How Asynchrony Affects Rumor Spreading Time

George Giakkoupis  
INRIA Rennes, France  
george.giakkoupis@inria.fr

Yasamin Nazari  
U. of Calgary, Canada  
ynazari@ucalgary.ca

Philipp Woelfel  
U. of Calgary, Canada  
woelfel@ucalgary.ca

July 9, 2015

Abstract

In standard randomized rumor spreading, nodes communicate in synchronized rounds. In each round every node contacts a random neighbor in order to exchange the rumor (i.e., either push the rumor to its neighbor or pull it from the neighbor). A natural asynchronous variant of this algorithm is one where each node has an independent Poisson clock with rate 1, and every node contacts a random neighbor whenever its clock ticks. This asynchronous variant is arguably a more realistic model in various settings, including message broadcasting in communication networks, and information dissemination in social networks.

In this paper we study how asynchrony affects rumor spreading time, that is, the time before a rumor originated at a single node spreads to all nodes in the graph. Our first result states that the asynchronous rumor spreading time is asymptotically bounded by the standard, synchronous, time. Precisely, we show that for any graph $G$ on $n$-nodes, where the synchronous protocol informs all nodes within $T^*(G)$ rounds with high probability, the asynchronous protocol needs at most time $O(T^*(G) + \log n)$ to inform all nodes with high probability. On the other hand, we show that the expected synchronous rumor spreading time is bounded by $O(\sqrt{n})$ times the expected asynchronous time.

These results improve upon the currently best known bounds for both directions, shown recently by Acan et al. (PODC 2015). An interesting implication of our first result is that in regular graphs, the weaker push-only variant of synchronous rumor spreading has the same asymptotic performance as the synchronous push-pull algorithm.
1 Introduction

Broadcasting information in large networks is a fundamental and well-studied problem. Desirable properties of broadcasting algorithms are efficiency, simplicity, decentralization, and tolerance to changes in the network topology. The classical abstraction is the randomized (synchronous) rumor spreading protocol [7, 22]: Initially, a piece of information, called rumor, is injected at a random or arbitrarily chosen node. After that, processes communicate in synchronous rounds to inform each other of the rumor. In every round, each node calls a random neighbour and establishes a communication with its callee in order to possibly exchange the rumor: In the push protocol, an informed caller pushes the rumor to its callee, while in the pull protocol, a non-informed caller receives the rumor from its callee, if the callee is informed. The push-pull protocol combines the push and pull communication and allows a bi-directional rumor exchange in each round between each caller and its callee.

Besides being of fundamental interest, rumor spreading protocols have many direct applications, such as in the maintenance of distributed replicated database systems [8, 14], failure detection [25], resource discovery [21], and data aggregation [3]. As such, the rumor spreading time, i.e., the number of rounds until all nodes in a network have received the rumor (either in expectation or with high probability), has been studied intensively. A large body of research work deals with the question how the rumor spreading time is influenced by the network topology (e.g., [1, 5, 9, 13, 16–18]), network parameters, such as expansion [6, 19, 20, 24], or the communication modes push, pull, and push-pull (e.g., [23]).

The synchrony assumption, according to which all processes establish connections simultaneously in a round-by-round fashion, has been criticized for not being plausible in many scenarios [10–12]. Real networks typically do not have a centralized clock, and individual links are affected by frequent changes in communication speed. Moreover, decentralization has been emphasized as one of the main advantages of the rumor spreading protocol, but this contradicts the model assumption of a centralized clock. More recently the performance of rumor spreading protocols in a natural asynchronous setting, initially proposed by Boyd, Ghosh, Prabhakar, and Shah [4], has been considered. Here, nodes establish communications with their neighbours at times determined by independent Poisson processes, as opposed to at fixed unit times. More precisely, each node is equipped with an independent Poisson clock with rate 1, and whenever a node’s clock ticks, it exchanges the rumor with a random neighbour (using either push, pull, or push-pull communication).

On the hypercube, the asynchronous push-pull protocol corresponds to Richardson’s model for the spread of disease, and has been investigated by studying first-passage percolation [2, 15]. But only recently has asynchronous rumor spreading found the interest of researchers in the area of networks, initially in order to understand information spread in social networks. It was observed that on common network topologies modelling social networks, such as certain power law graphs [17] or preferential attachment graphs [10], the push-pull protocol spreads the rumor to a large fraction of the nodes significantly faster in the asynchronous than in the synchronous model. There are even graphs for which the asynchronous protocol has poly-logarithmic rumor spreading time, whereas the synchronous protocol requires polynomially many steps [1]. On the other hand, there are also simple networks, where synchrony yields faster rumor spreading than asynchrony [1]: In an n-vertex star, it takes at most 2 rounds to spread the rumor to all nodes, because it takes at most one round to push the rumor to the centre node, and another round for all remaining nodes to pull the rumor. On the other hand, in the asynchronous model it takes with high probability $\Theta (\log n)$ time until sufficiently many different Poisson clocks have ticked for all nodes to get informed.

This raises the question, how big the gap between the asynchronous and the synchronous rumor spreading times can be. In the following discussion we restrict ourselves to push-pull communica-

1
tion for graphs with \( n \) vertices, unless mentioned otherwise. Acan, Collevecchio, Mehrabian, and Wormald [1] showed that for any graph the high probability rumor spreading time in the asynchronous model is at most an \( O(\log n) \) factor larger than that of the synchronous model. While this result is tight for the \( n \)-vertex star, it may not be tight for graphs that have super-constant synchronous rumor spreading time. In fact, Acan et al. conjectured that the high probability asynchronous rumor spreading time can be at most by an additive \( O(\log n) \) term larger than the synchronous one. Our first main result proves this conjecture up to a constant factor.

**Theorem 1.** Let \( G \) be a graph with \( n \) vertices, \( u \) a vertex of \( G \), and \( T^*(G) \) the smallest integer such that with probability at least \( 1 - 1/n \) it takes at least \( T^*(G) \) synchronous push-pull rounds for the rumor to spread from \( u \) to all other nodes. Then in the rumor spreading time of the asynchronous push-pull protocol for \( G \) is at most \( O(T^*(G) + \log n) \) with high probability.

This bound is asymptotically tight, and in particular it shows that for most graphs, namely those that have at least logarithmic high probability rumor spreading time, the asynchronous rumor spreading time is not asymptotically larger than the synchronous one.

Acan et al. [1] also showed that the high probability rumor spreading time in the synchronous model can be at most by a factor of \( O(n^{2/3}) \) larger than in the asynchronous model. They conjecture that this factor can be improved to \( n^{1/2} \cdot (\log n)^{O(1)} \). Our second main result is a proof of this conjecture:

**Theorem 2.** Let \( G \) be a graph with \( n \) vertices, \( u \) a vertex of \( G \), and \( T(G) \) the number of synchronous push-pull rounds it takes for the rumor to spread from \( u \) to all other nodes. Then the rumor spreading time of the asynchronous push-pull protocol for \( G \) has expectation \( \Omega(E[T(G)]/\sqrt{n}) \).

Using this theorem, the conjecture of Acan et al. follows from the fact that with high probability the synchronous rumor spreading time, \( T(G) \), is at most by an \( O(\log n) \) factor larger than its expectation. I.e., \( T(G) = O(E[T(G)] \cdot \log n) \) with high probability.

Note that in the synchronous model, the push-pull protocol can be significantly faster than the push protocol (but clearly it cannot be slower). For example, for an \( n \)-vertex star it takes with high probability \( \Theta(n \cdot \log n) \) synchronous rounds to push the rumor to all nodes, while it takes only two synchronous push-pull rounds (as discussed earlier). Theorem 1 has the interesting consequence, that push-pull communication can only have performance benefits on irregular graphs. This follows from the following observations: (1) For the push protocol it is known that the high probability synchronous rumor spreading time on any graph is bounded by the asynchronous one [23]. (2) It is not hard to see that on regular graphs, the high probability asynchronous rumor spreading time of the push protocol is not worse than that of the push-pull protocol.\(^1\) (3) Theorem 1 implies that on regular graphs the high probability asynchronous rumor spreading time is not larger than the synchronous one (because the latter is at least \( \log n \)). To summarize, we obtain the following, informally stated relations for the asymptotic high probability rumor spreading times on any regular graph:

\[^1\]We couple an asynchronous push algorithm that runs at double the speed (i.e., each node takes a step with rate 2 rather than 1), with the normal asynchronous push-pull algorithm. We assume a \( d \)-regular graph \( G \). For push, we have a Poisson clock \( C_{u,v} \) with rate \( 2/d \) for each directed pair of adjacent nodes \( u, v \), while for push-pull, we have a Poisson clock \( C'_{u,v} \) with rate \( 2/d \) for each undirected pair \( u, v \). When the first clock ticks, \( u \) pushes the rumor to \( v \) in push, if \( u \) is informed; when the second clock ticks, if in push-pull one of \( u, v \) is already informed, the other gets informed as well. Now we couple the clocks so that for each pair \( u, v \), if \( u \) gets informed before \( v \) in push, then from that point on, the ticks of clocks \( C_{u,v} \) and \( C'_{u,v} \) are perfectly synchronized. This coupling implies that the rumor spreading times of both protocols have the same distribution. Thus, if we considered the normal asynchronous push algorithm (not the sped-up one), its rumor spreading time has the same distribution as that of twice the rumor spreading time of push-pull.
graph:

\[
\text{synchronous push}^{(1)} \leq \text{asynchronous push}^{(2)} \leq \text{asynchronous push-pull}^{(3)} \leq \text{synchronous push-pull}.
\]

**Corollary 1.** Let \( G \) be a regular graph with \( n \) vertices, \( u \) a vertex of \( G \), and \( T_p^*(G) \) and \( T_{pp}^*(G) \) the smallest integers such that with probability at least \( 1 - 1/n \) the rumor to spreads from \( u \) to all other nodes within \( T_p^*(G) \) respectively \( T_{pp}^*(G) \) synchronous push respectively push-pull rounds. Then \( T_p^*(G) = \Theta(T_{pp}^*(G)) \).

## 2 Definitions

Let \( G = (V,E) \) be a connected undirected graph with \( |V| = n \) nodes. For each \( u \in V \), \( \deg_G(v) \) denotes the degree of \( v \), and \( \Gamma_G(u) \) is the set of neighbors of \( v \) in \( G \); we will often omit the subscript \( G \) when there is no ambiguity.

**Rumor Spreading Algorithms.** We consider two randomized rumor spreading algorithms on \( G \). The first one is the standard synchronous push-pull algorithm, or simply push-pull, which proceeds in synchronized rounds; we will denote by pp this algorithm. Initially, in round 0, a source node \( u \in V \) generates a rumor. In each subsequent round \( r = 1, 2, \ldots \), every node \( v \in V \) initiates a communication channel with a random neighbor \( w \in \Gamma(v) \) (we say \( v \) contacts \( w \)), and if exactly one of \( v, w \) knows the rumor (is informed) before the round, then the other node gets informed in this round as well. In particular, if \( v \) is already informed, we say that it pushes the rumor to \( w \) in the round, while if \( v \) is not informed and \( w \) is, we say that \( v \) pulls the rumor from \( w \). Note that each node contacts exactly one other node, but may be contacted by several nodes in the same round. In this case, we assume that the communications take place in parallel and independently. In the analysis, we will represent the pairwise communications that take place in a round by a set \( \{(v, w_v)\}_{v \in V} \) of \( n \) pairs, denoting that each node \( v \) contacts node \( w_v \) in the round.

The second algorithm we consider is the asynchronous push-pull algorithm, denoted pp-a. In this algorithm, each node \( v \in V \) has its own independent Poisson clock with rate \( \lambda = 1 \), and each time \( v \)'s clock ticks, \( v \) contacts a random neighbor \( w \in \Gamma(v) \). As before, if only one of \( v, w \) knows the rumor before the communication takes place, then the other node gets informed as well. We refer to this communication as a step of the algorithm, and we say that node \( v \) takes or executes this step. We will represent a step by a pair \( (v, w) \), denoting that node \( v \) contacts \( w \) in the step.

We will consider a couple of alternative, but equivalent views of asynchronous push-pull. Rather than assuming a Poisson clock with rate 1 on each node, we can assume that we have an independent Poisson clock for each (directed) pair of adjacent nodes \( (v, w) \) with rate \( 1/\deg(v) \), and each time this clock ticks, \( v \) contacts \( w \). A second alternative is to assume that we have a single Poisson clock with rate \( n \), and each time this clock ticks, a random node is chosen to perform a step, i.e., contact a random neighbor. The equivalence of these descriptions is immediate from the properties of the sum of independent Poisson random variables.

