time*; the expected time for all particles to coalesce into a single particle. In [Cox89] they study this on the torus. The recent paper [CEOR12] gives bounds on rather general graphs.

Our result should be compared with the cover time for multiple random walks on a graph ([ES11] and [AAK⁺07])). The basic question is how the cover time is reduced by using the combined ranges of k random walks. This is studied on a variety of different graphs, and the speedup depends on the graph structure. For the complete graph, [AAK⁺07] cites the folklore (we give a proof in Lemma 2 (IV)) that the speedup is linear, meaning the cover time for a single random walk is k times the cover time for k random walkers. All of the results for speedup of multiple walk take the worst-case scenario across every starting configuration for the k walkers. The frog model is different in that we have only one possible starting configuration, and just one active particle. However, the placement of sleeping frogs is optimal in the sense that activated particles are more likely to be near unexplored sites. We ask a question regarding this in Further Questions (i).

1.1. Main theorem and overview. Before stating the theorem we review *asymptotic notation*. We say that $f(n) = O(g(n))$ if there exists $c > 0$ such that for all sufficiently large n it holds that $f(n) \leq cg(n)$. We write $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ and $g(n) = O(f(n))$. The wakeup time, T_n , is formally defined in the next section.

Theorem 1. $\mathbf{E}T_n = \Theta(\log(n))$.

Additionally, in Section 4 we give an explicit recursive formula for the distribution of T_n . This is in Proposition 5. The formula involves some sophisticated combinatorial objects and, combined with the formula in [DK] for $\mathbf{E}T_n$, yields a bound on their growth that could be of independent interest.

Most of our work is done on K_n° , the complete graph with a self loop at each vertex. In Lemma 2 (IV) we show the frog model on K_n° has a stochastically larger wakeup time. We then prove Theorem 1 in two phases. First, we show that it takes logarithmic time to wake the first $n/2$ frogs. This is done by embedding a process that grows slower than the frog model, but still (on average) grows exponentially. The idea is to, when say k frogs are awake, only allow more frogs to wake up if at least αk asleep frogs are visited. This occurs

with some probability $q_{k,n}$. We show in Proposition 3 that

$$\inf_{n \geq 3} \left(\min_{k < n/2} q_{k,n} \right) \geq p_* > 0.$$

This lower bound is obtained by having the k awake frogs jump one at a time, and thinking of the number of single jumps to wake the $(i + 1)$ st frog as a Geometric with mean $\frac{n-k-i}{n}$ waiting time. An application of Markov's inequality gives a uniform bound in terms of α for all $k < n/2$ and $n \geq 3$.

Next, thinking of each time an α proportion wakes up as a Bernoulli($q_{k,n}$) trial, the number of successes after t steps is stochastically larger than a $\text{Bin}(t, p_*)$ random variable. Thus, the number of frogs awake at time t is stochastically larger than

$$(1 + \alpha)^{\text{Bin}(t, p_*)}.$$

Moreover, the time for this quantity to exceed $n/2$ is a sum of $O(\log n)$ geometric random variables with mean p_* . This is made formal in Lemma 2 (V). We can conclude that the expected time it takes to wake the first $n/2$ frogs is $O(\log n)$.

Once half the frogs are awake, we ignore the contribution of any new frogs added and show that $n/2$ frogs cover the remaining vertices in $O(\log n)$ steps. This is made precise in Lemma 2 (IV) by reducing to the coupon collector problem.

1.2. Further Questions. The wakeup time for the frog model is a largely unexplored topic. There are many further questions one could ask. We remark that [Pop03] and [HJJ15] discuss a few other problems on finite graphs.

- (i) It is interesting to compare the wakeup time for the frog model on G , a graph with n vertices, to the cover time for n independent random walks on G started in the least optimal starting configuration. Perhaps the frogs being evenly spread might overcome the disadvantage of starting with only one awake particle. *Are there graphs for which the frog model wakeup time is faster than the cover time for n -multiple random walks?* The full binary tree of height n is a good candidate. The expected cover time with 2^n particles started at the same leaf is $O(n^2(2/\sqrt{2})^n)$ (see [Sau10]), whereas the wakeup time is conjectured to be polynomial in n (see [HJJ15]).
- (ii) Let $G(n, p)$ denote an Erdős-Rényi graph (i.e. the random graph obtained by keeping each edge in K_n with probability p). *What is the wakeup time for the frog model on $G(n, p)$?* For fixed $p > 0$, this should still with high probability be $O(\log n)$, but for p_n decaying with n the graph is sparser and the wakeup time is possibly larger.
- (iii) *What is the wakeup time for other graphs?* For instance, the path, cycle, and grid.

