Gossip vs. Markov Chains, and Randomness-Efficient Rumor Spreading

Zeyu Guo*
California Institute of Technology
Pasadena, USA
zguo@caltech.edu

He Sun†
Max Planck Institute for Informatics
Saarbrücken, Germany
hsun@mpi-inf.mpg.de

Abstract

We study gossip algorithms for the rumor spreading problem which asks one node to deliver a rumor to all nodes in an unknown network. We present the first protocol for *any* expander graph G with n nodes such that, the protocol informs every node in $O(\log n)$ rounds with high probability, and uses $\widetilde{O}(\log n)$ random bits in total. The runtime of our protocol is tight, and the randomness requirement of $\widetilde{O}(\log n)$ random bits almost matches the lower bound of $\Omega(\log n)$ random bits for dense graphs. We further show that, for many graph families, polylogarithmic number of random bits in total suffice to spread the rumor in $O(\text{poly}\log n)$ rounds. These results together give us an almost complete understanding of the randomness requirement of this fundamental gossip process.

Our analysis relies on unexpectedly tight connections among gossip processes, Markov chains, and branching programs. First, we establish a connection between rumor spreading processes and Markov chains, which is used to approximate the rumor spreading time by the mixing time of Markov chains. Second, we show a reduction from rumor spreading processes to branching programs, and this reduction provides a general framework to derandomize gossip processes. In addition to designing rumor spreading protocols, these novel techniques may have applications in studying parallel and multiple random walks, and randomness complexity of distributed algorithms.

Keywords: distributed computing, rumor spreading, Markov chains, randomness complexity, branching programs

*This work is supported by NSF grant CCF-1116111. Part of this work was done while visiting Max Planck Institute for Informatics.

†This work has partially been funded by the Cluster of Excellence "Multimodal Computing and Interaction" within the Excellence Initiative of the German Federal Government. Part of this work was done while visiting California Institute of Technology.

1 Introduction

Gossip algorithms is one of the most important communication primitives in large networks, and has been studied under different names such as rumor spreading, information dissemination, or broadcasting. Efficient gossip algorithms for information spreading have wide applications in failure detection [38], resource discovery [30], replicated database systems [11, 18], and modeling the spread of computer viruses [3]. Besides computer science, the dynamics of such processes in social networks also constitutes a research topic in economics and sociology.

The simplest and widely studied form of gossip algorithms is the so-called *push model* of rumor spreading. Initially, a message, called *a rumor*, is placed on an arbitrary node of an unknown network with n nodes. In subsequent synchronous rounds, every node that knows the rumor picks a neighbor uniformly at random and sends the rumor to the chosen neighbor. This process continues until every node gets the rumor. It was shown that this simple protocol is very efficient on several network topologies [16–18, 23]. In particular, its *runtime*, the number of rounds required until every node gets the rumor with high probability, is logarithmic in the number of nodes in the graph. Graphs satisfying this property range from complete graphs, hypercubes, Erdős-Rényi random graphs, and "quasi-regular" expanders (i.e., expander graphs for which the ratio between the maximum and minimum degree is constant). In addition to its efficiency, the protocol is local (i.e., no knowledge of global graph structure is needed), simple, and can tolerate link failures. More recently, several variations of information spreading protocols have been proposed to allow information to spread efficiently on networks with weak expansion properties [6], arbitrary networks [7], and dynamic networks [15].

Most of these algorithms are inherently randomized in both their design and analysis in that they crucially rely on choosing neighbors *independently and uniformly at random* in each round, i.e., we assume that every node of the graph has access to a random source of unbiased and independent coins. However, it is not known how to physically realize this abstraction in the real world and, from a theoretical point of view, it is not clear if this randomization is essential for efficiently disseminating the rumor. Hence the randomness requirement, the number of random bits used in total in order to spread the rumor efficiently, becomes a key measurement to evaluate rumor spreading protocols. One of the most studied questions concerns the randomness requirement: how many random bits are sufficient to efficiently spread a rumor to all nodes in a graph? While for any graph with n nodes, the above-mentioned *fully-random* push protocol requires $O(T \cdot n \log n)$ random bits for spreading a rumor within T rounds, it is not difficult to show that for any graph G of n nodes, there is a protocol which uses $3 \log n$ random bits in total, and whose runtime is as fast as the standard fully-random protocol (cf. Corollary B.2). However, the explicit construction of such protocols is more complicated, and a long line of research has been devoted to finding randomness-efficient protocols, see [13, 24, 25] for instance.

1.1 Our Results

In this paper we establish a novel reduction from the problem of designing rumor spreading protocols of low randomness complexity to the problem of constructing pseudo-random generators (PRGs) for branching programs. To the best of our knowledge, this reduction gives the first application of the model of branching programs in the area of distributed computing and also provides a powerful tool for designing gossip algorithms.

At a high level, the connection between gossip processes and branching programs is natural because (1) random walks over branching programs resemble the rumor spreading process where nodes send messages to random neighbors, and (2) in a rumor spreading protocol, each node has access to only its own list of neighbors, and is oblivious to the structure of the network. This is an analogue of *oblivious derandomization* achieved by PRGs. However, rumor spreading appears much more complicated than small-space computation due to the following facts: (1) In

the rumor spreading process, rumors are "duplicated" every round, although every "existing" rumor viewed individually performs a random walk. Hence, instead of considering every single random walk performed by any fixed rumor, we need to study the dynamics of the whole rumor spreading process. (2) The state of the process at some time essentially depends on the past behavior of all nodes and is by no means computable in small space. Indeed, even knowing if a single node u gets the rumor at some round requires knowing the set of its neighbors having the rumor in the previous rounds, and may require $\deg(u) = \Theta(n)$ bits for dense graphs. For these reasons, this connection to small-space computation is delicate and not obvious.

Surprisingly, we show that such a reduction from designing rumor spreading protocols to constructing PRGs for branching programs exists. Hence the question of designing randomness-efficient rumor spreading protocols is now exposed to the numerous techniques used in PRG constructions for small-space computation. In particular, PRGs with optimal parameters yield protocols whose randomness complexity matches the lower bound or the best known upper bound of existential results from the probabilistic method (cf. Theorem 2.7). Our result is as follows:

Theorem 1.1 (**Main Result**). *Let G be a graph with n nodes, spectral gap $\alpha \in (0,1)$ and irregularity $\beta \triangleq \Delta/\delta$. Then there is an explicit protocol using $O((\log(1/\alpha) + \log\beta) \cdot \log n) + \widetilde{O}(\log n)$ random bits such that with high probability all nodes get the rumor in $T = O(C \log n)$ rounds, where $C = (1/\alpha) \cdot \beta^2 \max\{1, 1/(\alpha \cdot \Delta^{0.499})\}$.*

Theorem 1.1 implies that, for *any* expander graph G with n nodes, $\alpha = \Theta(1)$ and irregularity $\beta = O(1)$, the protocol finishes in $O(\log n)$ rounds and uses $\widetilde{O}(\log n)$ random bits in total. Note that any protocol needs at least $\Omega(\log n)$ rounds to spread the rumor to all nodes, hence our runtime for expander graphs is tight. For the randomness complexity, our result improves the previous best bound of $O(\log^2 n)$ random bits [25]. Since for any expander graph with minimum degree $\delta = n^{\Theta(1)}$, any protocol that finishes in $O(\log n)$ rounds with high probability needs at least $\Omega(\log n)$ random bits (cf. Theorem C.2), our bound is almost tight.