**Rumor Spreading Time.** Next we define the time complexity measure we will use. For a rumor spreading algorithm \( \alpha \), we define the rumor spreading time of \( \alpha \) on \( G = (V,E) \) for source \( u \in V \), denoted \( T(\alpha, G, u) \), to be the “time” before \( \alpha \) spreads a rumor originated at \( u \) to every node in \( G \). The notion of time is different for synchronous and asynchronous algorithms. For the former, time is measured in terms of rounds, while for asynchronous algorithms is measured in terms of
time units.\(^2\) For \(0 < q < 1\), we define \(T_q(\alpha, G, u) = \min\{t: \Pr(T(\alpha, G, u) \leq t) \geq 1 - q\}\), i.e., the time before all nodes are informed with probability \(1 - q\). We will be particularly interested in \(T_{1/n}(\alpha, G, u)\), the high-probability rumor spreading time of \(\alpha\).

Other Notation. We write \(X \sim Y\) to denote that random variables \(X\) and \(Y\) have the same distribution, and \(X \preceq Y\) to denote that \(Y\) stochastically dominates \(X\).

\(\text{Unif}(A)\) denotes the uniform distribution over set \(A\); \(\text{Geom}(p)\) is the geometric distribution with success probability \(p\); \(\text{Exp}(\lambda)\) is the exponential distribution with rate \(\lambda\); \(\text{NegBin}(k, p)\) is the negative-binomial distribution, i.e., the sum of \(k\) independent geometric random variables with probability parameter \(p\); and \(\text{Erl}(k, \lambda)\) is the Erlang distribution, i.e., the sum of \(k\) independent exponential random variables with rate \(\lambda\).

For any function \(f(x)\), we define the set \(\argmin_x f(x) = \{x: \forall y, f(x) \leq f(y)\}\). In case this set is a singleton set \(\{a\}\), we will write \(\argmin_x f(x) = a\).

3 Analysis Overview

Below we highlight the main ideas and techniques used in the analysis of our results.

Upper Bound. Our proof of Theorem 1 can be viewed as an extension of a basic coupling technique used in \([23]\) to relate the rumor spreading time of the asynchronous push algorithm (denoted push-a) with that of synchronous push. The following simple coupling was proposed there. After it gets informed, each node \(v\) contacts its neighbors in the exact same order in both algorithms, i.e., if \(v\) is the round when \(v\) gets informed in push, \(t_v\) is the time when \(v\) gets informed in push-a, and \(t_{v,i}\) is the \(i\)-th time that \(v\)’s clock ticks after \(t_v\), then \(v\) pushes the rumor to the same node in round \(r_v + i\) of push, and at time \(t_{v,i}\) in push-a. Consider now a path \(v_0 = v, v_1, \ldots, v_l = v\) through which \(v\) receives the rumor in push, where \(v_{i+1}\) learns the rumor from \(v_i\), and let \(d_i := r_{v_{i+1}} - r_{v_i}\). The time before \(v\) gets informed is then \(r_v = \sum_i d_i\). Consider also the same path in push-a and let \(\tau_i := t_{v_{i+1}} - t_{v_i}\); then \(t_v = \sum_i \tau_i\). The coupling implies that in push-a, \(v_{i+1}\) learns the rumor no later than in time \(t_{v_{i+1}}\) (when \(v_i\) pushes the rumor to it). And since in expectation, for any \(j\), \(t_{v,j} - t_v = j\), it follows that \(\text{E}[\tau_i | d_i] \leq d_i\), and thus \(\text{E}[t_v] \leq \text{E}[r_v]\). This bound can be turned into a high-probability one but we do not discuss this now.

This simple technique does not work for push-pull as there is no obvious way to couple pull operations. In fact, as far as we know, there are no coupling techniques in the rumor spreading literature that achieves such a coupling between pull operations. Our analysis does exactly that: it provides a method to couple pull operations to achieve the same effect as the above natural coupling between push operations. The coupling we propose and especially its analysis are somewhat involved, so we give the high-level ideas next.

For each node \(v\) and each neighbor \(w\) of \(v\) we define an independent exponential random variable \(Y_{v,w}\) with rate \(\lambda_v = 1/\text{deg}(v)\). In pp-a, we set equal to \(Y_{v,w}\) the time between the point \(t_w\) when \(v\) gets informed, and the point when \(v\) contacts \(w\) for the first time after \(t_w\) in order to pull the rumor, provided that \(v\) is still not informed by that time. For pp, we would like to set equal to \(Y_{v,w}\) (precisely, to \(\lfloor Y_{v,w}\rfloor\)) the number of rounds after \(r_w\), when \(v\) pulls the rumor from \(w\), provided \(v\) is still not informed by that time. This, however, has a number of issues:

\(^2\)An alternative measure for asynchronous rumor spreading would be the total number of steps before all nodes get informed. Note that the ratio of the number of steps over \(n\) is equal in expectation to the expected time units before all nodes get informed.
(1) The probability of pulling from any specific neighbor in a round is $1 - e^{-1/\deg(v)}$ which is slightly smaller than $1/\deg(v)$. We fix that by setting $\lambda_v = 2/\deg(v)$ (and using $2Y_{v,w} \sim \text{Exp}(1/\deg(v))$) rather than $Y_{v,w}$ in pp-a).

(2) Node $v$ may have to pull from more than one nodes in a round. However, this is not a real issue as it suffices to contact any one of them to get informed.

(3) The probability of a successful pull in a round is not exactly the right one: if $v$ has fewer than some constant fraction of informed neighbors the probability is larger that it should be, otherwise it is smaller. E.g., in the two extreme cases, if $v$ has only one informed neighbor the probability of pulling from it is $1 - e^{-2/\deg(v)} \approx 2/\deg(v)$ instead of $1/\deg(v)$; while if all of $v$’s neighbors are informed the probability of a successful pull is less than the correct value of 1. The former is not an actual problem, as it just speeds up pp which only makes our result stronger. The latter however is a problem. To solve it, we impose that as soon as $v$ has at least $\deg(v)/2$ informed neighbors, it pulls the rumor in the next round $r^*$ with probability 1 (if $v$ is not already informed).

This last modification requires some subtle handing in order to work: Among all its informed neighbors, we let $v$ pull the rumor from the neighbor $w^*$ that minimizes quantity $t_w + Y_{v,w}$. We show that this implies that the value of $r_w^* + Y_{v,w}^*$ is not much larger than $r^*$, in particular, $r_w^* + Y_{v,w}^* = r^* + O(1)$ in expectation. Hence, this case is not very different from the setting without the last modification, in which $v$ pulls the rumor from $w$ in a round $r$ such that $r_w + \lfloor Y_{v,w}\rfloor = r$. The intuition why we have $r_w^* + Y_{v,w}^* - r^* = O(1)$ in expectation is simple: for each of the at least $\deg(v)/2$ informed neighbors $w$ of $v$, the difference $r_w + Y_{v,w} - r^*$ is an independent exponential random variable with rate $2/\deg(v)$, and $r_w^* + Y_{v,w}^* - r^*$ is the minimum of them. This implies that $r_w^* + Y_{v,w}^* - r^*$ is exponentially distributed with a rate of at least 1.

**Lower Bound.** Our proof of Theorem 2 can be viewed as a refinement of the analysis technique used in [1] to prove a factor of $n^{-2/3}$ lower bound. However, our analysis introduces several new ideas, in order to improve that lower bound to $n^{-1/2}$.

We use a coupling argument which, roughly speaking, divides the sequence of steps in pp-a into blocks of consecutive steps. Each block is mapped to one, or more, rounds of pp in a way that the set of informed nodes in pp-a after each block is a subset of the set of informed nodes in pp after the round corresponding to that block.

We have two types of blocks, normal and special. A normal block has size at most $\sqrt{n}$ and is mapped to a single round of pp, while a special block has size exactly one and may be mapped to more than one rounds. Let us denote by $S_1 = (x_1, y_1), S_2 = (x_2, y_2), \ldots$ the sequence of steps in pp-a, where $S_i = (x_i, y_i)$ means that in step $i$ node $x_i$ contacts $y_i$. The first block into which the sequence of $S_i$ is partitioned is a normal block, and starts from $S_1$. A normal block $B$ starting from $S_i$ contains all elements up to element $S_{j-1}$, for the smallest $j$ such that one of the following three condition is met: (1) $B$ has reached its maximum size of $\sqrt{n}$; (2) the left element, $x_j$, of $S_j$ has appeared again in another pair in $B$, either as the left or as the right node, i.e., during some step in block $B$, $x_j$ has either contacted or been contacted by a node; we say that $S_j$ is left-incompatible with $B$; or (3) the right element, $y_j$, of $S_j$ got informed during a step in block $B$; we say that $S_j$ is right-incompatible with $B$.

Terminating conditions (2) and (3) ensure that each normal block can be simulated in a single round of pp. In particular, (2) ensures that a single nodes does not contact two neighbors in the same block, while (3) prevents the following bad scenario: node $v$ gets informed by some neighbor $u$, and then a non-informed neighbor $w$ of $v$ pulls the rumor from $v$ during the same block. (This scenario cannot happen in a single round.) The single round corresponding to a normal block contains all the communication pairs in that block.
If a block ends because it has reached the maximum size of $\sqrt{n}$, or a left-incompatible pair has been encountered then the next block is also normal. However, if a right-incompatible pair $S_j = (x_j, y_j)$ is encountered then the next block will be special. The reason for treating this case differently is that knowing that the next round $r$ contains an element that is right-incompatible with the previous block $B$, introduces dependencies between round $r$ and the round $r'$ that corresponds to $B$: we have that with probability 1, at least one node that was informed in round $r'$ will be contacted in the next round $r$. Note that this problem does not exist for left-incompatible pairs as each round contains one communication pair ($v, v'$) for every node $v$.

We prove that on any graph, the high-probability rumor spreading time of asynchronous push-pull is bounded by the high-probability rumor spreading time of synchronous push-pull plus, at most, a logarithmic term.

### 4 Upper Bound on Asynchronous Push-Pull

We prove that on any graph, the high-probability rumor spreading time of asynchronous push-pull is bounded by the high-probability rumor spreading time of synchronous push-pull plus, at most, some logarithmic term.

**Theorem 3.** For any connected $n$-node graph $G = (V, E)$, and any vertex $u \in V$ of $G$, we have $T_{1/n}(pp-a, G, u) = O(T_{1/n}(pp, G, u) + \log n)$.

The proof is based on a coupling argument. For the sake of comprehension, we define two auxiliary rumor spreading processes, $ppx$ and $ppy$, and present the coupling in three steps: first we couple $pp$ with $ppx$, then $ppx$ with $ppy$, and finally $ppy$ with $pp-a$.

Processes $ppx$ and $ppy$ are very similar to $pp$, except that they use different rules to decide from which neighbor a non-informed node pulls the rumor in a round. We point out that $ppx$ and $ppy$
are not realistic rumor spreading algorithms, as they assume that at any time, a node knows the set of its informed neighbors; we introduce these processes just to facilitate our analysis.

We start by describing process $ppx$.

**Definition 4 (Process $ppx$).** Process $ppx$ is a synchronous rumor spreading algorithm. For each round $r$ and $v \in V$, (1) if $v$ is informed before round $r$, then in round $r$, $v$ pushes the rumor to a random neighbor; and (2) if $v$ is not informed before round $r$, and has $k$ informed neighbors at that time, then with probability

$$p = \begin{cases} 
1 - e^{-2k/\deg(v)}, & \text{if } k < \deg(v)/2; \\
1, & \text{if } k \geq \deg(v)/2,
\end{cases}$$

$v$ pulls the rumor from a random informed neighbor in round $r$, while with the remaining probability, $1 - p$, $v$ does not pull the rumor in this round.

We show next that the rumor spreading time for $ppx$ is dominated by that for $pp$. The proof (in Appendix A.1) is straightforward and relies on the observation that, for the same set of informed nodes, a non-informed node is more likely to pull the rumor in $ppx$ than in $pp$, in the next round.

**Lemma 5.** $T(ppx, G, u) \leq T(pp, G, u)$.

The second auxiliary process we introduce, $ppy$, is identical to $ppx$ except for the probability with which a node pulls the rumor in a round: we no longer distinguish between the cases $k < \deg(v)/2$ and $k \geq \deg(v)/2$, and the first formula is used in both cases.

**Definition 6 (Process $ppy$).** Process $ppy$ is a synchronous rumor spreading algorithm. For each round $r$ and $v \in V$, (1) if $v$ is informed before round $r$, then in round $r$, $v$ pushes the rumor to a random neighbor; and (2) if $v$ is not informed before round $r$, and has $k$ informed neighbors at that time, then with probability

$$p = 1 - e^{-2k/\deg(v)},$$

$v$ pulls the rumor from a random informed neighbor in round $r$, while with the remaining probability, $1 - p$, $v$ does not pull the rumor in this round.