2. FORMAL MODEL AND COUPLINGS

Here we give a formal definition of the frog model. Then we describe in Lemma 2 the five couplings we depend on in proving our main theorem.

2.1. Formal definition of the frog model. We borrow much of our notation from [AMP02]. Let V be the vertex set of K_n . Consider the collection, $\{F_v(t) : v \in V\}$, of independent random walks on K_n each satisfying $F_v(0) = v$. These random walks correspond to the trajectory of each frog. We now introduce stopping times to account for the waking up that occurs. Define

$$t(v, u) = \min_t \{F_v(t) = u\},$$

the time that the frog originally at vertex v takes to reach vertex u . Also define

$$T(v, u) = \inf \left\{ \sum_{i=1}^k t(v_{i-1}, v_i) : v_0 = v, \dots, v_k = u \text{ for some } k \right\},$$

the first passage time from v to u in the frog model. Then $T(v_0, u)$ gives the time it takes for u to be woken. For each $u \in V$, the position of the frog originally at u at time t is defined to be

$$P_u(t) = \begin{cases} u, & t \leq T(v_0, u) \\ F_u(t - T(v_0, u)), & t > T(v_0, u) \end{cases}.$$

With this we can define $\Lambda(t) = \{u \in V : T(v_0, u) \leq t\}$, the set of sites that have been visited by time t or the set of awake frogs at time t . Define the number of frogs awake at time t to be $N_t = |\Lambda(t)|$. Thus, the time to wake all of the frogs is $T_n = \inf\{t : N_t = n\}$.

2.2. Couplings and stochastic dominance. The frog model only depends on the underlying random walk trajectories. It has the nice feature that restricting the range of frogs, or ignoring woken frogs yields models with monotonically slower waking behavior. This is made formal using couplings and stochastic dominance.

Let X and Y be two random variables. If for each $a > 0$ we have $\mathbf{P}[Y \geq a] \geq \mathbf{P}[X \geq a]$ then we say that Y *stochastically dominates* X , written $X \preceq Y$. A thorough reference on stochastic domination is [SS07]. Note that if $A \succeq B$, then $\mathbf{E}A \geq \mathbf{E}B$. An equivalent condition to stochastic dominance is that $X \preceq Y$ if and only if there exists a coupling (X, Y) with $X \leq Y$ a.s. Formally, a *coupling* is a probability space with (possibly dependent) random variables X' and Y' that have the same distribution as X and Y , respectively. Couplings can often be described intuitively and rigorously in words. In the following lemma we describe all of the couplings used in this paper.

Lemma 2. *The following stochastic dominance relations hold:*

(I) *Let $\{q_i\}_{i=1}^t$ be a sequence in $[0, 1]$ with $q_i > p$ for all $i = 1, 2, \dots, t$. It holds that*

$$\sum_{i=1}^t \text{Ber}(q_i) \succeq \text{Bin}(t, p).$$

Here $\text{Ber}(p)$ denotes a Bernoulli- p random variable and $\text{Bin}(t, p)$ is a binomial random variable with t trials.

(II) *Let K_n° be the complete graph with a self-loop added to each vertex. If T_n° is the wakeup time for the frog model on K_n° , then*

$$T_n \preceq T_n^\circ.$$

(III) *Let $\tau_{n/2}$ and $C_{n/2}$ be as defined in the proof of Theorem 1. It holds that*

$$T_n \preceq \tau_{n/2} + C_{n/2}.$$

(IV) *Let K_n° be the complete graph with a self-loop added to each vertex. Define C_k° to be the time for k random walks to collectively visit every vertex of K_n° , and define C_k analogously for K_n . It holds that*

$$C_k \preceq C_k^\circ \stackrel{d}{=} \frac{C_1^\circ}{k} \stackrel{d}{=} \frac{1}{k} \sum_{i=1}^n \text{Geo}\left(\frac{n-i}{n}\right). \quad (1)$$

Where $\text{Geo}(p)$ is the number of Bernoulli- p trials until a success occurs.