We further study the so-called *averaging process*, which is a generalization of rumor spreading process and can be considered as the random matching model of load balancing with a certain initial load vector (cf. [21, 36]). We show that this general averaging process can be modeled by branching programs as well, which leads to an explicit averaging protocol. This approach implies the following result (Theorem 1.2) for the rumor spreading problem, and has independent interest in studying other distributed algorithms, e.g. quasi-random load balancing [22]. Due to page limitation, we defer the formal discussion about the averaging process to Section E.

Theorem 1.2. *Let G be a graph, $\mathsf{List}(u)$ be the adjacency list of node u, and $N(u)$ be the set of neighbors of u. We assume that each node u knows the ID of its neighbors $v \in N(u)$, and its index in $\mathsf{List}(v)$ for any neighbor $v \in N(u)$.*[1] *Then there is an explicit rumor spreading protocol using $O((\log(1/\alpha) + \log\beta + \log\log n) \cdot \log n)$ random bits, such that with high probability all nodes get the rumor in $T = O((1/\alpha) \cdot \beta^2 \log n)$ rounds.*

Our third result is for general graph with conductance ϕ. In contrast to Theorem 1.1 and Theorem 1.2 that are based on branching programs, this result relies on the observation that the rumor spreading process enjoys nice locality when the maximum degree is small.

Theorem 1.3. *Let G be a graph with n nodes, conductance ϕ and irregularity β. Then there is an explicit protocol using $O\big((1/\phi) \cdot \beta \cdot \log n \cdot (\log\log n + \log\Delta)\big)$ random bits in total, such that with high probability all nodes get the rumor in $O((1/\phi) \cdot \beta \cdot \log n)$ rounds.*

[1] We remark that similar assumptions are also made in other references, e.g. [29], and one can deterministically use $O(\Delta)$ preprocessing time to guarantee this assumption.

The runtime in Theorem 1.3 matches the upper bound known in the truly random protocol, and is tight, in the sense that there are graphs with diameter $\Omega((1/\phi)\log n)$ [8]. For the randomness requirement, our result improves the previous best one in [25], which needs $O((1/\phi)\log^2 n)$ random bits in total and only holds for graphs with $\beta = O(1)$.

Our protocol takes advantage of the locality by using a "two-level hashing" construction: We use a family of objects called unbalanced expanders to hash the node IDs into a smaller space, and then apply the classical pairwise independent generators. This construction yields much smaller seed length than using pairwise independent generators alone. The protocol has the advantage of being very simple. Furthermore, a variant of this protocol using PRGs for combinatorial rectangles achieves the best possible runtime for strong expanders:

Theorem 1.4. *Let G be a graph such that $\Delta/\delta = 1 + o(1)$ and $\alpha = 1 - o(1)$. Then there is a protocol using $O(\log n \cdot (\log\log n + \log\Delta))$ random bits in total, such that with high probability all nodes get the rumor in $\log n + \ln n + o(\log n)$ rounds.*

The runtime in Theorem 1.4 matches the precise runtime for the truly random protocol [16–18], and is known to be tight [17]. Moreover, our protocol uses $O(\log n \cdot (\log\log n + \log\Delta))$ random bits in total, in contrast to $\Omega(\log^3 n)$ random bits used for all previous protocols, e.g. [19, 25]. These four results (Theorem 1.1–Theorem 1.4), together with the existential proof (Corollary B.2) and the lower bound analysis (Theorem C.2), give us an almost complete understanding of the randomness complexity of this fundamental gossip problem.

Remark 1.5. *One common feature of our protocols is that all randomness is picked by the initial node having the rumor, and the whole rumor spreading process becomes deterministic once the random seed is picked. We remark that, through our protocol, the whole rumor spreading dynamics is encoded in this short random seed, and any node can recover the rumor spreading process once it receives the random seed. This feature may have applications in studying algebraic gossip algorithms, and other settings.*

1.2 Techniques

To derive the results above, we develop several new techniques for studying gossip processes. We highlight some of them in this subsection.

Approximation via Random Walks. The usual analyses for fast rumor spreading proceed by showing some measure (e.g. the volume of the set of informed/uninformed nodes) increases or decreases over time. Our approach is fundamentally different from previous work. Roughly speaking, we approximate the rumor spreading process by a collection of random walks and then use the rapid mixing of the random walks to prove the property of fast rumor spreading. It turns out that the pieces of local information provided by these random walks give a surprisingly good control of the global behavior of rumor spreading, despite that the walks are complicated and highly correlated.

Formally, we approximate the rumor spreading process by various random walks, distinguished by whether the walks are lazy or non-lazy in each round. Each walk is associated with a positive number called its *weight*. A node u is informed if the total weights of random walks reaching u is positive. By the Cauchy-Schwarz inequality, we lower bound the probability of this event in terms of the expectation of the total weights reaching u as well as its second moment.

Analysis of Markov Chains. With the weights chosen intelligently, the expectation and the second moment of total weights reaching a node are computed by certain Markov chains. The expected total weights are computed by the chain $\mathbf{M}$ representing a lazy random walk in the graph. It follows from the rapid mixing of $\mathbf{M}$ that it can be well estimated using the stationary distribution of $\mathbf{M}$. The case for the second moments is more complicated as they correspond

to a *non-reversible* chain $\mathbf{M}'$. A key result we manage to show is that $\mathbf{M}'$ and $\mathbf{M} \otimes \mathbf{M}$ have very close stationary distributions and comparable mixing time. We remark that this result is interesting on its own since $\mathbf{M}'$ is a very natural Markov chain, closely related to the Doeblin coupling [32].

Simulating Pull by Push. While a randomness-efficient protocols using a global seed can be easily implemented in the push model, the "dual" protocol in the pull model is not physically realizable, as it is impossible for a node to perform random pulls before getting the seed. Using the technique called *simulating pull by push*, we are able to employ the analysis for the pull model while actually using the push model. This is crucial in our analysis, since when most nodes already have the rumor, the random walks defined via push operations become too congested and correlated, whereas the "reversed" random walks using pull operations work well.

1.3 Related Work

There is a large amount of literature devoted to various aspects of rumor spreading. The majority of research studies the rumor spreading time in terms of the graph properties, e.g. conductance [8, 23], mixing time [4], diameter [18] and degree [18]. For instance, the first explicit connection between randomized rumor spreading and graph expansion was established by Mosk-Aoyama and Shah [34], who proved that on any regular graph with conductance ϕ, the protocol finishes in $O((1/\phi) \cdot \log n)$ rounds. More recent work includes the study of rumor spreading in social networks [14, 20] and dynamic graphs [10, 15], and algebraic gossip algorithms [28].