In Lemma 8 below, we bound the rumor spreading time for $ppx$ in terms of the rumor spreading time for $ppy$. First we provide a technical lemma that we will need. This lemma computes the conditional distribution of the minimum of a collection of independent exponential random variables, given some limited information about them. The proof can be found in Appendix A.2.

**Lemma 7.** Let $Z_1, \ldots, Z_k$ be independent identical random variables with $Z_i \sim \text{Exp}(\lambda)$, and let $J = \text{argmin}_i Z_i$. For $\alpha_1, \ldots, \alpha_k$ arbitrary non-negative integers, let $Z = \text{min}_i \{Z_i - \alpha_i\}$, and let $A$ be the event: for all $i$, $Z_i > \alpha_i$. Then $(Z \mid J = j, A) \sim \text{Exp}(k\lambda)$, i.e., for any $t \geq 0$, we have that $\Pr[Z \leq t \mid J = j, A] = 1 - e^{-kt\lambda}$.

We remark that from the memoryless property of exponential distribution it is immediate that $(Z \mid A) \sim \text{Exp}(k\lambda)$. The lemma says that conditioning also on $J$ does not add any information.

We now proceed to the main lemma.

**Lemma 8.** $T_\delta(ppy, G, u) = O(T_\delta(ppx, G, u) + \log(n/\delta))$, for any $0 < \delta \leq 1/2$. 
Proof. We define a coupling of the random processes underlying $ppx$ and $ppy$. For each $v \in V$, let $r_v$ and $r'_v$ denote the rounds in which $v$ gets informed in $ppx$ and $ppy$, respectively (for the source $u$, $r_u = r'_u = 0$). We will show that for any fixed $v \in V$, with probability at least $1 - \delta/2n$ we have $r'_v = O(r_v + \log(n/\delta))$. A union bound then completes the proof.

To facilitate the coupling we introduce the following collection of random variables. For each $v \in V$, let $X_{v,i}$ and $Y_{v,w}$ be random variables with $X_{v,i} \sim \text{Unif}(\Gamma(v))$ and $Y_{v,w} \sim \text{Exp}(\lambda_v)$, with $\lambda_v = 2/\deg(v)$ (i.e., $X_{v,i}$ is a random neighbor of $v$, and $Y_{v,w}$ is an exponential random variable with rate $\lambda_v$). We assume all these random variables to be mutually independent.

For push operations, the coupling states that each node pushes the rumor to the same neighbor in both processes, in the $i$-th round after the node gets informed. Formally, for each $v \in V$ and $i \geq 1$, $v$ pushes the rumor to node $X_{v,i}$ in round $r_v + i$ of $ppx$, as well as in round $r'_v + i$ of $ppy$.

For pull operations, the coupling is more involved. For $ppy$, for each pair of adjacent nodes $v,w$, if $w$ gets informed before $v$ (i.e., $r'_w < r'_v$), and $v$ is still not informed before round $r'_w + [Y_{v,w}]$, then we let $v$ pull the rumor from $w$ in that round. The formal definition takes also into account the possibility that $r'_w + [Y_{v,w}] = r'_x + [Y_{v,x}]$ for two distinct neighbors $w, x$ of $u$, in which case the tie is broken using the actual values of $Y_{v,w}$ and $Y_{v,x}$ (rather than their rounded up values). Precisely, for any $v \in V \setminus \{u\}$, if $v$ does not get informed by a push operation before round $t = \min_{w \in \Gamma(v)} \{r'_w + [Y_{v,w}]\}$, then in round $t$, $v$ pulls the rumor from node $\arg\min_{w \in \Gamma(v)} \{r'_w + Y_{v,w}\}$, i.e., the neighbor $w$ that minimizes $r'_w + Y_{v,w}$. Clearly, for this neighbor $w$, $r'_w + [Y_{v,w}] = t$.

For $ppx$, we use a similar coupling rule except for we need to enforce that, as soon as half of $v$’s neighbors get informed, $v$ will pull the rumor in the next round with probability 1, if it is not already informed. The neighbor $w$ from which $v$ pulls the rumor in this case is the (currently informed) neighbor that minimizes $r_w + Y_{v,w}$. Precisely, for any $v \in V \setminus \{u\}$, if $t = \min_{w \in \Gamma(v)} \{r'_w + [Y_{v,w}]\}$ and $z$ is the first round by the end of which at least $\deg(v)/2$ of $v$’s neighbors have been informed, we distinguish two cases:

(i) If $t \leq z$ and $v$ does not get informed by a push operation before round $t$, then in round $t$, $v$ pulls the rumor from node $\arg\min_{w \in \Gamma(v)} \{r_w + Y_{v,w}\}$. So, this case is completely analogous to the rule for $ppy$.

(ii) If $t > z$ and $v$ does not get informed by a push operation before round $z_v + 1$, then in round $z_v + 1$, $v$ pulls the rumor from node $\arg\min_{w \in \Gamma(v)}: r_w \leq z_v \{r_w + Y_{v,w}\}$.

It is not hard to check that the above coupling is valid, in the sense that the marginal distributions of the two processes are the correct ones: For push operations there is nothing to argue about, so we focus on pull operations. In $ppy$, if before round $r$ node $v$ is still not informed, and its set of informed neighbors is $S$ with $|S| = k$, then the probability that $v$ pulls the rumor in round $r$ is the same as the conditional probability that $\min_{w \in S} \{r'_w + Y_{v,w}\} \leq r$, given that $\min_{w \in S} \{r'_w + Y_{v,w}\} > r - 1$. Since $Y_{v,w} \sim \text{Exp}(\lambda_v)$, it follows from the memoryless property of the exponential distribution that the above conditional probability is

$$1 - (1 - \Pr[r'_w + Y_{v,w} > r \mid r'_w + Y_{v,w} > r - 1])^k = 1 - e^{-k\lambda_v} = 1 - e^{-2k/\deg(v)},$$

which is the right probability according to Definition 6. Moreover, if $v$ does pull the rumor in round $r$, then it is equally likely to pull it from any of its informed neighbors in $S$, as the conditional random variables $(r'_w + Y_{v,w} \mid r'_w + Y_{v,w} > r - 1)$, for $w \in S$, have the same distribution and are independent. A very similar argument shows that our coupling yields the correct distribution for the pull operations in $ppx$, as well.

\footnote{Since $r'_w + Y_{v,w}$ is a continuous random variable, the probability this quantity is the same for two distinct $w$ is 0.}
Next we show that with probability at least $1 - \delta/2n$, $r'_v = O(r_v + \log(n/\delta))$.

Let $\pi_v$ be a path through which the rumor reaches node $v \in V \setminus \{u\}$ in $ppx$. Formally, $\pi_v = v_0 v_1 \ldots v_l$, where $v_0 = u$, $v_l = v$, and for each $0 \leq i < l$, $v_i$ is a node from which $v_{i+1}$ receives the rumor for the first time (i.e., in round $r_{v_{i+1}}$). We express $r_v$ and $r'_v$ as

$$r_v = \sum_{0 \leq i < l} (r_{v_{i+1}} - r_{v_i}), \quad r'_v = \sum_{0 \leq i < l} (r'_{v_{i+1}} - r'_{v_i}).$$

(Note that these equations hold for any collection of $v_i$, not just for the specific definition of $\pi_v$.)

For each $0 \leq i < l$, we compare the random variables $d_i := r_{v_{i+1}} - r_{v_i}$ and $d'_i := r'_{v_{i+1}} - r'_{v_i}$, and show that $d'_i - d_i$ is dominated by a geometric random variable with constant expectation. We distinguish three cases, depending on how $v_{i+1}$ gets the rumor from $v_i$ in $ppx$.

**Case 1:** $v_i$ pushes the rumor to $v_{i+1}$, in round $r_{v_{i+1}}$ of $ppx$. In this case, we have $r_{v_{i+1}} = r_{v_i} + \min\{j : X_{v_i,j} = v_{i+1}\}$, and thus $d_i = \min\{j : X_{v_i,j} = v_{i+1}\}$. Similarly, in $ppx$, $v_i$ pushes the rumor to $v_{i+1}$ in round $r'_{v_{i+1}} + \min\{j : X_{v_i,j} = v_{i+1}\}$. Thus $v_{i+1}$ gets informed in $ppx$ no later than in this round, i.e., $r'_{v_{i+1}} \leq r'_{v_{i+1}} + \min\{j : X_{v_i,j} = v_{i+1}\}$, and so, $d'_i \leq \min\{j : X_{v_i,j} = v_{i+1}\} = d_i$.

**Case 2:** $v_{i+1}$ pulls the rumor from $v_i$, in round $r_{v_{i+1}}$ of $ppx$, and before that round fewer than half of $v_i$'s neighbors are informed. In this case, $r_{v_{i+1}} = r_{v_i} + Y_{v_{i+1},v_i}$, thus $d_i = Y_{v_{i+1},v_i}$. Similarly, in $ppx$, $v_{i+1}$ gets informed no later than in round $r'_{v_{i+1}} + Y_{v_{i+1},v_i}$, because if $v_{i+1}$ is still not informed before that round, it will pull the rumor in round $r'_{v_{i+1}} + Y_{v_{i+1},v_i}$ (from $v_i$ or some other informed neighbor). It follows that $d'_i \leq [Y_{v_{i+1},v_i}] = d_i$.

**Case 3:** $v_{i+1}$ pulls the rumor from $v_i$, in round $r'_{v_{i+1}}$ of $ppx$, and before that round at least half of $v_i$'s neighbors are informed. This case is more involved. As in case 2, we have $d'_i \leq [Y_{v_{i+1},v_i}]$, but now it is possible that $d_i < [Y_{v_{i+1},v_i}]$. In this case we will use Lemma 7 to bound $d'_i - d_i$. Let $z$ be the first round in $ppx$ after which at least half of the neighbors of $v_{i+1}$ are informed, and let $S$ be the set of those informed neighbors (so $|S| \geq \deg(v_{i+1})/2$). Then $r_{v_{i+1}} = z + 1$. Also, from the coupling (case (ii)), $v_i = \arg\min_{w \in S} \{r_w + Y_{v_{i+1},v_i}\}$, thus $r_{v_i} + Y_{v_{i+1},v_i} < r_w + Y_{v_{i+1},v_i}$ for all $w \in S \setminus \{v_i\}$. Note, it is possible that $r_{v_i} + Y_{v_{i+1},v_i} > z + 1$ (implying $d_i < Y_{v_{i+1},v_i}$).

In preparation for applying Lemma 7, let us fix the random choices in all rounds of $ppx$, and for each $w \in S$, let $Z_w = r_w + Y_{v_{i+1},v_i} - z$. The set $\{Z_w\}_{w \in S}$ is then a collection of independent random variables with distribution $\text{Exp}(\lambda_{v_{i+1}})$ each, and $\arg\min_{w \in S} \{Z_w\} = v_i$.

Consider now process $ppx$. To simplify exposition we shift all round numbers in $ppx$ by an appropriate offset, so that the round in which $v_i$ gets informed has the same number in both processes. E.g., if $v$ gets informed after $k$ rounds in $ppx$ and after $\ell$ rounds in $ppx$, we add an offset of $(k - \ell)$ to all round numbers in $ppx$. So, the $i$-th round in $ppx$ has number $i + (k - \ell)$, and the round when $v$ gets informed is $r'_{v_i} = k = r_{v_i}$.

If it is $r'_{v_{i+1}} \leq z$, then $d'_i = r'_{v_{i+1}} - r'_{v_i} = r_{v_{i+1}} - r_{v_i} \leq z - r_{v_i} = d_i - 1$, thus $d'_i < d_i$.

In the following we assume that $r'_{v_{i+1}} > z$, and bound the difference $r'_{v_{i+1}} - z$ using Lemma 7.

Let us fix the random choices in the first $z$ rounds of $ppx$, in a way that respects our coupling (recall we have already fixed the random choices in all rounds of $ppx$). Revealing these random choices in $ppx$ discloses additional information about variables $Y_{v_{i+1},w}$, and thus about $Z_w$. Precisely, the additional information is that for each $w \in S$, $r'_w + Y_{v_{i+1},v_i} - z > 0$, which follows from assumption $r'_{v_{i+1}} > z$. This implies $Z_w > r_w - r'_w$, and since we also have $Z_w > 0$ (from $ppx$), it follows that for any $w \in S$, $Z_w > \alpha_w$, where $\alpha_w = \max\{0, r_w - r'_w\}$. Letting $Z = \min_{w \in S} \{Z_w - \alpha_w\}$, and applying Lemma 7, gives for $t \geq 0$,

$$\Pr[Z \leq t] = 1 - e^{-|S|\lambda_{v_{i+1}}t} \geq 1 - e^{-t},$$

as $|S| \geq \deg(v_{i+1})/2$ and $\lambda_{v_{i+1}} = 2/\deg(v_{i+1})$. 