- (V) Consider a modified frog model on K_n° where, when k frogs are active, more frogs wake up only if at least αk sleeping frogs are visited at the next step (for some fixed $\alpha > 0$). When this occurs we select an arbitrary subset of $\lceil \alpha k \rceil$ of these frogs and allow them to wake up. The others remain asleep. Thus, on this event there are at least $(1 + \alpha)k$ frogs awake. Proposition 3 shows that waking at least αk frogs occurs with probability at least $p_* > 0$ for any k and and sufficiently large n . We then have

$$(1 + \alpha)^{\text{Bin}(t, p_*)} \leq N_t,$$

and for $\tau_{n/2} := \inf\{t: N_t \geq n/2\}$ and $n_* := \lfloor \frac{\log(n/2)}{\log(1+\alpha)} \rfloor$ we have

$$\tau_{n/2} \leq \inf\{t: (1 + \alpha)^{\text{Bin}(t, p_*)} \geq n/2\} \stackrel{d}{=} \sum_1^{n_*} \text{Geo}(p_*).$$

Proof. All of the proofs establish stochastic dominance via couplings.

- (I) Define $X = \sum_{i=1}^t \text{Ber}(q_i)$ and $Y = \sum_{i=1}^t \text{Ber}(p) = \text{Bin}(t, p)$. Let $\{U_i\}_{i=1}^t$ be uniform $[0, 1]$ random variables, so that

$$X \stackrel{d}{=} \sum_{i=1}^t 1\{U_i \leq q_i\},$$

$$Y \stackrel{d}{=} \sum_{i=1}^t 1\{U_i \leq p\}.$$

Our hypothesis $q_i > p$ guarantees that $X \geq Y$ for all realizations of the U_i .

- (II) Pair the frogs on K_n and K_n° in the natural way. Whenever a frog on K_n° moves to a new vertex, have the corresponding frog on K_n follow it. In this way, the frogs on each graph perform random walks, but those on K_n° possibly spend extra steps traveling self-loops. This coupling ensures that $T_n \leq T_n^\circ$ in every realization of the model.
- (III) Run the frog model up to time $\tau_{n/2}$. Of the $N_{\tau_{n/2}}$ frogs awake choose a batch of $n/2$ of them. Now think of this batch as paired to another frog model in the same configuration as ours at time $\tau_{n/2}$. Our $n/2$ -batch frogs follow their counterparts. The time, $C_{n/2}$, it takes for the batch to visit all n vertices of K_n is at least as large as the $T_n - \tau_{n/2}$ steps taken by the frog model they are coupled with. In this way, the model restricted to a batch spends $\tau_{n/2} + C_{n/2}$ steps, which is at least the T_n steps taken by the frog model.
- (IV) A similar coupling as in (II) gives $C_k \leq C_k^\circ$. Observe that on K_n° every site is accessible in one step. Thus, the set of sites visited by k random walks has the same law as the range of a single random walk in k steps. It follows that $kC_k^\circ \stackrel{d}{=} C_1^\circ$. The last equality

$$C_1^\circ \stackrel{d}{=} \sum_{i=1}^n \text{Geo}\left(\frac{n-i}{n}\right)$$

follows from the observation that the waiting time for a single random walk on K_n° to increase its range from i to $i+1$ is the waiting time to have a success in a sequence of Bernoulli $((n-i)/n)$ trials. This is distributed as a $\text{Geo}((n-i)/n)$. As increases in the range are independent and skip-free, the claimed formula follows.

- (V) By (II) we preserve the dominance relation by working on K_n° . Since each successful increase in the number of frogs is a Bernoulli trial with probability at least p_* it

follows from (I) that this further dominates the random quantity $(1 + \alpha)^{\text{Bin}(t, p_*)}$. The fact that N_t dominates this quantity is a straightforward consequence of the fact that we are only ignoring frogs. And, the fact that $\tau_{n/2}$ is less than the time for $(1 + \alpha)^{\text{Bin}(t, p_*)}$ to exceed $n/2$ follows immediately from the relationship between $(1 + \alpha)^{\text{Bin}(t, p_*)}$ and N_t .