The study of determining and reducing the amount of randomness required for rumor spreading has been studied extensively in the past years. Doerr et al. [12] proposed a *quasi-random* version of the rumor spreading push protocol. In contrast to $O(n \log^2 n)$ random bits that used in the standard push model, the quasi-random rumor spreading model uses $\Theta(n \log n)$ random bits, and has been shown to be efficient on several graph topologies [13, 19]. Further progress along this line include [24, 25]. Besides this, researchers also studied the question of designing randomness-efficient or deterministic protocols for similar problems. For instance, Haeupler [29] presented one deterministic gossip algorithm for the k-local broadcast and the global broadcast problem. However, the algorithms in [29] require that all nodes in the graph have unique identifiers (UID), and every node knows its own and the neighbors' UIDs. Hence the techniques developed there cannot be applied to our setting.

In addition to rumor spreading, the technique of pseudorandomness was also studied in other settings of online algorithms, e.g., in the context of Local Computation Algorithms (LCA) [1], and complexity analysis of information spreading in dynamic networks [15].

1.4 Notations

Let $G = (V, E)$ be a connected, undirected, and simple graph with n nodes. For any node u, the degree of u is represented by $\deg(u)$. Let Δ, δ and d be the maximum, minimum and average degree of G, respectively, and call $\beta \triangleq \Delta/\delta$ the *irregularity* of G. We use $\mathbf{A}_G$ to express the adjacency matrix of G, and $\mathbf{N}_G \triangleq \mathbf{D}^{-1/2}\mathbf{A}_G\mathbf{D}^{-1/2}$, where $\mathbf{D}$ is the $n \times n$ diagonal matrix defined by $\mathbf{D}_{uu} = \deg(u)$ for $u \in V[G]$. Define the n real eigenvalues of $\mathbf{N}_G$ by $1 = \lambda_1 \geqslant \cdots \geqslant \lambda_n \geqslant -1$, and let $\lambda_{\max} \triangleq \max\{\lambda_2, |\lambda_i|\}$. The spectral gap α is defined by $\alpha \triangleq 1 - \lambda_2$, whereas the absolute spectral gap is defined as $1 - \lambda_{\max}$. For simplicity, we also use α to express the spectral expansion of a reversible Markov chain if the chain is clear from the context.

By $\log x$ we denote the binary logarithm of x. For any integer m, define $[m] \triangleq \{0, \ldots, m-1\}$. With high probability stands for with probability $1 - n^{-\Theta(1)}$.

2 Gossip vs. Markov Chains

Let $G = (V, E)$ be an undirected and simple graph with $V[G] = [n]$. We consider only T'-round protocols for G, in which nodes send rumors only for the first T' rounds, and assume that $T' = O(n^c)$ for a constant $c > 0$. Through this section, we assume that each node has a unique identifier (ID), and each node initially solely knows its own ID, which is from 0 to n^c for a constant c. Let s be the initial node having the rumor. For simplicity, we assume the adjacency list of each node u has length Δ, and the last $\Delta - \deg(u)$ neighbors are u itself, i.e. we add $\Delta - \deg(u)$ self-loops for ever node u. However, we use $\deg(u)$ and $N(u)$ to represent the degree and the set of neighbors of u respectively in the underlying simple graph.

2.1 Preliminaries

Given $d \in \mathbb{N}$ and a finite set $S = \prod_{i\in[d]} S_i$, define $\mathsf{CR}_S \triangleq \left\{\prod_{i\in[d]} A_i : A_i \subseteq S_i\right\}$. The members of CR_S are called *combinatorial rectangles in* S and d is their *dimension*. For $\varepsilon > 0$, $d \in \mathbb{N}$, and a finite set $S = \prod_{i\in[d]} S_i$, we call $\mathcal{G} : \{0,1\}^\ell \to S$ an ε-PRG for CR_S with seed length ℓ if $\left|\mathbf{Pr}_{x\in\{0,1\}^\ell}[\,\mathcal{G}(x) \in A\,] - |A|/|S|\right| \leqslant \varepsilon$ for any $A \in \mathsf{CR}_S$.

The second family of PRGs that we will use is PRGs for Branching Programs[2]. Let $\mathcal{B}$ be a branching program of length L, width W and degree D. For $x = (x_1, \ldots, x_L) \in [D]^L$ and a node $(s, 0)$ on the first layer, define $\mathcal{B}(s, x) \in [W]$ such that the random walk that starts from $(s, 0)$ and takes the edge with label x_i at the ith step for $1 \leqslant i \leqslant L$ finally arrives at $(\mathcal{B}(s, x), L)$. We call a function $\mathcal{G} : \{0,1\}^\ell \to [D]^L$ an ε-PRG for (L, W, D)-branching programs if for any (L, W, D)-branching program, and any node $(s, 0)$ on the first layer, it holds that

$$\sum_{u\in[W]} \left|\mathbf{Pr}_{x\in\{0,1\}^\ell}[\,\mathcal{B}(s, \mathcal{G}(x)) = u\,] - \mathbf{Pr}_{x\in[D]^L}[\,\mathcal{B}(s, x) = u\,]\right| \leqslant \varepsilon.$$

2.2 Analysis of the Prototype Protocol

In this subsection we relate rumor spreading processes to Markov chains, and show how the mixing time of certain Markov chains relates to the rumor spreading time. We first analyze the following prototype of rumor spreading protocols, which includes the standard push protocol as a special case.

Protocol 1 (Prototype of Rumor Spreading Protocols)**.** *Let $\mathcal{D}$ be a distribution over the set of functions $f : [T] \times V[G] \to [\Delta]$. Sample f according to $\mathcal{D}$. In the ith round, an informed node u sends the message to its $f(i, u)$th neighbor in its adjacency list.*

We are primarily interested in analyzing Protocol 1 when $\mathcal{D} = \mathcal{U}$ is the uniform distribution, i.e. $f(i, u)$ are chosen from $[\Delta]$ independently and uniformly at random for all i and u.

Approximation via Random Walks. To analyze the runtime of Protocol 1, we compare the process of rumor spreading with a random walk on a branching program. For random walks, a walk always stays at a single node throughout the process, although this node keeps changing. On the other hand, in the process of rumor spreading, each informed node u randomly sends the rumor to one of its neighbors v in each round, and then u, v are both informed subsequently. So we may think of rumor spreading as many random walks in parallel: When node u sends the rumor to v, one random walk moves from u to v whereas another one stays at u. In order to characterize this behavior, we introduce the notion of forward and reversed random walks. For any round $i \in [T]$ and node $u \in V[G]$, denote by $\widetilde{f}(i, u)$ the $f(i, u)$th neighbor of u in its adjacency list.

[2]See Definition D.7 for the formal definition of branching programs.