9
The random variable \( Z \) is closely related to \( r'_{v_i+1} \):

\[
Z = \min_{w \in S} \{ Z_w - \alpha_w \} = \min_{w \in S} \{(r_w + Y_{v_i+1,w} - z) - (\max\{0, r_w - r'_w\})\}
\]

\[
= \min_{w \in S} \{ \min\{r_w, r'_w\} + Y_{v_i+1,w} \} - z
\]

\[
= \min_{w \in S} \{ r'_w + Y_{v_i+1,w} \} - z,
\]

where the last equation holds because for any \( w \in S \setminus \{v_i\}, r_w + Y_{v_i+1,w} > r_{v_i} + Y_{v_i+1,v_i} \) (since \( v_i = \arg\min_{w \in S} \{ r_w + Y_{v_i+1,w} \} \)), thus if \( r_w < r'_w \), the term \( \min\{r_w, r'_w\} + Y_{v_i+1,w} \) does not affect the value of \( \min \) in the second-to-last line above, and we may as well replace it by the larger term \( r'_w + Y_{v_i+1,w} \).

Now since \( \min_{w \in S} \{ r'_w + [Y_{v_i+1,w}] \} = \min_{w \in \Gamma(v_{i+1})} \{ r'_w + [Y_{v_i+1,w}] \} = r'_{v_i+1} \), the equation we showed above yields \( Z \geq r'_{v_i+1} - z \). From this and the bound \( \Pr[Z \leq t] \geq 1 - e^{-t} \) shown earlier, it follows that for any integer \( t \), \( \Pr[r'_{v_i+1} - z \leq t] \geq 1 - e^{-t} \).

Last, substituting \( d'_i = r'_{v_i+1} - r'_{v_i} = r'_{v_i+1} - r_{v_i} \) and \( d_i = (z + 1) - r_{v_i} \), yields \( \Pr[d'_i - d_i + 1 \leq t] \geq 1 - e^{-t} \). This completes case 3.

In each of the above cases 1–3, we have that either (1) \( d'_i \leq d_i \), or (2) conditionally on the random choices in all rounds of \( \text{ppx} \) and the first \( r'_{v_i} + d_i - 1 \) rounds of \( \text{ppy} \), we have \( d'_i - d_i + 1 \leq t \) with probability at least \( 1 - e^{-t} \), for any integer \( t \geq 0 \). Note that (1) implies (2).

Let us now fix all random choices in \( \text{ppx} \) (and thus all \( d_i \)). Then from (2) it follows that \( \Pr[d'_i - d_i + 1 \leq t | d'_1 \ldots d'_{i-1}] \geq 1 - e^{-t} \). This says that the random variables \( Z_i := d'_i - d_i + 1 \) are dominated by Geom\((1 - e^{-t})\), independently of \( Z_1 \ldots Z_{i-1} \). Using that \( \sum_{0 \leq i \leq l} d_i \) and \( r'_v = \sum_{0 \leq i \leq l} d_i \), we obtain \( r'_v - r_v + l = \sum_{0 \leq i \leq l} (d'_i - d_i + 1) \), and applying then Lemma 10 (in the appendix), for the sum of the (dependent) random variables \( Z_i, 0 \leq i < l \), we obtain that \( r'_v - r_v + l \leq \text{NegBin}(l, 1 - 1/e) \). From this, it follows that with probability at least \( 1 - \delta/2n \), \( r'_v \leq r_v + O(l + \log(n/\delta)) \), and thus \( r'_v = O(r_v + \log(n/\delta)) \), as \( l \leq r_v \).

Taking the union bound over all \( v \), gives that with probability at least \( 1 - \delta/2 \), we have for all \( v \) that \( r'_v = O(r_v + \log(n/\delta)) \). And since by definition, with probability \( 1 - \delta/2 \), for all \( v \) we have \( r_v \leq T_{\delta/2}(\text{ppx}, G, u) \), another union bound gives that with probability at least \( 1 - \delta \), we have for all \( v \) that \( r'_v = O(T_{\delta/2}(\text{ppx}, G, u) + \log(n/\delta)) \). This means \( T_{\delta}(\text{ppy}, G, u) = O(T_{\delta/2}(\text{ppx}, G, u) + \log(n/\delta)) \). Finally, observing that \( T_{\delta/2}(\text{ppx}, G, u) \leq 2T_{\delta}(\text{ppx}, G, u) \) for \( \delta \leq 1/2 \), concludes the proof of Lemma 8.

It remains to couple process \( \text{ppy} \) with the asynchronous rumor spreading algorithm \( \text{pp-a} \).

**Lemma 9.** \( T_{\delta}(\text{pp-a}, G, u) \leq O(T_{\delta}(\text{ppy}, G, u) + \log(n/\delta)) \), for any \( 0 < \delta \leq 1/2 \).

The proof, which can be found in Appendix A.4, has a similar structure as that for Lemma 8. However, the coupling is more straightforward and easier to analyse in this case. Essentially, it captures the intuition that \( \text{ppy} \) is just a discretized version of \( \text{pp-a} \).

**Proof of Theorem 3.** Combining Lemma 9, Lemma 8, and Lemma 5, we obtain that for any \( 0 < \delta \leq 1/2, T_{\delta}(\text{pp-a}, G, u) = O(T_{\delta}(\text{ppy}, G, u) + \log(n/\delta)) = O(T_{\delta}(\text{ppx}, G, u) + \log(n/\delta)) = O(T_{\delta}(\text{pp}, G, u) + \log(n/\delta)) \). Setting \( \delta = 1/n \), yields the statement of the theorem.
References

rumour spreading. In Proceedings of the 34th ACM Symposium on Principles of Distributed


actions on Information Theory and IEEE/ACM Transactions on Networking, 52:2508–2530,
2006.


pp. 399–408. 2010.

and D. Terry. Epidemic algorithms for replicated database maintenance. In Proceedings of the

and D. Terry. Epidemic algorithms for replicated database maintenance. In Proceedings of the


315. 2012.


Cayley graphs. In Proceedings of the 24th International Symposium on Theoretical Aspects of


APPENDIX

A Omitted Proofs

A.1 Proof of Lemma 5

We define a simple coupling of the random choices of the two algorithms in a round. Let $S_r$ and $S'_r$ denote the set of informed nodes after the first $r$ rounds of $pp$ and $ppx$, respectively. Let us fix $r$ and sets $S_r, S'_r$, and suppose that $S_r \subseteq S'_r$. We will describe a coupling of round $r + 1$ of the two algorithms, which guarantees that $S_{r+1} \subseteq S'_{r+1}$. Applying this coupling to all rounds (assuming the same source $u$ for both algorithm), we obtain that $S_r \subseteq S'_r$ for all $r \geq 0$. The claim then follows.

The coupling is based on the following observation. For any $v \in V \setminus S'_r$, the probability that $v$ pulls the rumor in round $r + 1$ is at least as large for $ppx$ as it is for $pp$: Let $p_v$ and $p'_v$ be these probabilities for $pp$ and $ppx$, respectively, and let $k$ and $k'$ be the number of $v$’s neighbors that belong to $S_r$ and $S'_r$, respectively. Since $S_r \subseteq S'_r$, it follows that $k \leq k'$. We have two cases:

If $k' < \deg(v)/2$, then

$$p'_v = 1 - e^{-2k'/\deg(v)} \geq 1 - (1 - k'/\deg(v)) \geq k/\deg(v) = p_v,$$

where the first inequality is obtained using the fact that $e^{-x} \leq 1 - x/2$ for $x \leq 1$, and that $2k'/\deg(v) < 1$; and the second inequality above holds because $k \leq k'$.

If $k' \geq \deg(v)/2$, then $p'_v = 1$ and thus clearly $p'_v \geq p_v$.

Therefore in both cases above, it holds $p'_v \geq p_v$.

The coupling between the random choices of the two algorithms in round $i + 1$ is as follows: (1) each node $v \in S_r$ pushes the rumor to the same node in both algorithms; and (2) for each node $v \in V \setminus S'_r$, we couple the choice of $v$ in $pp$ and its choice in $ppx$ in a way that $v$ pulls the rumor in $pp$ (from some informed neighbor) only if $v$ pulls the rumor in $ppx$ (maybe from a different informed neighbor); this coupling is possible because $p'_v \geq p_v$, as we argued earlier. Note that we do not need to couple the choices in the two algorithms of a node $v \in S'_r \setminus S'_r$: $v$ is already informed in $ppx$ before round $r + 1$, while in $ppx$ it is not informed and thus cannot inform any other node in round $r + 1$.

From the above coupling it is immediate that $S_{r+1} \subseteq S'_{r+1}$: For each $v \in V \setminus S'_r$, if $v$ gets informed in $pp$ by a push operation (performed by some $v \in S_r$), then the same happens in $ppx$, because of rule (1) of the coupling; while if $v$ gets informed in $pp$ by pulling the rumor then the same happens in $ppx$, because of rule (2).

A.2 Proof of Lemma 7

By Bayes’ rule

$$\Pr[Z > t \mid J = j, A] = \frac{\Pr[J = j \mid Z > t, A]}{\Pr[J = j \mid A]} \cdot \Pr[Z > t \mid A].$$

We will compute the three probabilities on the right-hand side.

Conditionally on $A$, the random variables $Z'_i = Z_i - \alpha_i$ have distribution $\text{Exp}(\lambda)$ (by the memoryless property of the exponential distribution) and are independent, thus the conditional distribution of their minimum, $Z$, is $\text{Exp}(k\lambda)$. It follows that $\Pr[Z > t \mid A] = e^{-k\lambda t}$.

By the memoryless property, $\Pr[Z_i \leq z \mid Z_i > \alpha_i] = \Pr[Z_i \leq z + t \mid Z_i > \alpha_i + t]$. Using this
equality we obtain
\[
\Pr[J = j \mid A] = \Pr[\arg\min_i Z_i = j \mid \{Z_i > \alpha_1\}_i]
\]
\[
= \Pr[\arg\min_i (Z_i - t) = j \mid \{Z_i > \alpha_1 + t\}_i] \quad \text{(by the equation above)}
\]
\[
= \Pr[\arg\min_i Z_i = j \mid \{Z_i > \alpha_1 + t\}_i]
\]
\[
= \Pr[J = j \mid Z > t, A].
\]
Thus, \[\frac{\Pr[J = j | Z > t, A]}{\Pr[J = j | A]} = 1.\]
Combining all the above yields \[\Pr[Z > t | J = j, A] = e^{-k\lambda t}.\]

\section{A Domination Lemma}

The next lemma bounds a sum of random variables in which, roughly speaking, each random variable has a geometric distribution conditionally on all previous variables.

\begin{lemma}
Let \(Z_1, \ldots, Z_k\) be random variables such that for each \(1 < i \leq k\) and any \(j \geq 0\), \(\Pr[Z_i \leq j \mid Z_1 \ldots Z_{i-1}] \geq 1 - q^j\). Then \(\sum Z_i \sim \text{NegBin}(k, 1 - q)\).
\end{lemma}

\begin{proof}
A standard coupling argument shows that \(\sum Z_i\) is dominated by the sum of \(k\) independent random variables \(Z_1', \ldots, Z_k'\) with \(Z_i' \sim \text{Geom}(1 - q)\). The claim then follows because \(\sum Z_i' \sim \text{NegBin}(k, 1 - q)\).
\end{proof}

\section{Proof of Lemma 9}

The structure of the proof is similar to that for Lemma 8. The coupling is similar as well, even though the justification is simpler in this case.

For each \(v \in V\), let \(t_v\) be the \textit{time} in which \(v\) gets informed in \(pp-a\), and let \(r_v\) be the \textit{round} in which \(v\) gets informed in \(ppy\). We will show that for any \(v \in V\), with probability at least \(1 - \delta/2n\), \(t_v = O(r_v + \log(n/\delta))\). Then a union bound completes the proof.

For the coupling between \(pp-a\) and \(ppy\) we use the same random variables as in the coupling between \(pp-a\) and \(ppx\), in the proof of Lemma 8: For any \(v \in V\), \(i \geq 1\), and \(w \in \Gamma(v)\), let \(X_{v,i}\) and \(Y_{v,w}\) be independent random variables with \(X_{v,i} \sim \text{Unif}(\Gamma(v))\) and \(Y_{v,w} \sim \text{Exp}(\lambda_v)\), where \(\lambda_v = 2/\text{deg}(v)\). For \(ppy\), we have the same coupling rules as in the proof of Lemma 8, while for \(pp-a\) we use a continuous-time variant of those rules.