Taking the log of each side, we have the stopping time $\inf\{t: (1 + \alpha)^{\text{Bin}(t, p_*)} \geq n/2\}$ is equivalent to the time for a $\text{Bin}(t, p_*)$ random variable to exceed $\log(n/2)/\log(1 + \alpha)$. The claimed distributional equality is just the fact that $\text{Bin}(t, p_*)$ is a skip-free process which increases independently at each increment after $\text{Geo}(p_*)$ steps. \square

3. PROVING THEOREM 1

We start by elaborating a bit on Lemma 2 (V). Let $\alpha < 1$ be a yet to be chosen parameter. Define the probabilities $q_{k,n} = q_{k,n}(\alpha)$ that the frog model on K_n° with k frogs awake wakes at least αk frogs in one time step. Let $p_* = p_*(\alpha) = \inf_{n \geq 3} (\min_{k < n/2} q_{k,n})$. Our first goal is to establish that p_* is bounded away from 0.

Proposition 3. *For $\alpha = 1/10$ it holds that $p_* \geq 1/37$.*

Proof. It is useful to decompose one time step in the frog model with k frogs awake on K_n° into k steps by a single random walk. Notice that the number of steps by the single random walk to visit the first sleeping frog is the waiting time for a success in a sequence of $\text{Ber}(\frac{n-k}{n})$ trials (i.e. a $\text{Geo}(\frac{n-k}{n})$ random variable). Similarly, once i of the $n - k$ frogs have been woken the waiting time is $\text{Geo}(\frac{n-k-i}{n})$. Let $X = X(k, n, \alpha) = \sum_{i=0}^{\lceil \alpha k \rceil} \text{Geo}(\frac{n-k-i}{n})$. This represents the number of the k awake frogs that jump in order to wake $\lceil \alpha k \rceil$ more. It follows that $q_{k,n} = 1 - \mathbf{P}[X > k]$. And, by Markov's inequality

$$q_{k,n} \geq 1 - \frac{\mathbf{E} X}{k}. \quad (2)$$

We can use linearity and the fact that the mean of a $\text{Geo}(p)$ random variable is $1/p$ to estimate $\mathbf{E} X$:

$$\mathbf{E} X = \sum_{i=0}^{\lceil \alpha k \rceil} \frac{n}{n-k-i} = n \sum_{i=n-\lceil \alpha k \rceil}^n \frac{1}{i} \leq n \frac{n - (n - \lceil \alpha k \rceil)}{n - \lceil \alpha k \rceil} = \frac{n \lceil \alpha k \rceil}{n - \lceil \alpha k \rceil}.$$

Since $\lceil \alpha k \rceil \leq \alpha k + 1$ we can bound $\mathbf{E} X/k$ by

$$\begin{aligned} \frac{\mathbf{E} X}{k} &\leq \frac{\alpha k + 1}{k} \frac{n}{n - \alpha k - 1} \\ &= \frac{\alpha n}{n - \alpha k - 1} + \frac{1}{k} \frac{n}{n - \alpha k - 1}. \end{aligned}$$

Note that $q_{1,n} = 1 - \frac{1}{n} \geq \frac{2}{3}$ since we assume $n \geq 3$. Thus, we can work with $k \geq 2$. Also by assumption k is no larger than $n/2$. We then preserve the above bound by setting $k = n/2$ in the negative terms and $k = 2$ in the $\frac{1}{k}$ term. This results in

$$q_{k,n} \geq 1 - \frac{\alpha n}{n - \alpha n/2 - 1} - \frac{n}{2(n - \alpha n/2 - 1)} = 1 - \frac{\alpha + \frac{1}{2}}{1 - \frac{\alpha}{2} - 1/n}.$$

As $n \geq 3$ we arrive at $p_* \geq 1 - \frac{\alpha + \frac{1}{2}}{\frac{2}{3} - \frac{\alpha}{2}}$. If we evaluate at $\alpha = 1/10$, then $p_* \geq 1/37$, which completes the proof. \square