Definition 2.1 (Forward random walks)**.** *Consider a random rumor spreading process in T rounds on a graph G using Protocol 1 determined by $f \sim \mathcal{D} = \mathcal{U}$. A forward random walk of length $k \in [T]$ with pattern $S = (s_0, \ldots, s_{k-1}) \in \mathcal{C}_k \triangleq \{lazy, non\text{-}lazy\}^k$ is a sequence of $k+1$ nodes $(p_0, \ldots, p_k)$ of G, such that for all $i \in [k]$: (i) if $s_i = lazy$, then $p_{i+1} = p_i$; (ii) if $s_i = non\text{-}lazy$, then $p_{i+1} = \widetilde{f}(i, p_i)$.*

We also define reversed random walks, tailored to the idea of simulating pull using push. Roughly speaking, a reversed random walk takes a step from node v to u if u is the unique node pushing to v. For technical reasons, we introduce auxiliary random variables $r_{i,u}$ uniformly distributed over $[0,1]$ for each $i \in [T]$ and $u \in V[G]$ to equalize the probabilities of successful steps of reversed random walks made from different nodes. These random variables only appear in the analysis, not in the protocol constructions. Then the reversed random walks are determined by the randomness $f \sim \mathcal{D}$ together with $r_{i,u}$, whereas the forward walks are solely determined by f. See Definition D.10 for the formal definition of reversed random walks.

For $k \in [T/4]$, $u, v \in V[G]$ and $S \in \mathcal{C}_k = \{\text{lazy}, \text{non-lazy}\}^k$, let $X^S_{u,v}$ (resp. $Y^S_{u,v}$) be the indicator random variable of the event that the unique forward (resp. reversed) walk with pattern S and initial node u is at node v in the kth round. For $\gamma \in (0,1)$, let $\mathcal{D}_{\gamma,k}$ be the distribution over $\mathcal{C}_k$ where entries are independently chosen to be lazy with probability $1-\gamma$.

We fix an arbitrary node $w \in V[G]$, and study the probability that node w is informed in T rounds. Clearly, if there exist a forward random walk p from s to some node u and a reversed random walk p' from w to u, then the rumor is sent from s to u following p and then from u to w following the reversal of p'. Also note that the two walks exist if and only if $X^S_{s,u} Y^{S'}_{w,u} > 0$ for some S, S' and u. Therefore it holds for any $k \in [T/4]$ that

$$\mathbf{Pr}\left[\, w \text{ receives the message in } T \text{ rounds} \,\right] \geqslant \mathbf{Pr}\left[\sum_{S,S' \in \mathcal{C}_k, u \in V[G]} X^S_{s,u} Y^{S'}_{w,u} > 0 \right], \tag{2.1}$$

where the probability is taken over the randomness $f \sim \mathcal{D}$ and $r_{i,u}$.

We want to reduce the global event $\sum_{S,S' \in \mathcal{C}_k, u \in V[G]} X^S_{s,u} Y^{S'}_{w,u} > 0$ to local events $X^S_{s,u}$ and $Y^{S'}_{w,u}$. By using Cauchy-Schwarz inequality, and linearity of expectation, we show that (2.1) is lower bounded by

$$\frac{\sum_{u,v \in V[G]} \mathbf{E}_{r,S}\left[X^S_{s,u} \right] \mathbf{E}_{r,S}\left[X^S_{s,v} \right] \mathbf{E}_{r,S}\left[Y^S_{w,u} \right] \mathbf{E}_{r,S}\left[Y^S_{w,v} \right]}{\sum_{u,v \in V[G]} \mathbf{E}_{r,S,S'}\left[X^S_{s,u} X^{S'}_{s,v} \right] \mathbf{E}_{r,S,S'}\left[Y^S_{w,u} Y^{S'}_{w,v} \right]}. \tag{2.2}$$

Hence the runtime of Protocol 1 can be derived by analyzing multiple random walks individually or pairwisely. See Lemma D.11 for detailed analysis.

Analysis using Markov Chains. We study the expectations in (2.2) in terms of finite-state Markov chains. For simplicity, we represent these Markov chains by stochastic matrices. Recall that a stochastic matrix $\mathbf{M}'' \in \mathbb{R}^{n \times n} \otimes \mathbb{R}^{n \times n}$ is a *coupling* of $\mathbf{M}, \mathbf{M}' \in \mathbb{R}^{n \times n}$ if (i) $\sum_{x \in [n]} \mathbf{M}''_{(u,w)(v,x)} = \mathbf{M}_{u,v}$ for any $u, w, v \in [n]$, and (ii) $\sum_{v \in [n]} \mathbf{M}''_{(u,w)(v,x)} = \mathbf{M}'_{w,x}$ for any $u, w, x \in [n]$.

We define the "bi-lazy" analogue of lazy Markov chains with respect to a coupling where the two chains choose to be lazy or non-lazy independently.

Definition 2.2. *For $\gamma \in [0,1]$, let $\mathcal{L}_\gamma(\mathbf{M}) \triangleq (1-\gamma)\mathbf{I} + \gamma\mathbf{M}$ be the lazy Markov chain.*

Definition 2.3 (Lazy coupling)**.** *Let $\mathbf{M}''$ be a coupling of $\mathbf{M}, \mathbf{M}' \in \mathbb{R}^{n \times n}$. For $\gamma, \gamma' \in [0,1]$, define $\mathcal{L}_{\gamma,\gamma'}(\mathbf{M}'') \triangleq (1-\gamma)(1-\gamma')(\mathbf{I} \otimes \mathbf{I}) + (1-\gamma)\gamma'(\mathbf{I} \otimes \mathbf{M}') + \gamma(1-\gamma')(\mathbf{M} \otimes \mathbf{I}) + \gamma\gamma'\mathbf{M}''$. That is, $\mathcal{L}_{\gamma,\gamma'}(\mathbf{M}'')$ is a coupling of $\mathcal{L}_\gamma(\mathbf{M})$ and $\mathcal{L}_{\gamma'}(\mathbf{M}')$.*

Definition 2.4 (Doeblin coupling [32]). *Let $\mathbf{M} \in \mathbb{R}^{n\times n}$ be a stochastic matrix. The Doeblin coupling $\mathcal{Q}(\mathbf{M})$ of two copies of $\mathbf{M}$ is defined as*

$$\mathcal{Q}(\mathbf{M})_{(u,w)(v,x)} \triangleq \begin{cases} (\mathbf{M}\otimes\mathbf{M})_{(u,w)(v,x)} & u \neq w, \\ \mathbf{M}_{uv} & u = w, v = x, \\ 0 & u = w, v \neq x. \end{cases}$$

Using the above definitions, we are able to characterize the expectations in (2.2) in terms of Markov chains. For instance, the first and the second moments $\mathbf{E}_{r,S}\left[X^S_{u,v}\right]$ and $\mathbf{E}_{r,S,S'}\left[X^S_{u,v}X^{S'}_{w,x}\right]$ about forward random walks are characterized by the chains $\mathcal{L}_\gamma\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)$ and $\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)$ respectively, and similar results hold for reversed walks. Hence we reduce the problem of lower bounding (2.1) to the study of these Markov chains.