Precisely, for \textit{push} operations the coupling is as follows. For each \(v \in V\) and \(i \geq 1\), \(v\) pushes the rumor to node \(X_{v,i}\) in round \(r_v + i\) of \(ppy\), and similarly \(v\) pushes the rumor to \(X_{v,i}\) at time \(r_{v,i}\) in \(pp-a\), where \(r_{v,i}\) denotes the time at which \(v\) does its \(i\)-th step after time \(t_v\) (i.e., its \(i\)-th push operation after it gets informed).

For \textit{pull} operations, the coupling is as follows. In \(ppy\), for each \(v \in V \setminus \{u\}\), if \(v\) does not get informed by a push operation before round \(r = \min_{w \in \Gamma(v)} \{r_w + |Y_{v,w}|\}\), then in round \(r\), \(v\) pulls the rumor from node \(\arg\min_{w \in \Gamma(v)} \{Y_{v,w}\}\). In \(pp-a\), for any \(v \in V \setminus \{u\}\), if \(v\) does not get informed by a push operation before time \(t = \min_{w \in \Gamma(v)} \{t_w + 2Y_{v,w}\}\), then at time \(t\), \(v\) pulls the rumor from node \(\arg\min_{w \in \Gamma(v)} \{t_w + 2Y_{v,w}\}\) (the factor of 2 is explained next).

The above coupling yields the correct marginal distribution for the two processes: For \(ppy\) the same argument applies as in the proof of Lemma 8. For \(pp-a\) we just need to argue about pull operations. For that, it is convenient to consider the view of \(pp-a\) in which for each (directed) pair \(v, w\) of adjacent nodes, there is an independent poisson clock \(C_{v,w}\) with rate \(1/\text{deg}(v)\), and for each
time \( t \) at which the clock ticks, if right before \( t \), \( v \) is still not informed and \( w \) is informed, then \( v \), pulls the rumor from \( w \) at time \( t \). Our coupling sets the length \( L_w \) of the interval between time \( t_w \) and the next tick of clock \( C_{v,w} \) to be \( L_w = 2Y_{v,w} \). Since \( Y_{v,w} \sim \text{Exp}(2/\deg(v)) \) it follows that \( 2Y_{v,w} \sim \text{Exp}(1/\deg(v)) \), and this is the correct distribution for \( L_w \).

Next we show that with probability at least \( 1 - \delta/2n \), \( t_v = O(r_v + \log(n/\delta)) \).

Let \( \pi_v \) be a path through which the rumor reaches node \( v \in V \setminus \{u\} \) in \( ppy \). Formally, \( \pi_v = v_0 v_1 \ldots v_l \), where \( v_0 = u \), \( v_l = v \), and for each \( 0 \leq i < l \), \( v_i \) is a node from which \( v_{i+1} \) receives the rumor for the first time (i.e., in round \( r_{v_{i+1}} \)). Following the proof of Lemma 8, we compare the random variables \( d_i := r_{v_{i+1}} - r_v \) and \( \tau_i := t_{v_{i+1}} - t_v \). We have two cases:

**Case 1:** \( v_{i+1} \) pulls the rumor from \( v_i \) in round \( r_{v_{i+1}} \) of \( ppy \). In this case, \( r_{v_{i+1}} = r_v + [Y_{v_{i+1}, v_i}] \), thus \( d_i = [Y_{v_{i+1}, v_i}] \). Similarly, in \( pp-a \), \( v_{i+1} \) gets informed no later than in time \( t_v + 2Y_{v_{i+1}, v_i} \), because if \( v_{i+1} \) is still not informed by that time, it will pull the rumor from \( v_i \). It follows that

\[
\tau_i = t_{v_{i+1}} - t_v \leq 2Y_{v_{i+1}, v_i} \leq 2d_i.
\]

**Case 2:** \( v_i \) pushes the rumor to \( v_{i+1} \) in round \( r_{v_{i+1}} \) of \( ppy \). In this case, \( r_{v_{i+1}} = r_v + x \), where \( x = \min\{j : X_{u,j} = v_{i+1}\} \), thus \( d_i = x \). Similarly in \( pp-a \), \( v_i \) pushes the rumor to \( v_{i+1} \) at time \( t_{v_i,x} \) (i.e., in its \( x \)-th step after it gets informed). Thus, \( t_{v_{i+1}} \leq t_{v_i,x} \), and \( \tau_i \leq t_{v_i,x} - t_v \). Given all random choices in \( ppy \), and the random choices in \( pp-a \) up to time \( t_{v_i,x} \), we have \( t_{v_i,x} - t_v \sim \text{Erl}(x, 1) \), and thus \( \tau_i \leq \text{Erl}(x, 1) \).

Therefore, in both of the above two cases, either (1) \( \tau_i \leq 2d_i \), or (2) conditionally on all random choices in \( ppy \) and the random choices in \( pp-a \) up to time \( t_{v_i} \), we have \( \tau_i \leq \text{Erl}(d_i, 1) \). Hence in either case it holds \( \tau_i - 2d_i \leq \text{Erl}(d_i, 1) \).

Let us now fix all random choices in \( ppy \) (and thus all \( d_i \)). From the above it follows that

\[
\sum_i \tau_i - \sum_i 2d_i \leq \text{Erl}(d_i, 1) \sim \text{Erl}(\sum_i d_i, 1),
\]

and substituting \( \sum_i \tau_i = t_v \) and \( \sum_i d_i = r_v \), yields

\[
t_v - 2r_v \leq \text{Erl}(r_v, 1).
\]

Using the fact that \( \text{Erl}(k, \lambda) \leq \text{NegBin}(k, 1 - e^{-\lambda}) \), we obtain \( t_v - 2r_v \leq \text{NegBin}(r_v, 1 - 1/e) \). From this, it follows that \( t_v = O(r_v + \log(n/\delta)) \), with probability at least \( 1 - \delta/2n \). Using now the union bound exactly as in the end of the proof of Lemma 8, we complete the proof of Lemma 9.

\[\square\]

## B Lower Bound on Asynchronous Push-Pull

We prove that the expected rumor spreading time of asynchronous push-pull is at least \( O(1/\sqrt{n}) \) times the expected rumor spreading time of its synchronous counterpart.

**Theorem 11.** For any connected \( n \)-node graph \( G = (V, E) \), and any vertex \( u \in V \) of \( G \), we have

\[
E[T(pp-a, G, u)] = \Omega\left((1/\sqrt{n}) \cdot E[T(pp, G, u)]\right).
\]

The proof is based on a coupling argument which, roughly speaking, maps \( \sqrt{n} \) steps of \( pp-a \) to one round of \( pp \), except for some “special” steps which are mapped to possibly several rounds. However the total number of rounds mapped to those special steps is shown to be bounded by \( O(\sqrt{n}) \) in expectation. We start by describing this coupling between \( pp-a \) and \( pp \).

**Coupling Construction.** We will use the following simple notion. By \( S \) we will denote a random variable that is a pair \((x, y)\) with \( x \sim \text{Unif}(V) \) and \((y | x) \sim \text{Unif}(\Gamma(x))\), i.e., \( x \) is a random node and \( y \) a random neighbor of \( x \). By \( R \) we will denote a random variable that is an \( n \)-set of pairs \( \{(v, z_v)\}_{v \in V} \), with one pair per node, such that \( z_v \sim \text{Unif}(\Gamma(v)) \) and the \( z_v \), \( v \in V \), are independent.

\[\text{Recall, } \text{Erl}(x, \ell) \text{ is the sum of } k \text{ independent exponential random variables with rate } \ell.\]
For the coupled copies of the two processes we will use the following notation. For \( t \geq 1 \), let \( a_t \) denote the node that takes the \( t \)-th step in \( pp-a \), and let \( b_t \) be the neighbor that \( a_t \) contacts in that step. For \( r \geq 1 \) and each \( v \in V \), let \( c_{r,v} \) denote the neighbor that node \( v \) contacts in round \( r \) of \( pp \). Note, it must be \( (a_t, b_t) \sim S \) and \( \{(v, c_{r,v})\}_{v \in V} \sim R \). We define also the following random variables, which will facilitate the coupling. Let \( S_i \) and \( R_{i,j} \), for \( i, j \geq 1 \), be a collection of independent random variables such that \( S_i = (x_i, y_i) \sim S \), and \( R_{i,j} = \{(v, z_{i,j,v})\}_{v \in V} \sim R \).

Our coupling partitions the sequence \( S_1, S_2, \ldots \) into blocks of consecutive sequence elements, such that, roughly speaking, the result on rumor spreading of the pairwise communications described by the block’s elements is the same, regardless of whether communications take place sequentially or in parallel. To each block will correspond to a sequence of one or more steps of \( pp-a \), and a sequence of one or more rounds of \( pp \). We distinguish two types of blocks: normal and special blocks. A normal block has size at most \( \sqrt{n} \). The sequence of \( S_i \) in a normal block describes the corresponding sequence of steps \( (a_t, b_t) \) in \( pp-a \), and also (a subset of) the communications \( \{(v, c_{r,v})\}_{v \in V} \) in a single round \( r \) of \( pp \).

Each special block is preceded by a normal block, and has size exactly 1. Its single element satisfies certain conditions with respect to the elements of the preceding normal block: Roughly speaking, if \( S_i = (x_i, y_i) \) is the element of the special block, then \( y_i \) must have got informed in a step of \( pp-a \) corresponding to the previous block. The special block corresponds to a single step in \( pp-a \), though not the step described by pair \( S_i \) (as explained later). Further, the block corresponds to one or more rounds of \( pp \), and the communication in these rounds is described by the sequence \( R_{i,1}, R_{i,2}, \ldots \) of random variables. Roughly speaking, the number of rounds we have for this special the block equals the minimum \( j \) such that \( R_{i,j} \) contains a pair \( (v, z_{i,j,v}) \) satisfying the same condition as \( S_i \) with respect to the preceding normal block; this pair is used as the \( i \)-th step \( (a_t, b_t) \) of \( pp-a \). Special blocks are necessary in our coupling to ensure the correct marginal distribution for the coupled copy of the \( pp \) process.

We provide now the details of the coupling. We will use the following terminology. Let \( H = \{(u_i, v_i)\} \) be a set of pairs of adjacent nodes, and let \( I \subseteq V \). We say that a pair \((x, y)\) of adjacent nodes is left-incompatible with \( H \), if \( x \in \{u_i\} \cup \{v_i\} \); and \((x, y)\) is right-incompatible with \( H \) and \( I \), if (1) \((x, y)\) is not left-incompatible with \( H \), and (2) in a run of \( pp-a \) in which the set \( I \) of nodes know the rumor initially, and the pair of nodes \( u_i, v_i \) communicate in the \( i \)-th step, we have that node \( y \) gets informed in one of these steps.\(^5\) We say that \( H \) is incompatible-free with respect to \( I \subseteq V \) if there is no pair \((u, v) \in H \) that is left-incompatible with the remaining set \( H' = H \setminus \{(u, v)\} \), or right-incompatible with \( H' \) and \( I \).

**Remark 12.** If \( H \) is incompatible-free with respect to \( I \), then after applying the pairwise communications described in \( H \) starting from a set \( I \) of informed nodes, the resulting set of informed nodes is the same whether these communications take place sequentially (as in \( pp-a \)), or in parallel (as in a single round of \( pp \)).

As mentioned earlier, we partition the sequence of \( S_i \) into blocks of consecutive elements, and distinguish between normal and special blocks. The first block, which starts with element \( S_1 \), is normal. A normal block starting from \( S_i \) contains all elements up to (and including) \( S_{j-1} \), for the smallest \( j \) such that one of the following three conditions is met:

1. \( j \geq i + \sqrt{n} \);
2. \( S_j \) is left-incompatible with \( H = \{S_1, \ldots, S_{j-1}\} \);

\(^5\)In the following we will often say just left-incompatible or right-incompatible, without explicitly specifying the sets \( H \) and/or \( I \), when there is no ambiguity.
3. $S_j$ is right-incompatible with $H$ and the set $I$ of informed nodes before the $i$-th step in $pp$-$a$.

Note that this block is incompatible-free with respect to $I$. If the element $S_j$ after the last element of the above normal block is right-incompatible with $H, I$, then the next block is a special one, containing the single element $S_j$; while if $S_j$ is not right-incompatible (even if it is left-incompatible), the next block is a normal one. Finally, a block that is right after a special block is normal.