3.1. Proof of Theorem 1. With Lemma 2 and Proposition 3 we can prove our main theorem.

Proof of Theorem 1. Since the number of frogs can at most double at each step we have $T_n \geq \log_2(n)$. The lower bound immediately follows. As for the asymptotic upper bound, let $\tau_{n/2} = \inf\{t: N_t \geq n/2\}$ be the time to wake at least $n/2$ frogs, and let $C_{n/2}$ be the time for $n/2$ walkers to visit every vertex of K_n (taken to be the maximum such time over all possible starting configurations of walkers). Lemma 2 (III) describes a coupling where we ignore the benefit of waking more frogs after time $\tau_{n/2}$ to conclude

$$T_n \preceq \tau_{n/2} + C_{n/2}. \quad (3)$$

Here ‘ \preceq ’ denotes stochastic domination, see Section 2.2 for the definition. The couplings in Lemma 2 (IV) and (V) along with Proposition 3 imply that there exists $\alpha, p_* > 0$ such that

$$\tau_{n/2} \preceq \sum_{i=1}^{\lfloor \frac{\log(n/2)}{\log(1+\alpha)} \rfloor} \text{Geo}(p_*) \quad \text{and} \quad C_{n/2} \preceq \frac{2}{n} \sum_{i=1}^n \text{Geo}\left(\frac{n-i}{n}\right).$$

Using the fact that the expectation of a Geometric(p), random variable is $1/p$ we can take the expectation of both sides of (3) to obtain

$$\mathbf{E}T_n \leq (1/p_*) \left\lfloor \frac{\log(n/2)}{\log(1+\alpha)} \right\rfloor + \frac{2}{n} \sum_{i=1}^{n-1} \frac{n}{n-i} = O(\log n).$$

Note the first summand above is $O(\log n)$ because p_* and α are positive. The second summand is $O(\log n)$ by canceling the factors of $1/n$ and n then comparing to the harmonic numbers $\sum_{i=1}^n \frac{1}{i} \approx \log n$. \square

4. EXACT DISTRIBUTION OF T_n

Let $\sigma_k = \sigma_k(n)$ be the time to wake up all n frogs on the complete graph given that there are k frogs currently awake. So, $\sigma_1 = T_n$.

Proposition 4. *Let*

$$p_{j,k} = \mathbf{P}[k \text{ awake frogs visit } j \text{ sleeping frogs on the next step}].$$

It holds that

$$p_{j,k} = \frac{1}{(n-1)^k} \sum_{\ell=j}^k \binom{k}{\ell} \binom{n-k}{j} S_{j,k}^\ell (k-1)^{k-\ell},$$

where $S_{j,k}^\ell$ the number of ways to distribute ℓ balls into k boxes so that no box is empty. This is given by the formula

$$S_{j,k}^\ell = j! \cdot S(k, \ell) = \sum_{k=1}^j \binom{j}{j-k} (-1)^{j-k} k^\ell,$$

where $S(k, \ell)$ is Stirling’s number of the second kind.

Proof. Observe that to wake up j more frogs, we can use between j and all k frogs, and distribute them onto j unvisited vertices. So let ℓ range between j and k . There are $\binom{k}{\ell}$ ways to choose ℓ of the already awoken frogs to visit the j new vertices. We can choose j new vertices in $\binom{n-k}{j}$ ways. Once we have chosen the ℓ frogs and j new vertices, we can

distribute them in (by definition) $S_{j,k}^\ell$ ways. Finally the remaining $k - \ell$ frogs can go to any of the $k - 1$ already visited vertices they are adjacent to, so they can move in $(k - 1)^{k - \ell}$ ways. Also, since there are k frogs and we are thinking of them as distinguishable, each one of these events happens with probability $1/(n - 1)^k$. Thus we get, by summing over all ℓ , that

$$p_{j,k} = \frac{1}{(n - 1)^k} \sum_{\ell=j}^k \binom{k}{\ell} \binom{n - k}{j} S_{j,k}^\ell (k - 1)^{k - \ell}.$$