Notice that matrix $\mathcal{Q}(\mathbf{M})$ agrees with $\mathbf{M}\otimes\mathbf{M}$ except on the rows indexed by (u,u), $u \in V[G]$. This is a manifestation of the fact that the "non-lazy" steps from the same node made by two different forward/reversed random walks are not independent, i.e., every informed node can only send the rumor to one neighbor in each round. Despite this complication, we show that $\mathcal{Q}(\mathbf{M})$ is actually quite close to $\mathbf{M}\otimes\mathbf{M}$:

Lemma 2.5. *Suppose $\mathbf{M} \in \mathbb{R}^{n\times n}$ is a doubly-stochastic matrix with spectral gap $\alpha > 0$, and suppose $\mathbf{M}_{uv} \leqslant \eta$ for any distinct $u, v \in V[G]$. Then for any distribution $\mathbf{u}$ over $V[G]\times V[G]$, $k \in \mathbb{N}$, and $0 \leqslant \gamma \leqslant \min\left\{1/3, \alpha\eta^{-1/2}/9\right\}$, we have*

$$\left\|\mathbf{u}\left(\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})\right)^k - \boldsymbol{\pi}\otimes\boldsymbol{\pi}\right\|_2 \leqslant (1-\gamma\alpha/2)^k + 2\sqrt{2}\gamma\alpha^{-1}n^{-3/2},$$

where $\boldsymbol{\pi}$ denotes the uniform distribution over $V[G]$.

One corollary of Lemma 2.5 states that the stationary distribution of the Markov chain $\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})$ is very close to $\boldsymbol{\pi}\circ\boldsymbol{\pi}$, and its mixing rate is comparable to that of $\mathbf{M}\otimes\mathbf{M}$ (see Corollary D.15). Using the rapid mixing of $\mathcal{L}_\gamma\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)$ and $\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})$ (and similar chains for reversed random walks), we obtain an upper bound of the runtime of Protocol 1, which holds for general graphs with spectral gap α and irregularity β. Our result in this subsection is summarized as follows:

Theorem 2.6. *Suppose G has spectral gap α and irregularity β. Using Protocol 1 with distribution $\mathcal{D} = \mathcal{U}$, with high probability all nodes get the rumor in $T = O(C\log n)$ rounds, where $C = (1/\alpha)\cdot\beta^2\max\{1, 1/(\alpha\cdot\Delta^{0.499})\}$.*

We remark that our analysis above provides a fundamentally new approach to analyze the rumor spreading time of general graphs and, as shown in Theorem 2.6, the result is tight for certain graph families, e.g. $T = O(\log n)$ for any expander graph with n nodes and $\beta = O(1)$.

2.3 A Randomness-Efficient Protocol

The discussion above relates rumor spreading processes to multiple random walks. The transitions of these random walks from a fixed node only depend on local information and are characterized by combinatorial rectangles. Moreover the memoryless feature of random walks/Markov chains allow us to compute them in log-space, or branching programs with polynomial width. Using PRGs for combinatorial rectangles and those for branching programs, we obtain a distribution that is samplable with a short seed and has almost the same performance as the distribution $\mathcal{D} = \mathcal{U}$ in Protocol 1. This gives Protocol 2 that corresponds to Theorem 1.1.

Protocol 2. *Pick the following objects:*

- *an explicit ε-PRG $\mathcal{G} = (\mathcal{G}_0, \ldots, \mathcal{G}_{n-1}) : \{0,1\}^\ell \to [m]^n$ for $\mathsf{CR}_{[m]^n}$ with seed length ℓ, and*

- *an explicit ε'-PRG $\mathcal{G}' = (\mathcal{G}'_0, \ldots, \mathcal{G}'_{T/2-1}) : \{0,1\}^{\ell'} \to \left(\{0,1\}^{\ell}\right)^{T/2}$ for $(T/2, n^2, 2^\ell)$-branching programs with seed length ℓ'*

where $\varepsilon^{-1}, \varepsilon'^{-1}, m = n^{\Theta(1)}$ are sufficiently large.

The initial node having the rumor independently chooses random strings $x, y \in \{0,1\}^{\ell'}$. These random strings are appended with the rumor and sent to other nodes.

- *In the ith round for $0 \leqslant i < T/2$, an informed node u sends the rumor to the neighbor with index $\mathcal{G}_u(\mathcal{G}'_i(x)) \bmod \Delta$ in its adjacency list.*
- *In the ith round for $T/2 \leqslant i < T$, let $j = \lfloor \frac{T-i-1}{2} \rfloor$. For $u \in V[G]$, let $(r_0, r_1) = \mathcal{G}_u(\mathcal{G}'_j(y)) \bmod \Delta^2 \in [\Delta]^2$. Then u sends the rumor to the r_0th neighbor if $i = T-1-2j$, and to the r_1th neighbor if $i = T-2-2j$.*

Setting $C = (1/\alpha) \cdot \beta^2 \max\{1, \alpha^{-1}/\Delta^{0.499}\}$, Protocol 2 uses $2\ell'$ random bits, and with high probability informs all nodes in $T = O(C \log n)$ rounds. As a consequence, we obtain the following reduction:

Theorem 2.7. *Given an explicit ε-PRG for $\mathsf{CR}_{[m]^n}$ with seed length ℓ and an explicit ε'-PRG for $(T/2, n^2, 2^\ell)$-branching programs with seed length ℓ', where $\varepsilon^{-1}, \varepsilon'^{-1}, m = n^{\Theta(1)}$ are sufficiently large, there exists an explicit protocol using $2\ell'$ random bits such that, with high probability all nodes get the rumor in $T = O(C \log n)$ rounds. In particular, given an explicit ε-PRG for (L, W, D)-branching programs with seed length $O(\log n)$ where $L = \max\{T/2, n\}$, $W = n^2$, and $D, \varepsilon^{-1} = n^{\Theta(1)}$ sufficiently large, there exists an explicit protocol using $O(\log n)$ random bits, and with high probability informs all nodes in $T = O(C \log n)$ rounds.*[3]

Combining the reduction above with known explicit constructions of PRGs (Theorem D.5, Theorem D.8), we obtain Theorem 1.1.

Remark 2.8. *We remark here that, by allowing every node to have $O(\Delta)$ preprocessing time before the protocol starts, the rumor spreading time can be improved to $T = O((1/\alpha) \cdot \beta^2 \log n)$, which corresponds to Theorem 1.2. See Section E for formal discussions.*

3 Two-Level Hashing Protocols

In this section we present two protocols. Our protocols are based on pairwise independent generators and unbalanced expanders with near-optimal expansion. Here different rounds use different random bits. In contrast to $O(n \log n)$ random bits per round used in the truly random protocol, we show that $O(\log \log n + \log \Delta)$ random bits per round suffice to spread the rumor efficiently on general graphs G. In contrast to protocols in Section 2, the protocols in this section do not need to assume that nodes have initial IDs, and we can combine the protocols with an ID distribution mechanism so that every node gets a unique ID once it gets the rumor. Formally, in round 0 there is one arbitrary node having the rumor, and the ID of this node is set to be 0. We assume that node 0 knows the maximum degree Δ, and an upper bound $n' \triangleq n^c$ $(c \geqslant 1)$ of the number of nodes n. Moveover, node 0 chooses a binary string, called *seed*, uniformly at random, and the seed is appended to the rumor. In subsequent rounds, whenever one node with ID u sends the rumor to one of its neighbors in round t, it also sends a unique string consisting of the ID u, parameters n', Δ, and current round number t. A node is *uninformed* as long as it has not received a rumor. Once a node receives the first rumor from an informed node with ID u in round t, it becomes *informed* and gets a unique ID defined by $g_t(u) \triangleq 2^{t-1} + u$. If one node becomes informed from multiple informed nodes, then this node chooses an arbitrary node with

[3] This follows from the simple observation that combinatorial rectangles in $[m]^n$ can be computed by $(n, 2, m)$-branching programs.