Let $B_1, B_2, \ldots$ be the sequence of blocks as defined above. The coupling is as follows. Suppose that $B_k = \{S_i, \ldots, S_{j-1}\}$ is a normal block. In $pp$-$a$ we just set $(a_i, b_i) = (x_t, y_t)$, for all $i \leq t < j$. In $pp$ we have a single round, $r$, corresponding to $B_k$, and we set $c_{r,x_t} = y_t$ for all $i \leq t < j$, while for the remaining nodes $v \in V \setminus \{x_i, \ldots, x_{j-1}\}$, we can assume for now that they do not start any communication, as this can only increase the rumor spreading time of $pp$. Later (right before Lemma 16), we describe how to choose the values of $c_{r,v}$, for $v \in V \setminus \{x_i, \ldots, x_{j-1}\}$.

Suppose now that $B_k = \{S_i\}$ is a special block, i.e., $S_i$ is right-incompatible with $B_{k-1}$ and $I$, for $I$ the set of informed nodes in $pp$-$a$ before the first step in block $B_{k-1}$. Let $j^* = \min\{j : R_{i,j} \text{ contains an element that is right-incompatible with } B_{k-1}, I\}$. For block $B_k$ then we have $j^*$ rounds in $pp$, and if $r$ is the index of the round corresponding to the previous block $B_{k-1}$, then we set $\{(v, c_{r+j,v})\}_{v \in V} = R_{i,j^*}$ for $1 \leq j \leq j^*$. For $pp$-$a$ we set $(a_i, b_i)$ to be a pair from $R_{i,j^*}$ that is right-incompatible with $B_{k-1}, I$. Precisely, if $A$ is the set of all possible pairs that are right-incompatible with $B_{k-1}, I$, then we set $(a_i, b_i)$ to be a element from the set $D = A \cap R_{i,j^*}$, drawn at random according to a distribution $\mu_{A|D}$ with the property that

$$\forall (a, b) \in A, \quad \mu_A(a, b) = \Pr[S = (a, b) \mid S \in A], \quad (1)$$

where $\mu_A(a, b) = \sum_{C \subseteq A} \mu_{A|C}(a, b) \cdot \Pr[D = C]$. I.e., distribution $\mu$ ensures that the pair $(a, b)$ chosen has the same distribution as the pair $S$ in a random step of $pp$-$a$, given that $S$ is right-incompatible ($\mu_A$ averages over all possible rounds that contain at least one right-incompatible pair.) In Lemma 15 we show that such a probability distribution $\mu$ exists.

In the following, we will overload the meaning of ‘block’ to refer also the set of steps of $pp$-$a$, or the set of rounds of $pp$ corresponding to the actual block. For example we will say ‘a step in a block,’ or ‘the set of informed nodes in $pp$ after a block.’

Distribution $\mu$. We now argue that a distribution $\mu$ with the properties described above exists. We do so in the next two lemmas. The first lemma states that given a collection of (possibly dependent) binary random variables $X_1, \ldots, X_k$, we can define a new random variable, $Y$, that chooses one among those variables that have value 1, in a way that the overall probability that $X_j$ is chosen, given that at least one of the $X_i$ is 1, is proportional to the expectation of $X_j$. Here is a simple numerical example: Say $k = 2$, $\Pr[X_1 = 1] = 1$, and $\Pr[X_2 = 1] = 1/2$; so, $p_1 = 1$, $p_2 = 1/2$, and $p = 3/2$ (in the statement of the lemma below). This means that with probability 1/2, $X_1 = X_2 = 1$, and with the remaining 1/2, $X_1 = 1$ and $X_2 = 0$. Defining $Y$ such that $\Pr[Y = 1 \mid (X_1, X_2) = (1,1)] = 1/3$ and $\Pr[Y = 1 \mid (X_1, X_2) = (1,0)] = 1$ satisfies the lemma, as $\Pr[Y = 1 \mid X_1 + X_2 > 0] = (1/2 \cdot 1/3) + 1/2 = 2/3 = p_1/p$ (note that when $X_1 = X_2 = 1$, $X_2$ must be selected with larger probability than $X_1$).

**Lemma 13.** Let $X_1, \ldots, X_k$ be 0/1 random variables with $E[X_i] = p_i > 0$. Let $K = \{i : X_i = 1\}$. We can define a random variable $Y$ that takes values in the set $[k] = \{1, \ldots, k\}$, in such a way that $\Pr[Y \in K \mid K \not= \emptyset] = 1$, and for any $i \in [k]$, $\Pr[Y = i \mid K \not= \emptyset] = p_i/p$, where $p = \sum_j p_j$.

**Proof.** We compile a system of linear equations whose solution would imply the existence of $Y$, and then describe an iterative method for solving this system which converges to a solution.
For \( N \subseteq [k] \) and \( i \in [k] \), let \( p_{N,i} = \Pr[(K = N) \land (Y = i) \mid K \neq \emptyset] \), and \( p_N = \Pr[K = N \mid K \neq \emptyset] \). Let \( \mathcal{K} = \{ N \subseteq [k] : p_N > 0 \} \) be the collection of all non-empty sets \( N \) with positive probability, and let \( \mathcal{K}_i = \{ N \in \mathcal{K} : i \in N \} \) be the subset of those \( N \) containing \( i \). The following set of equations must then hold:

\[
\forall N \in \mathcal{K}, \quad \sum_{i \in N} p_{N,i} = p_N; \quad (2)
\]

\[
\forall i \in [k], \quad \sum_{N \in \mathcal{K}_i} p_{N,i} = p_i/p. \quad (3)
\]

The set of equations (2) is equivalent to the lemma’s requirement that \( \Pr[Y \in K \mid K \neq \emptyset] = 1 \), as this is equivalent to: \( \forall N \in \mathcal{K}, \Pr[(K = N) \land (Y \in N) \mid K \neq \emptyset] = p_N \), and we also have \( \Pr[(K = N) \land (Y \in N) \mid K \neq \emptyset] = \sum_{i \in N} p_{N,i} \). The second set of equations, (3), is equivalent to the other requirement of the lemma, that for any \( i \in [k] \), \( \Pr[Y = i \mid K \neq \emptyset] = p_i/p \), because \( \Pr[Y = i \mid K \neq \emptyset] = \sum_{N \in \mathcal{K}_i} p_{N,i} \).

Therefore, to prove that \( Y \) exists we must show that the system of linear equations (2) and (3), with unknowns \( p_{N,i} \), \( N \in \mathcal{K} \) and \( i \in N \), has a non-negative solution (at least one).

We describe an iterate approach that, in the limit, converges to such a solution. We start by assigning to the unknowns arbitrary non-negative values such that all equations (2) are satisfied (but possibly not all equation (3) are). E.g., we can set \( p_{N,i}^0 = \frac{p_N}{|N|} \), for each \( N \in \mathcal{K} \) and \( i \in N \). By \( p_{N,i}^t \) we denote the value for \( p_{N,i} \) after the \( t \)-th iteration. For each \( t \), we maintain the invariants that for all \( N \in \mathcal{K} \), \( \sum_{i \in N} p_{N,i}^t = p_N \), and for all \( i \in N \), \( p_{N,i}^t \geq 0 \).

For each \( i \in [k] \), let \( \phi_{t,i} = \sum_{N \in \mathcal{K}_i} p_{N,i}^t - p_i/p \) be the discrepancy between the value of the sum in (3) after the \( t \)-th iteration and the desired value of this sum. We measure the distance from a solution after the \( t \)-th iteration using the potential function \( \Phi_t = \sum_i (\phi_{t,i})^2 \). Clearly, \( \Phi_t = 0 \) implies that a solution has been reached.

Assuming that \( \Phi_t > 0 \), in the \( (t+1) \)-th iteration we do the following: We choose some \( N \in \mathcal{K} \) and \( i, j \in N \), such that \( \phi_{t,i} > \phi_{t,j} \) and \( p_{N,i}^t > 0 \), and we set \( p_{N,i}^{t+1} = p_{N,i}^t - \epsilon \) and \( p_{N,j}^{t+1} = p_{N,j}^t + \epsilon \), for some \( \epsilon \leq p_{t,N,i}^t \), i.e., we move some probability mass \( \epsilon \) from \( p_{N,i}^t \) to \( p_{N,j}^t \); the rest of the values stay the same. (The requirement that \( \epsilon \leq p_{t,N,i}^t \) is necessary to ensure that the variables remain non-negative.) Among the, possibly, several choices for \( N, i, j, \epsilon \), we choose one that maximizes the drop in the potential \( \Phi_t - \Phi_{t+1} \). This potential drop is

\[
\Phi_t - \Phi_{t+1} = (\phi_{t,i})^2 - (\phi_{t+1,i})^2 + (\phi_{t,j})^2 - (\phi_{t+1,j})^2
= (\phi_{t,i} - \phi_{t+1,i})(\phi_{t,i} + \phi_{t+1,i}) + (\phi_{t,j} - \phi_{t+1,j})(\phi_{t,j} + \phi_{t+1,j})
= \epsilon(2\phi_{t,i} - \epsilon) - \epsilon(2\phi_{t,j} + \epsilon)
= 2\epsilon(\phi_{t,i} - \phi_{t,j} - \epsilon).
\]

We will prove the follow statement.

**Claim 14.** If \( \Phi_t > 0 \), there is \( N \in \mathcal{K} \) and \( i, j \in N \), such that \( \phi_{t,i} - \phi_{t,j} \geq \frac{\sqrt{\Phi_t}}{k^{3/2}} \) and \( p_{N,i}^t \geq \frac{\sqrt{\Phi_t}}{2|K| \cdot k^{5/2}} \).

Applying this result to the equation above, letting \( \epsilon = \frac{\sqrt{\Phi_t}}{2|K| \cdot k^{3/2}} \), yields

\[
\Phi_t - \Phi_{t+1} \geq 2 \cdot \frac{\sqrt{\Phi_t}}{2|K| \cdot k^{3/2}} \left( \frac{\sqrt{\Phi_t}}{k^{3/2}} - \frac{\sqrt{\Phi_t}}{2|K| \cdot k^{5/2}} \right) \geq \frac{\Phi_t}{2|K| \cdot k^4}.
\]

From this, it follows that \( \Phi_t \leq \left( 1 - \frac{\Phi_t}{2|K| \cdot k^4} \right)^t \Phi_0 \), and letting \( t \to \infty \), we have \( \Phi_t \to 0 \). Therefore, the method converges to a solution of the system, as promised.

It remains to prove Claim 14.
Proof of Claim 14. Without loss of generality we assume that \( \phi_t,1 \leq \phi_t,2 \leq \cdots \leq \phi_t,k \). There must be then some \( 1 \leq i^* < k \) such that \( \phi_t,i^*+1 - \phi_t,i^* \geq \frac{\phi_t,k-\phi_t,1}{k} \). Let \( \delta = \frac{\phi_t,k-\phi_t,1}{2k} \). It follows that at least one of the next two inequalities must hold: \( \phi_t,i^* \leq -\delta \) or \( \phi_t,i^*+1 \geq \delta \). Based on that we distinguish the two cases below. In both cases, we show that there are \( N \in \mathcal{K} \) and \( i,j \in N \), such that \( \phi_t,i - \phi_t,j \geq 2\delta \) and \( p_{N,i}^j \geq \frac{\delta}{|\mathcal{K}|} \).

**Case 1:** \( \phi_t,i^* \leq -\delta \). Let \( \mathcal{K}^* = \bigcup_{i \leq i^*} \mathcal{K}_i \). We have

\[
\sum_{N \in \mathcal{K}^*} p_N = \sum_{N \in \mathcal{K}^*} \sum_{i \in N: i \leq i^*} p_{N,i}^i + \sum_{N \in \mathcal{K}^*} \sum_{i \in N: i > i^*} p_{N,i}^i.
\]

The first double sum on the right side is

\[
\sum_{N \in \mathcal{K}^*} \sum_{i \in N: i \leq i^*} p_{N,i}^i = \sum_{i < i^*} \sum_{N \in \mathcal{K}_i} p_{N,i}^i = \sum_{i < i^*} \left( \frac{p_i}{p} + \phi_t,i \right) \leq \sum_{i < i^*} \left( \frac{p_i}{p} - \delta \right) = \sum_{i < i^*} \frac{p_i}{p} - i^* \delta = \sum_{N \in \mathcal{K}^*} \sum_{i \in N: i \leq i^*} p_{N,i} - i^* \delta \leq \sum_{N \in \mathcal{K}^*} p_N - i^* \delta.
\]

Substituting this above, yields

\[
\sum_{N \in \mathcal{K}^*} \sum_{i \in N: i > i^*} p_{N,i}^i \geq i^* \delta.
\]

Since the total number of terms in this double sum is at most \( |\mathcal{K}| \cdot k \), it follows that there is some \( N \in \mathcal{K}^* \) and \( i \in N \) with \( i > i^* \), for which \( p_{N,i}^i \geq \frac{i^* \delta}{|\mathcal{K}|} \). And since \( N \in \mathcal{K}^* \) there is also some \( j \leq i^* \), such that \( j \in N \). For this pair of \( i,j \) we have \( \phi_t,i - \phi_t,j \geq \phi_t,i^*+1 - \phi_t,i^* \geq 2\delta \).