We claim that for fixed ℓ :

$$S_{j,k}^\ell = \sum_{k=1}^j \binom{j}{j - k} (-1)^{j - k} k^\ell$$

To see this we proceed by inclusion exclusion principle. Observe j^ℓ counts the number of ways to distribute ℓ distinguishable balls to j distinguishable boxes with some boxes left empty possibly. So to count the number of ways with no boxes left empty, we should subtract the number with at least one empty, which is $\binom{j}{j - 1} (j - 1)^\ell$ since we have $j - 1$ boxes that we will possibly place balls in, and have $j - 1$ choices for each of the balls. But now all the ways with exactly two boxes left empty have been added once and subtracted twice (since we counted them $\binom{j}{1}$ times in the subtraction), so we should add $\binom{j}{j - 2} (j - 2)^\ell$ to count the number of ways with at least two left empty. Keep going in this fashion to get

$$S_{j,k}^\ell = \sum_{k=1}^j \binom{j}{k} (-1)^{j - k} (j - k)^\ell = \sum_{k=1}^j \binom{j}{j - k} (-1)^{j - k} (j - k)^\ell.$$

□

An explicit formula for time to wake all n frogs on the complete graph with n vertices is defined recursively as follows:

Proposition 5. *Let $p_{j,k}$ be as in Proposition 4 and let $p'_{j,k} = \frac{1 - p_{0,k}}{p_{0,k}}$. The random variables σ_k satisfy the following recursive distributional relationship*

$$\sigma_1 \stackrel{d}{=} 1 + \sigma_2, \\ \sigma_k \stackrel{d}{=} \begin{cases} \text{Geo}(1 - p_{0,k}) + \sum_{j=k+1}^{2k} p'_{j,k} \sigma_{k+j}, & 2 \leq k \leq \frac{n}{2} \\ \text{Geo}(1 - p_{0,k}) + \sum_{j=k+1}^n p'_{j,k} \sigma_{k+j}, & \frac{n}{2} < k \leq n - 1 \end{cases}.$$

Recall that $T_n = \sigma_1$.

Proof. The expression for σ_1 is the observation that after one step there will always be two frogs awake. When $k \geq 2$ frogs are awake the time to wake more frogs is a geometric random variable with mean $1 - p_{0,k}$. Conditioned that the k frogs wake another frog, we obtain j more awake frogs with probability $p'_{j,k}$. In this situation we now must wait σ_{k+j} steps. □

5. APPENDIX

Phases for argument in [FG85].

- i. The time inform N residents for some fixed constant N (not growing with n).
- ii. The time to go from N to ζn informed residents with $0 < \zeta < 1$ a fixed constant.

- iii. The time to go from ζn to $(1 - \epsilon)n$ with $\epsilon > 0$ a small fixed constant.
- iv. The time to go from $(1 - \epsilon)n$ to $n - R$ informed residents where R is a large fixed constant.
- v. The time to go from $n - R$ to n informed residents.

Deterministic equation in [Pit87]. Letting $N(t)$ be the number of informed residents at time t :

$$N(t+1) = n - (n - N(t)) \exp(-N(t)/n).$$

Phases for argument in [DK].

- i. The time inform \sqrt{n} residents.
- ii. The time to go from \sqrt{n} to $n/2$ informed residents.
- iii. The time to go from $n/2$ to n informed residents.