ID u that informs it and uses $g_t(u)$ as its ID. It was shown in [25] that, through this protocol above, all informed nodes have different IDs, and all the IDs are in $[2^T]$ if the protocol finishes in T rounds.

3.1 Protocol For Graphs with Certain Conductance

Our first protocol in this section corresponds to Theorem 1.3, and holds for graphs with conductance ϕ. Formally, for a graph G of n nodes, the *conductance* $\phi(G)$ of G is defined by

$$\phi(G) \triangleq \min_{S \subseteq V, 0<|S|<n} \frac{e(S, V \setminus S)}{\min\{\mathrm{vol}(S), \mathrm{vol}(V \setminus S)\}},$$

where $\mathrm{vol}(S) \triangleq \sum_{u \in S} \deg(u)$ is the volume of S, and $e(S,T) \triangleq |\{\{u,v\} : u \in S \text{ and } v \in T\}|$ is the number of edges between S and T. The formal description of our protocol is as follows:

Protocol 3 (Protocol for Graphs with Certain Conductance)**.** *Let $\varepsilon = \Delta^{-\Theta(1)}$ be sufficiently small and $m = 2^{\lceil \log(4/\varepsilon) \rceil}$. Pick the following objects:*

- *An explicit $(K, (1-\varepsilon^2/4)D)$-expander $\Gamma : [n^c] \times [D] \to \bigsqcup_{i \in [D]} [M_i]$, where $K = 2$, $D = ((\log n)/\varepsilon)^{O(1)}$ and $M_0 = \cdots = M_{D-1} = M \leqslant D$.*
- *An explicit pairwise independent generator $\mathcal{G} = (\mathcal{G}_1, \ldots, \mathcal{G}_M) : \{0,1\}^\ell \to [m]^M$, where $\ell = O(\log m + \log M) = O(\log\log n + \log \Delta)$.*

These two objects $\mathcal{G}$ and Γ can be uniquely constructed from n^c and $\Delta^{\Theta(1)}$, and hence are known to every informed node.

The initial node having the rumor chooses a random string $(s_1, \ldots, s_T)$ where every s_i is of the form $(x_i, y_i) \in [D] \times \{0,1\}^\ell$. This random string is appended with the rumor and sent to other nodes. Once one node gets the rumor, it gets the ID u. In the ith round, node u computes $r = \Gamma(u, x_i)$ that is in $[M_u]$, the uth copy of $[M]$. Node u computes $y \triangleq \mathcal{G}_r(y_i) \bmod \Delta$, and chooses the neighbor with index y in its adjacency list to send the rumor if $y \leqslant \deg(u)$.

Protocol 3 presents a nice "two-level hashing" framework: The first level is based on a pairwise independent generator $\mathcal{G}$. While the PRG-based protocol in [25] needs to generate $O(n)$ blocks and different nodes need to use different blocks, our protocol only needs $M = (\Delta \log n)^{O(1)}$ blocks and hence $O(\log\log n + \log\Delta)$ random bits suffice for this purpose. The second level uses unbalanced expanders to map the node with ID $u \in [n^c]$ to $r \in [\Delta^{O(1)}]$ by using $O(\log\log n + \log\Delta)$ random bits. After these, node u uses the value of the rth block of $\mathcal{G}$ to choose the neighbors. It is easy to see that every informed node u only needs $O(\mathrm{poly}\log n)$ arithmetic operations per round in order to determine its neighbor.

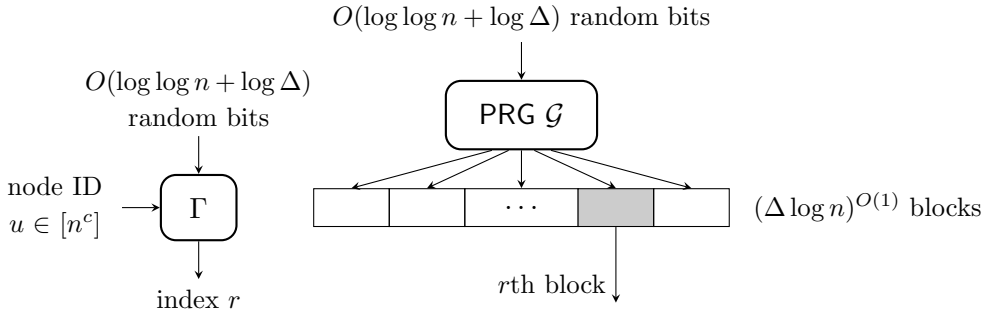

Figure 1: Illustration of the protocol for general graphs. Every node u uses an unbalanced expander Γ to generate an index r, and uses the rth block of PRG $\mathcal{G}$ to choose a neighbor to send the rumor.

Proposition 3.1. *Assume that Protocol 3 finishes in T rounds. Then it uses $O(T \cdot (\log\log n + \log\Delta))$ random bits in total.*

Remark 3.2. *Using the explicit constructions of unbalanced expanders in [27] and pairwise independent generators in [5], our protocol is very simple and can be described as follows: Assign each node with ID $u \in [n^c]$ with a distinct polynomial p_u of degree at most $\lceil c \log_q n \rceil$ over a finite field $\mathbb{F}_q$ of size $q = (\Delta \log n)^{\Theta(1)}$. The protocol then uses the random string $(s_1, \ldots, s_T)$ where every s_i is of the form $(x_i, a_i, b_i) \in \mathbb{F}_q^3$. Then node u computes $z = a_i \cdot p_u(x_i) + b_i$ (over $\mathbb{F}_q$) in the ith round, and chooses the neighbor with index $(z \bmod \deg(u))$ in its adjacency list to send the rumor.*

3.2 Protocol For Strong Expander Graphs

In this subsection we present one protocol for strong expander graphs, and prove Theorem 1.4.