**Case 2:** \( \phi_t,i^*+1 \geq \delta \). Let \( \mathcal{L}^* = \bigcup_{i > i^*} \mathcal{K}_i \). We have

\[
\sum_{N \in \mathcal{L}^*} p_N = \sum_{N \in \mathcal{L}^*} \sum_{i \in N: i \leq i^*} p_{N,i}^i + \sum_{N \in \mathcal{L}^*} \sum_{i \in N: i > i^*} p_{N,i}^i.
\]

The second double sum on the right side is

\[
\sum_{N \in \mathcal{L}^*} \sum_{i \in N: i > i^*} p_{N,i}^i = \sum_{i > i^*} \sum_{N \in \mathcal{K}_i} p_{N,i}^i = \sum_{i > i^*} \left( \frac{p_i}{p} + \phi_t,i \right) \geq \sum_{i > i^*} \left( \frac{p_i}{p} + \delta \right) = \sum_{i > i^*} \frac{p_i}{p} + (k-i^*) \delta = \sum_{N \in \mathcal{L}^*} \sum_{i \in N: i > i^*} p_{N,i} + (k-i^*) \delta = \sum_{N \in \mathcal{L}^*} \sum_{N \in \mathcal{K}_i} p_{N,i} + (k-i^*) \delta
\]

Substituting this above, yields

\[
\sum_{N \in \mathcal{L}^*} \sum_{i \in N: i \leq i^*} (p_{N,i} - p_{N,i}^i) \geq (k-i^*) \delta.
\]

This implies that there is some \( N \in \mathcal{L}^* \) for which \( \sum_{i \in N: i \leq i^*} (p_{N,i} - p_{N,i}^i) \geq \frac{(k-i^*) \delta}{|\mathcal{K}|} \). And since it is \( \sum_{i \in N} p_{N,i} = \sum_{i \in N} p_{N,i}^i = p_N \), it follows that \( \sum_{i \in N: i > i^*} (p_{N,i}^i - p_{N,i}) \geq \frac{(k-i^*) \delta}{|\mathcal{K}|} \). From this it follows that there is some \( i \in N \) with \( i > i^* \), for which \( p_{N,j}^i \geq p_{N,j} + \frac{(k-i^*) \delta}{|\mathcal{K}|} \), and also there is some \( j \in N \) with \( j \leq i^* \). For this pair of \( i,j \) we have \( \phi_t,i - \phi_t,j \geq \phi_t,i^*+1 - \phi_t,i^* \geq 2\delta \).
In both cases above, we have shown that there is \( N \in \mathcal{K} \) and \( i, j \in N \), such that \( \phi_{t,i} - \phi_{t,j} \geq 2\delta \) and \( p'_{N,i} \geq \frac{\delta}{|X|/k} \). To complete the proof it suffices to show that \( \delta \geq \frac{\sqrt{\delta_1}}{2k^{3/2}} \).

Recall that \( \delta = \frac{\phi_{t,k} - \phi_{t,1}}{2k} \), and \( \phi_{t,1} \leq \phi_{t,2} \leq \cdots \leq \phi_{t,k} \). From definition \( \Phi_t = (\phi_{t,1})^2 + \cdots + (\phi_{t,k})^2 \), it follows that \((\phi_{t,1})^2 + (\phi_{t,k})^2 \geq \frac{\delta^2}{k} \). And since \( \phi_{t,1} \leq 0 \leq \phi_{t,k} \), we have \((\phi_{t,k} - \phi_{t,1})^2 \geq (\phi_{t,1})^2 + (\phi_{t,k})^2 \geq \frac{\delta^2}{k} \). Hence, \( \phi_{t,k} - \phi_{t,1} \geq \sqrt{\Phi_t/k} \), and substituting to the definition of \( \delta \), we obtain that \( \delta \geq \frac{\sqrt{\delta_1}}{2k^{3/2}} \).

This concludes the proof of Claim 14, and of Lemma 13. \( \square \)

We now use the result above to prove our main lemma.

**Lemma 15.** Let \( A \) be a set of pairs of adjacent nodes. There exists a collection of distributions \( \mu_{A|C} \), for \( C \subseteq A \) and \( C \neq \emptyset \), such that \( \mu_{A|C}(a, b) \neq 0 \) only if \( (a, b) \in C \), and the distribution \( \mu_A \) with \( \mu_A(a, b) = \sum_{C \subseteq A} (\mu_{A,C}(a, b) \cdot \Pr[A \cap R = C \mid A \cap R \neq \emptyset]) \), for \( (a, b) \in A \), satisfies Eq. (1).

**Proof.** Let \( k = |A| \), and for \( 1 \leq i \leq k \), let \((u_i, v_i)\) denote the \( i \)-th element in \( A \) (with respect to some arbitrary order). Let \( X_i \) be the indicator random variable with \( X_i = 1 \) if \((u_i, v_i) \in R \), and \( X_i = 0 \) otherwise. Then \( \Pr[X_i] = 1/\deg(u_i) \). Let \( K = \{i: X_i = 1\} \). From Lemma 13, it follows that there is random variable \( Y \), such that \( \Pr[Y \in K \mid K \neq \emptyset] = 1 \), and for any \( 1 \leq i \leq k \),

\[
\Pr[Y = i \mid K \neq \emptyset] = \frac{1/\deg(u_i)}{\sum_j (1/\deg(u_j))}.
\] (4)

For any non-empty set \( C \subseteq A \), we define

\[
\mu_{A|C}(u_i, v_i) = \Pr[Y = i \mid A \cap R = C] = \Pr[Y = i \mid K = \{j: (u_j, v_j) \in C\}].
\]

Note that this definition satisfies the lemma’s requirement that \( \mu_{A|C}(a, b) \neq 0 \) only if \( (a, b) \in C \), as this follows from the fact that \( \Pr[Y \in K \mid K \neq \emptyset] = 1 \) mentioned earlier. It remains to show that it satisfies (1) as well. We have

\[
\mu_A(u_i, v_i) = \sum_{C \subseteq A} (\mu_{A,C}(u_i, v_i) \cdot \Pr[A \cap R = C \mid A \cap R \neq \emptyset])
\]

\[
= \sum_{C \subseteq A} (\Pr[Y = i \mid A \cap R = C] \cdot \Pr[A \cap R = C \mid A \cap R \neq \emptyset])
\]

\[
= \Pr[Y = i \mid A \cap R \neq \emptyset]
\]

\[
= \Pr[Y = i \mid K \neq \emptyset].
\]

For the right side of (1) we have

\[
\Pr[S = (u_i, v_i) \mid S \in A] = \frac{\Pr[S = (u_i, v_i) \wedge S \in A]}{\Pr[S \in A]} = \frac{\Pr[S = (u_i, v_i)]}{\Pr[S \in A]}
\]

\[
= \frac{(1/n) \cdot (1/\deg(u_i))}{\sum_j ((1/n) \cdot (1/\deg(u_j)))}
\]

\[
= \frac{1/\deg(u_i)}{\sum_j (1/\deg(u_j))}.
\]

From the last two results above and (4), it follows \( \mu_A(u_i, v_i) = \Pr[S = (u_i, v_i) \mid S \in A] \), as desired. This completes the proof of Lemma 15. \( \square \)
Correctness and Domination. Next we argue that the coupling yields the correct marginal distributions for the two processes, and thus we have a proper coupling. Recall, we have assumed that in a round \( r \) that corresponds to a normal block \( B = \{S_t, \ldots, S_{j-1}\} \), nodes \( v \in V \setminus \{x_i, \ldots, x_{j-1}\} \) do not start a communication, and we have neglected defining \( c_{r,v} \) for them. We do so now, in order to facilitate the next lemma. We distinguish two cases:

(C1) If \( |B| = \sqrt{n} \), or \( S_j \) is left-incompatible with \( B \), then for each \( v \in V \setminus \{x_i, \ldots, x_{j-1}\} \), \( c_{r,v} \) is chosen independently at random from \( \text{Unif}(\Gamma(v)) \).

(C2) If \( |B| < \sqrt{n} \) and \( S_j \) is right-incompatible with \( B, I \) (\( I \) being the set of informed nodes before the \( i \)-th step of \( pp-a \)), then \( c_{r,x_j} = y_j \), and for each \( v \in V \setminus \{x_i, \ldots, x_j\} \), \( c_{r,v} \) is chosen independently at random from \( \text{Unif}(\Gamma(v)) \).

The only difference between the two cases is that in the latter, we set \( c_{r,x_j} = y_j \) instead of choosing its value independently. To see why we need to treat that second case differently consider the following. Suppose that the current block is \( S_t, \ldots, S_t \), and that node \( v \) got informed in one of the steps \( i \) up to \( t \) in \( pp-a \), while its neighbor \( u \) did not participate in any communication. Suppose also that \( u \) has another neighbor \( v' \) that did not get informed in those steps. Then we have that \( S_{t+1} \) is equally likely to be \((u,v')\) or \((u,v)\). In the first case \((u,v')\) will be added to the current block and we will have \( c_{r,u} = v' \). But in the second case, if we did not set \( c_{r,u} = v' \), as in (C2), but instead we let \( c_{r,u} \) be a uniformly random neighbor of \( u \), as in (C1), then overall (taking into account both cases), \( c_{r,u} \) would be more likely to be \( v' \) than \( v \) in the round.

**Lemma 16.**

(a) Let \( X_i = (a_t, b_t) \). The random variables \( X_1, X_2, \ldots \) are independent and \( X_i \sim S \).

(b) Let \( Y_r = \{(v, c_{r,v})\}_{v \in V} \). The random variables \( Y_1, Y_2, \ldots \) are independent and \( Y_r \sim R \).

**Proof.** (a): Suppose we fix \( X_1, \ldots, X_t \), for some \( t \geq 0 \). To prove the claim it suffices to show that \( X_{t+1} \sim S \). If \( t = 0 \) then \( X_{t+1} = S_1 \sim S \); so suppose that \( t > 0 \). Consider the partition into blocks of the variables \( S_1, \ldots, S_t \), and let \( B = \{S_t, \ldots, S_t\} \) denote the last one of these blocks. Let \( I \) denote the set of informed nodes in \( pp-a \) before the \( k \)-th step. Note that to determine \( B \) and \( I \), it suffices to know the variables \( S_i \) and \( R_{i,j} \) for \( i \leq t \); we do not need to know any of the variables \( S_i \) or \( R_{i,j} \) for \( i > t \).

Next we distinguish the following cases.

Case 1: If \( B \) is a special block, or it is a normal block with \( |B| = \sqrt{n} \), then a new normal block starts with \( S_{t+1} \), and we have \( X_{t+1} = S_{t+1} \sim S \).

Case 2: If neither of the above conditions hold, then \( B \) is a normal block, and we expose whether or not \( X_{t+1} \) is right-compatible with \( B, I \). If \( X_{t+1} \) is not right-incompatible, then either \( S_{t+1} \) is added to \( B \) or (if \( X_{t+1} \) is left-incompatible with \( B \)) a new normal block starts with \( S_{t+1} \), and we have \( X_{t+1} = S_{t+1} \); if \( S_{t+1} \) is right-incompatible, and thus \( \{S_{t+1}\} \) is a special block, then our coupling lets \( X_{t+1} \) be some right-incompatible pair (determined based on the variables \( R_{i,j} \)), which has the same distribution as the conditional distribution of \( S \) give that \( S \) is right-incompatible with \( B, I \) (see Eq. (1)). Combining these two subcases, we obtain that \( X_{t+1} \sim S \).

(b): Suppose we fix \( Y_1, \ldots, Y_r \), for some \( r \geq 0 \). It suffices then to show that \( Y_{r+1} \sim R \). We distinguish the following cases.

Case 1: If \( r = 0 \), or \( Y_r \) corresponds to a normal block of size \( \sqrt{n} \), or \( Y_r \) corresponds to a special block \( \{S_t\} \) and was set equal to some \( R_{t,j} \) containing an element that is right-incompatible with the previous block \( B \) and the set \( I \) of informed nodes in \( pp-a \) before \( B \), then in all these cases
Lemma 17. For any \( k \geq 0 \), \( I_k(pp-a) \subseteq I_k(pp) \).

Proof. For \( k = 0 \), \( I_k(pp-a) = I_k(pp) = \emptyset \). Suppose that \( I_{k-1}(pp-a) \subseteq I_{k-1}(pp) \), for some \( k > 0 \). We will show that \( I_k(pp-a) \subseteq I_k(pp) \).