REFERENCES

- [AAK⁺07] N. Alon, C. Avin, M. Koucky, G. Kozma, Z. Lotker, and M. R. Tuttle, *Many Random Walks Are Faster Than One*, ArXiv e-prints (2007).
- [AMP02] O. S. M. Alves, F. P. Machado, and S. Yu. Popov, *The shape theorem for the frog model*, Ann. Appl. Probab. **12** (2002), no. 2, 533–546. MR 1910638 (2003c:60159)
- [AMPR01] O. S. M. Alves, F. P. Machado, S. Yu. Popov, and K. Ravishankar, *The shape theorem for the frog model with random initial configuration*, Markov Process. Related Fields **7** (2001), no. 4, 525–539. MR 1893139 (2003f:60171)
- [Arr81] Richard Arratia, *Limiting point processes for rescalings of coalescing and annihilating random walks on \mathbb{Z}^d* , Ann. Probab. **9** (1981), no. 6, 909–936.
- [Arr83] ———, *Site recurrence for annihilating random walks on \mathbb{Z}^d* , Ann. Probab. **11** (1983), no. 3, 706–713.
- [BG80] Maury Bramson and David Griffeath, *Asymptotics for interacting particle systems on \mathbb{Z}^d* , Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **53** (1980), no. 2, 183–196 (English).
- [CEOR12] Colin Cooper, Robert Elsässer, Hirotaka Ono, and Tomasz Radzik, *Coalescing random walks and voting on graphs*, CoRR [abs/1204.4106](https://arxiv.org/abs/1204.4106) (2012).
- [CF09] Collin Cooper and Alan Frieze, *Multiple random walks and interacting particle systems*, ICALP (2009), 399–410.
- [Cox89] J. T. Cox, *Coalescing random walks and voter model consensus times on the torus in \mathbb{Z}^d* , Ann. Probab. **17** (1989), no. 4, 1333–1366.
- [DK] Benjamin Doerr and Marvin Knemmann, *Tight analysis of randomized rumor spreading in complete graphs*, ch. 7, pp. 82–91.
- [EN74] P. Erdos and P. Ney, *Some problems on random intervals and annihilating particles*, Ann. Probab. **2** (1974), no. 5, 828–839.
- [ES11] Robert Elsässer and Thomas Sauerwald, *Tight bounds for the cover time of multiple random walks*, Theoretical Computer Science **412** (2011), no. 24, 2623 – 2641, Selected Papers from 36th International Colloquium on Automata, Languages and Programming (ICALP 2009).
- [FG85] A.M. Frieze and G.R. Grimmett, *The shortest-path problem for graphs with random arc-lengths*, Discrete Applied Mathematics **10** (1985), no. 1, 57 – 77.
- [Gri78] David Griffeath, *Annihilating and coalescing random walks on \mathbb{Z}^d* , Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **46** (1978), no. 1, 55–65 (English).
- [HJJ15] C. Hoffman, T. Johnson, and M. Junge, *From transience to recurrence with Poisson tree frogs*, ArXiv e-prints (2015), *To appear in Annals of Applied Probability*.
- [Pit87] Boris Pittel, *On spreading a rumor*, SIAM J. Appl. Math. **47** (1987), no. 1, 213–223.
- [Pop03] Serguei Yu. Popov, *Frogs and some other interacting random walks models*, Discrete random walks (Paris, 2003), Discrete Math. Theor. Comput. Sci. Proc., AC, Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2003, pp. 277–288 (electronic). MR 2042394
- [RS04] Alejandro F. Ramírez and Vladas Sidoravicius, *Asymptotic behavior of a stochastic combustion growth process*, J. Eur. Math. Soc. (JEMS) **6** (2004), no. 3, 293–334. MR 2060478 (2005e:60234)

- [Sau10] Thomas Sauerwald, *Expansion and the cover time of parallel random walks*, Proceedings of the 29th ACM SIGACT-SIGOPS symposium on Principles of distributed computing (New York, NY, USA), PODC '10, ACM, 2010, pp. 315–324.
- [SS07] Moshe Shaked and J. George Shanthikumar, *Stochastic orders*, Springer Series in Statistics, Springer, New York, 2007. MR 2265633 (2008g:60005)
- [TW99] András Telcs and Nicholas C. Wormald, *Branching and tree indexed random walks on fractals*, J. Appl. Probab. **36** (1999), no. 4, 999–1011. MR 1742145 (2001m:60199)
- [vdBK00] J. van den Berg and Harry Kesten, *Asymptotic density in a coalescing random walk model*, Ann. Probab. **28** (2000), no. 1, 303–352.

UNIVERSITY OF PORTLAND
E-mail address: `cartern16@up.edu`

SEATTLE PACIFIC UNIVERSITY
E-mail address: `bddygert@gmail.com`

UNIVERSITY OF WASHINGTON
E-mail address: `jungem@math.washington.edu`

UNIVERSITY OF OKLAHOMA
E-mail address: `stephen.d.lacina-1@ou.edu`

UNIVERSITY OF WASHINGTON
E-mail address: `collin.litterell@gmail.com`

UNIVERSITY OF WASHINGTON
E-mail address: `austinjstromme@gmail.com`

DUKE UNIVERSITY
E-mail address: `andrew.kh.you@gmail.com`