Let $\mathcal{G} = \{G\}_i$ be a family of graphs. We call $\mathcal{G}$ a family of *strong expander graphs* if every G_i in $\mathcal{G}$ has spectra gap $\alpha = 1 - o(1)$, and irregularity $\beta = 1 + o(1)$. This graph family includes several interesting graphs, e.g. Ramanujan graphs, complete graphs, random graphs $G(n, p)$ with $p = \omega(\log n/n)$, and random d-regular graph where d is any increasing function of n. The formal description of our protocol is as follows:

Protocol 4 (Protocol for Strong Expander Graphs)**.** *Let $\varepsilon = \Delta^{-\Theta(1)}$ be sufficiently small, $\varepsilon' = 2^{-\sqrt{\log\log n}}$, and $m = \Theta((\log n)/\varepsilon)$ a power of 2. Pick the following objects:*

- *An explicit $(\leqslant K, (1 - \varepsilon^2/4)D)$-expander $\Gamma : [n^c] \times [D] \to \bigsqcup_{i\in[D]}[M_i]$, where $K = \Delta$, $D = ((\log n)/\varepsilon)^{O(1)}$ and $M_0 = \cdots = M_{D-1} = M \leqslant \max\{D, \Delta^{O(1)}\}$.*
- *An explicit function $\mathcal{G} = (\mathcal{G}_1, \ldots, \mathcal{G}_M) : \{0,1\}^\ell \to [m]^M$ that is both a pairwise independent generator and an ε'-PRG for $\mathsf{CR}_{[m]^M}$, where $\ell = O(\log m + \log M) + \widetilde{O}(\log(1/\varepsilon')) = O(\log\log n + \log\Delta)$.*

These two objects $\mathcal{G}$ and Γ can be uniquely constructed from n^c and $\Delta^{\Theta(1)}$, and hence are known to every informed node.

The initial node having the rumor chooses a random string $(s_1, \ldots, s_T)$ where every s_i is of the form $(x_i, y_i) \in [D] \times \{0,1\}^\ell$. This random string is appended with the rumor and sent to other nodes. Once one node gets the rumor, it gets the ID u. In the ith round, node u computes $r = \Gamma(u, x_i)$ that is in $[M_u]$, the uth copy of $[M]$. It then chooses the neighbor with index $\mathcal{G}_r(y_i) \bmod \deg(u)$ in its adjacency list to send the rumor.

Proposition 3.3. *Assume that Protocol 4 finishes in T rounds. Then it uses $O(T \cdot (\log\log n + \log\Delta))$ random bits in total.*

Acknowledgement. We are grateful to Chris Umans for many hours of stimulating discussion and improving the presentation of the paper. We would like to thank Luca Trevisan and Avi Wigderson for helpful discussion about our work.

References

[1] N. Alon, R. Rubinfeld, S. Vardi, and N. Xie. Space-efficient local computation algorithms. In *23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'12)*, pages 1132–1139, 2012.

[2] R. Armoni, M. Saks, A. Wigderson, and S. Zhou. Discrepancy sets and pseudorandom generators for combinatorial rectangles. In *37th Annual IEEE Symposium on Foundations of Computer Science (FOCS'96)*, pages 412–421, 1996.

[3] N. Berger, C. Borgs, J. T. Chayes, and A. Saberi. On the spread of viruses on the internet. In *16th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'05)*, pages 301–310, 2005.

[4] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory and IEEE/ACM Transactions on Networking*, 52(6): 2508–2530, 2006.

[5] J. Carter and M. N. Wegman. Universal classes of hash functions. *Journal of Computer and System Sciences*, 18(2):143 – 154, 1979.

[6] K. Censor-Hillel and H. Shachnai. Fast information spreading in graphs with large weak conductance. In *43rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'11)*, pages 440–448, 2011.

[7] K. Censor-Hillel, B. Haeupler, J. A. Kelner, and P. Maymounkov. Global computation in a poorly connected world: fast rumor spreading with no dependence on conductance. In *44th Annual ACM Symposium on Theory of Computing (STOC'12)*, pages 961–970, 2012.

[8] F. Chierichetti, S. Lattanzi, and A. Panconesi. Almost tight bounds on rumour spreading by conductance. In *42nd Annual ACM Symposium on Theory of Computing (STOC'10)*, pages 399–408, 2010.

[9] F. R. K. Chung. Spectral graph theory. *Regional Conference Series in Mathematics, American Mathematical Society*, 92:1–212, 1997.

[10] A. E. F. Clementi, P. Crescenzi, C. Doerr, P. Fraigniaud, M. Isopi, A. Panconesi, F. Pasquale, and R. Silvestri. Rumor spreading in random evolving graphs. In *21st Annual European Symposium on Algorithms (ESA'13)*, pages 325–336, 2013.

[11] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry. Epidemic algorithms for replicated database maintenance. In *6th Annual ACM-SIGOPT Principles of Distributed Computing (PODC'87)*, pages 1–12, 1987.

[12] B. Doerr, T. Friedrich, and T. Sauerwald. Quasirandom rumor spreading. In *19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'08)*, pages 773–781, 2008.

[13] B. Doerr, T. Friedrich, and T. Sauerwald. Quasirandom rumor spreading: Expanders, push vs. pull and robustness. In *36th International Colloquium on Automata, Languages, and Programming (ICALP'09)*, pages 366–377, 2009.

[14] B. Doerr, M. Fouz, and T. Friedrich. Social networks spread rumors in sublogarithmic time. In *43rd Annual ACM Symposium on Theory of Computing (STOC'11)*, pages 21–30, 2011.

[15] C. Dutta, G. Pandurangan, R. Rajaraman, Z. Sun, and E. Viola. On the complexity of information spreading in dynamic networks. In *24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'13)*, pages 717–736, 2013.

[16] R. Elsässer and T. Sauerwald. Broadcasting vs. mixing and information dissemination on cayley graphs. In *24th International Symposium on Theoretical Aspects of Computer Science (STACS'07)*, pages 163–174. 2007.

[17] R. Elsässer and T. Sauerwald. On the runtime and robustness of randomized broadcasting. *Theoretical Computer Science*, 410(36):3414–3427, 2009.

[18] U. Feige, D. Peleg, P. Raghavan, and E. Upfal. Randomized broadcast in networks. *Random Structures and Algorithms*, 1(4):447–460, 1990.

[19] N. Fountoulakis and A. Huber. Quasirandom rumor spreading on the complete graph is as fast as randomized rumor spreading. *SIAM Journal on Discrete Mathematics*, 23(4): 1964–1991, 2009.

[20] N. Fountoulakis, K. Panagiotou, and T. Sauerwald. Ultra-fast rumor spreading in social networks. In *23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'12)*, pages 1642–1660, 2012.

[21] T. Friedrich and T. Sauerwald. Near-perfect load balancing by randomized rounding. In *41st Annual ACM Symposium on Theory of Computing (STOC'09)*, pages 121–130, 2009.

[22] T. Friedrich, M. Gairing, and T. Sauerwald. Quasirandom load balancing. *SIAM J. Comput.*, 41(4):747–771, 2012.

[23] G. Giakkoupis. Tight bounds for rumor spreading in graphs of a given conductance. In *28th International Symposium on Theoretical Aspects of Computer Science (STACS'11)*, pages 57–68, 2011.

[24] G. Giakkoupis and P. Woelfel. On the randomness requirements of rumor spreading. In *22nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'11)*, pages 449–461, 2011.

[25] G. Giakkoupis, T. Sauerwald, H. Sun, and P. Woelfel. Low randomness rumor spreading via hashing. In *29th International Symposium on Theoretical Aspects of Computer Science (STACS'12)*, pages 314–325, 2012.

[26] P. Gopalan, R. Meka, O. Reingold, L. Trevisan, and S. Vadhan. Better pseudorandom generators from milder pseudorandom restrictions. In *53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'12)*, pages 120–129, 2012.