If \( B_k = \{S_t\} \) is a special block, then by our coupling we have that the pair of nodes that communicate in step \( t \) of \( pp-a \) is set to be one of the communication pairs in the last round \( r \) of \( pp \) that corresponds to \( B_k \), i.e., \( (a_t,b_t) \in \{(v,c_{r,v})\}_{v \in V} \). The claim then follows.

If \( B_k = \{S_i, \ldots, S_{j-1}\} \) is a normal block, then the node pairs communicating in the steps of \( pp-a \) that correspond to \( B_k \) is a subset of the communication pairs in the (single) round \( r \) of \( pp \) that corresponds to \( B_k \), i.e., \( \{(a_t,b_t)\}_{i \leq t < j} \subseteq \{(v,c_{r,v})\}_{v \in V} \). And since the set to the left is incompatible-free with \( I_{k-1}(pp-a) \) (see Remark 12), the claim follows.

Bounding Rounds by Steps. Next we bound the expected number of rounds in \( pp \) that correspond to the steps in \( pp-a \) before all nodes get informed. Let \( I_t \) denote the set of informed nodes in \( pp-a \) after the first \( t \) steps, and let \( \tau = \min\{t: I_t = V\} \) be the number of those steps before all nodes get informed. Let \( \rho_t \) denote the number of rounds in \( pp \) that correspond to the first \( t \) steps of \( pp-a \); precisely, \( \rho_t \) is the number of rounds that correspond to the first \( k \) blocks, where \( k \) is the index of the block containing \( S_t \).

Lemma 18. \( E[\rho_T] = O\left(\frac{E[\tau]}{\sqrt{n}} + \sqrt{n}\right) \).

Proof. Let \( B_1, \ldots, B_k \) denote the blocks into which \( S_1, \ldots, S_t \) are partitioned. We break down \( \rho_t \) into the following four terms:

\( Y_{r+1} \) corresponds to a new normal block \( B' \). Further, no information about \( B' \)'s elements (and in particular about its first element) has been exposed. Recall now that \( B' \) ends as soon as (i) \( \sqrt{n} \) elements have been added to it, or (ii) an element that is left-incompatible with the rest of the set is found, or (iii) an element that is right-incompatible is found. Because of (ii), \( B' \) contains no two pairs with the same first node, thus all communications in \( B' \) can be executed in parallel in a single round. If terminating condition (i) applies, then from this and the assumption (C1) for choosing the remaining \( c_{r+1,v} \), it follows that \( Y_{r+1} \sim R \). Similarly, from (ii) and (C1) we get that \( Y_{r+1} \sim R \) because the fact that the last pair \( S_t = (x_t,y_t) \) exposed is left-incompatible with the rest and is not included in \( Y_{r+1} \), does not affect the distribution of \( c_{r+1,x_t} \). Last, from (iii) and (C2) we obtain \( Y_{r+1} \sim R \) because all pairs exposed are included in \( Y_{r+1} \).

Case 2: If \( Y_r \) corresponds to a normal block \( B \) with \( |B| < \sqrt{n} \), and the element \( S_t \) after \( B \) is left-incompatible with \( B \), then \( Y_{r+1} \) corresponds to a new normal block \( B' \). This case is similar to Case 1, expect that some information about the first element \( S_t = (x_t,y_t) \) of \( B' \) has been exposed, namely that it is left-incompatible with \( B \). However, this information does not change the distribution of \( c_{r+1,x_t} \). The rest of the proof is identical to that for Case 1.

Case 3: If \( Y_r \) corresponds to a normal block \( B \) with \( |B| < \sqrt{n} \), and the element \( S_t \) after \( B \) is right-incompatible with \( B \), then \( Y_{r+1} \) corresponds to a new special block, \( \{S_t\} \), and \( Y_{r+1} = R_{t,1} \sim R \).

Case 4: If \( Y_r \) corresponds to a special block \( \{S_t\} \) and was set equal to some \( R_{t,j} \) containing no right-incompatible elements with the previous block \( B \), then \( Y_{r+1} = R_{t,j+1} \sim R \).
1. The number $\rho_{t,\text{full}}$ of rounds corresponding to (normal) blocks $B_k$, $k \leq k_t$, with $|B_k| = \sqrt{n}$.

2. The number $\rho_{t,\text{left}}$ of rounds corresponding to normal blocks $B_k = \{S_i, \ldots, S_{j-1}\}$, with $k \leq k_t$ and $|B_k| < \sqrt{n}$, such that $S_j$ is left-incompatible with $B_k$.

3. The number $\rho_{t,\text{right}}$ of rounds corresponding to normal blocks $B_k = \{S_i, \ldots, S_{j-1}\}$, with $k \leq k_t$ and $|B_k| < \sqrt{n}$, such that $S_j$ is right-incompatible with $B_k, I_{k-1}(pp-a)$.

4. The number $\rho_{t,\text{special}}$ of rounds corresponding to special blocks.

We have $\rho_t = \rho_{t,\text{full}} + \rho_{t,\text{left}} + \rho_{t,\text{right}} + \rho_{t,\text{special}}$. Since $\rho_{t,\text{full}} \leq \frac{4}{\sqrt{n}}$, and $\rho_{t,\text{right}} \leq \rho_{t,\text{special}} + 1$, it follows $\rho_t \leq \frac{4}{\sqrt{n}} + \rho_{t,\text{left}} + 2\rho_{t,\text{special}} + 1$. Thus,

$$E[\tau] = O\left(\frac{E[\tau]}{\sqrt{n}} + E[\rho_{t,\text{left}}] + E[\rho_{t,\text{special}}] + 1\right).$$

We show that $E[\rho_{t,\text{left}}] = E[\rho_{t,\text{special}}] \leq 2\sqrt{n}$. Substituting these above yields the claim.

The bound on $\rho_{t,\text{left}}$ is based on the following observation. For any $t \geq 1$, and any way of fixing the first $t-1$ steps of $pp-a$, the probability that $S_t = (x_t, y_t)$ is left-incompatible with $\{S_i, \ldots, S_{t-1}\}$ (i.e., $x_t \in \{x_1, \ldots, x_{t-1}\} \cup \{y_1, \ldots, y_{t-1}\}$), where $S_t$ denotes the first element in the block containing $S_{t-1}$, is at most $\frac{2(t-i)}{n} \leq \frac{2}{\sqrt{n}}$. The reason is that $x_t$ is chosen uniformly at random among all $n$ nodes, and at most $2(t-i)$ distinct nodes appear in the pairs $S_i, \ldots, S_{t-1}$, while the number of those pairs is $t - i \leq \sqrt{n}$.

For $t \geq 0$, define $Z_t = \rho_{t,\text{left}} - \frac{2t}{\sqrt{n}}$. The sequence $Z_0, Z_1, \ldots$ is a supermartingale with respect to $X_1 = (a_1, b_1), X_2 = (a_2, b_2), \ldots$ because: (1) $\rho_{t,\text{left}}$ (and thus $Z_t$) is a function of $X_1, \ldots, X_t$, as this sequence completely determines the collection of simple blocks into which $S_1, \ldots, S_t$ are divided; and (2) for $t \geq 1$, $E[Z_t - Z_{t-1} | X_1 \ldots X_{t-1}] = \Pr[\rho_{t,\text{left}} - \rho_{t-1,\text{left}} = 1 | X_1 \ldots X_{t-1}] - \frac{2}{\sqrt{n}} \leq 0$, as we argued above.

We can now use the optional stopping theorem for this supermartingale sequence and the stopping time $\tau$. Since for any $t \geq 1$, the different $|Z_t - Z_{t-1}| \leq 1$ is bounded, and $E[\tau]$ is finite (bounded by $n^2 \log n$), the optional stopping theorem yield $E[Z_\tau] \leq E[Z_0]$. Substituting $Z_\tau = \rho_{t,\text{left}} - \frac{2\tau}{\sqrt{n}}$ and $Z_0 = 0$, we obtain $E[\rho_{t,\text{left}}] \leq \frac{2\tau}{\sqrt{n}}$.

Next we bound $\rho_{t,\text{special}}$. In fact, the $O(\sqrt{n})$ bound we prove holds for any $t$, not just for $t = \tau$. For each node $v$, we bound the expected number of rounds that correspond to special blocks $\{S_t\}$ with $y_t = v$, and we sum these expectations over all $v$ to obtain a bound on $E[\rho_{t,\text{special}}]$. Note that for each $v$, there is at most one special block $\{S_t\}$ with $y_t = v$, as this block must follow right after the block in which $v$ gets informed in $pp-a$. To prove the bound we just use a weaker fact: If $t_v$ denotes the step when $v$ gets informed in $pp-a$, and $\{S_t\}$ is a special block with $y_t = v$, then it must be $t \leq t_v + \sqrt{n}$, as the maximum block size is $\sqrt{n}$.

Let us fix $X_1, \ldots, X_{t-1}$, for some $t_v < t \leq t_v + \sqrt{n}$, and let $S_t$ be the first element in the block containing $S_{t_v}$. The probability that $\{S_t\}$ is a special block with $y_t = v$, is the probability that $v$ is contacted in the step by one of its neighbors in $V_t(v) = V(v) \setminus \{x_v, \ldots, x_{t-1}\} \cup \{y_t, \ldots, y_{t-1}\}$, and this probability is $\pi_t(v) = \frac{1}{n} \sum_{w \in V_t(v)} \frac{1}{\deg(w)}$ (we do not take into account $v$’s neighbors that have already appeared in a pair in $\{S_i, \ldots, X_{t-1}\}$, as $S_t$ should not be left-incompatible). Given now that $\{S_t\}$ is a special block with $y_t = v$, the expected number of rounds that correspond to this

\[\text{[The sequence may not determine the element of a special block, as this element is replaced by another right-incompatible pair.]}\]
block is at most $1/q_t(v)$, where $q_t(v)$ is the probability that $v$ is contacted by a neighbor from $\Gamma_t(v)$ in a given round, and

$$q_t(v) = 1 - \prod_{w \in \Gamma_t(v)} \left(1 - \frac{1}{\deg(w)}\right) \leq 1 - e^{-\sum_{w \in \Gamma_t(v)} \frac{1}{\deg(w)}} = 1 - e^{-n\pi_t(v)}.$$ 

Therefore, the expected number of rounds as a result of the possibility that \(\{S_t\}\) is a special block with \(y_t = v\), is at most

$$\frac{\pi_t(v)}{q_t(v)} \leq \frac{\pi_t(v)}{1 - e^{-n\pi_t(v)}} = \frac{1}{n} \cdot \frac{n\pi_t(v)}{1 - e^{-n\pi_t(v)}} \leq \frac{1}{n} \cdot (1 + n\pi_t(v)) = \frac{1}{n} + \pi_t(v) \leq \frac{1}{n} + \pi(v),$$

where $\pi(v) = \frac{1}{n} \sum_{w \in \Gamma(v)} \frac{1}{\deg(w)}$, and the middle inequality above was obtained using the fact that $e^{-x} \leq \frac{1}{1+x}$, for $x \geq 0$. Summing now over all $t_v < t \leq t_v + \sqrt{n}$, and over all $v$, we obtain

$$E[\rho_{r,special}] \leq \sqrt{n} \cdot \sum_{v \in V} \left(\frac{1}{n} + \pi(v)\right) = \sqrt{n} \left(1 + \sum_{v \in V} \pi(v)\right) = 2\sqrt{n}.$$

This completes the proof of Lemma 18. \qed

**Proof of Theorem 11.** We have described a coupling between processes $pp$ and $pp-a$. In Lemma 15 and Lemma 16 we have proved the correctness of this coupling. In Lemma 17 we have shown that after each block, defined by the coupling, the set of informed nodes in $pp-a$ is a subset of the set of informed nodes in $pp$. In Lemma 18, we have shown that the expected number of rounds of $pp$ in the blocks until all nodes get informed in $pp-a$, is $E[\rho_r] = O \left(\frac{E[\tau]}{\sqrt{n}} + \sqrt{n}\right)$, where $\tau$ is the number of steps before all nodes get informed in $pp-a$. From the last two results it follows that the expected number of rounds before all nodes get informed in $pp$ is $E[T(pp, G, u)] \leq E[\rho_r] = O \left(\frac{E[\tau]}{\sqrt{n}} + \sqrt{n}\right)$. Finally, the expected time before all nodes get informed in $pp-a$ is $E[T(pp-a, G, u)] = \tau/n$, because the expected time between any two consecutive steps is $1/n$, and the times between steps are independent and also independent of $\tau$. It follows that $E[T(pp, G, u)] = O \left(\sqrt{n} \cdot E[T(pp-a, G, u)] + \sqrt{n}\right)$. \qed