[27] V. Guruswami, C. Umans, and S. Vadhan. Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes. *Joural of ACM*, 56(4):20:1–20:34, 2009.

[28] B. Haeupler. Analyzing network coding gossip made easy. In *43rd Annual ACM Symposium on Theory of Computing (STOC'11)*, pages 293–302, 2011.

[29] B. Haeupler. Simple, fast and deterministic gossip and rumor spreading. In *24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'13)*, pages 705–716, 2013.

[30] M. Harchol-Balter, F. T. Leighton, and D. Lewin. Resource discovery in distributed networks. In *18th Annual ACM-SIGOPT Principles of Distributed Computing (PODC'99)*, pages 229–237, 1999.

[31] R. Impagliazzo, N. Nisan, and A. Wigderson. Pseudorandomness for network algorithms. In *26th Annual ACM Symposium on Theory of Computing (STOC'94)*, pages 356–364, 1994.

[32] T. Lindvall. *Lectures on the Coupling Method*. John Wiley & Sons Inc., New York, 2002.

[33] C.-J. Lu. Improved pseudorandom generators for combinatorial rectangles. *Combinatorica*, 22(3):417–434, 2002.

[34] D. Mosk-Aoyama and D. Shah. Fast distributed algorithms for computing separable functions. *IEEE Transactions on Information Theory*, 54(7):2997–3007, 2008.

[35] L. Saloff-Coste. Lectures on finite markov chains. In P. Bernard, editor, *Lectures on Probability Theory and Statistics*, volume 1665 of *Lecture Notes in Mathematics*, pages 301–413. Springer, 1997.

[36] T. Sauerwald and H. Sun. Tight bounds for randomized load balancing on arbitrary network topologies. In *53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'12)*, pages 341–350, 2012.

[37] A. Ta-Shma, C. Umans, and D. Zuckerman. Lossless condensers, unbalanced expanders, and extractors. *Combinatorica*, 27(2):213–240, 2007.

[38] R. van Renesse, Y. Minsky, and M. Hayden. A gossip-style failure detection service. In *15th IFIP Intl. Conf. on Distributed Systems Platforms (Middleware)*, pages 55–70, 1998.

A Notations & Useful Lemmas

In this section we list all notations used in the paper. Let $G = (V, E)$ be a connected, undirected, and simple graph with n nodes. For any node u, $\deg(u)$ stands for the degree of u. The maximum, minimum, and average degree of G are represented by Δ, δ, and d. Let $\beta \triangleq \Delta/\delta$ be the *irregularity* of graph G. The set of neighbors of an node u is represented by $N(u)$. Moreover, for any set $S \subseteq V$, let $N(S) \triangleq \bigcup_{u\in S} N(u)$, and $\mathrm{vol}(S) \triangleq \sum_{u\in S} \deg(u)$. For any set $S, T \subseteq V$, we define $E(S,T) \triangleq \{\{u,v\} : u \in S \text{ and } v \in T\}$ and $e(S,T) \triangleq |E(S,T)|$.

We use $\mathbf{A}_G$ to express the adjacency matrix of G. Let $\mathbf{D}$ the $n \times n$ diagonal matrix defined by $\mathbf{D}_{uu} = \deg(u)$ for $u \in V[G]$. Let $\mathbf{M}_G = \mathbf{D}^{-1}\mathbf{A}_G$ be the transition matrix for the random walk over G, and $\mathbf{N}_G \triangleq \mathbf{D}^{-1/2}\mathbf{A}_G\mathbf{D}^{-1/2}$. Define the n real eigenvalues of $\mathbf{N}_G$ by $1 = \lambda_1 \geqslant \cdots \geqslant \lambda_n \geqslant -1$, and let $\lambda_{\max} \triangleq \max\{\lambda_2, |\lambda_i|\}$. The spectral gap α is defined by $\alpha \triangleq 1 - \lambda_2$, whereas the absolute spectral gap is defined as $1 - \lambda_{\max}$. For simplicity, we also use α to express the spectral expansion of a reversible Markov chain if the chain is clear from the context.

For $m \in \mathbb{N}$, vector $\mathbf{u} \in \mathbb{R}^m$ and real number $p \geqslant 1$, define the ℓ_p-norm $\|\mathbf{u}\|_p = (\sum_{i=1}^m |\mathbf{u}_i|^p)^{1/p}$. In addition, we define $\|\mathbf{u}\|_\infty = \max_{1\leqslant i\leqslant m} |\mathbf{u}_i|$. The inner product of two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^m$ is $\langle \mathbf{u}, \mathbf{v}\rangle = \sum_{i=1}^m \mathbf{u}_i\mathbf{v}_i$. We write $\mathbf{1}_m$ for the vector in $\mathbb{R}^m$ having ones in all entries, or simply $\mathbf{1}$ if the dimension is clear from the context. Similarly write $\mathbf{0}_m$ or $\mathbf{0}$ for the zero vector. Let $\mathbf{e}_i$ be the vector that has an one in the ith entry and zero elsewhere. Write $\mathbf{I}_m$ or $\mathbf{I}$ for the $m \times m$ identity matrix. For a matrix $\mathbf{M} \in \mathbb{R}^{m\times m'}$, we use $\mathbf{M}_{ij}$ to denote the entry on $\mathbf{M}$'s ith row and jth column. For $p \in [1,\infty) \cup \{\infty\}$, define

$$\|\mathbf{M}\|_p = \sup_{\mathbf{u}\in\mathbb{R}^m\setminus\{\mathbf{0}\}} \frac{\|\mathbf{u}\mathbf{M}\|_p}{\|\mathbf{u}\|_p}.$$

It is easy to show that $\|\mathbf{M}\|_1$ equals the maximum of the ℓ_1-norms of the rows of $\mathbf{M}$. And $\|\mathbf{M}\|_\infty$ equals the maximum of the ℓ_1-norms of the columns of $\mathbf{M}$, or equivalently $\|\mathbf{M}^\intercal\|_1$. We say a square matrix $\mathbf{M}$ is stochastic if all of its entries are non-negative and all of its rows have ℓ_1-norm 1. Clearly if $\mathbf{M}$ is stochastic, then $\|\mathbf{M}\|_1 = 1$. We say $\mathbf{M}$ is doubly-stochastic if both $\mathbf{M}$ and $\mathbf{M}^\intercal$ are stochastic.

By $\log x$ we denote the binary logarithm of x. For any integer m, define $[m] \triangleq \{0, \ldots, m-1\}$. The disjoint union of a family of sets $\{A_i : i \in I\}$ indexed by I is denoted by $\bigsqcup_{i\in I} A_i \triangleq \bigcup_{i\in I}\{(x,i) : x \in A_i\}$. With high probability stands for with probability $1 - n^{-\Theta(1)}$.

Lemma A.1. *Fix any $0 < p < 1$ and let $X_1, \ldots, X_n$ be independent geometric random variables on $\mathbb{N}$ with $\mathbf{Pr}[X_i = k] = (1-p)^{k-1}p$ for every $k \in \mathbb{N}$. Let $X = \sum_{i=1}^n X_i$, and $\mu = \mathbf{E}[X]$. Then it holds for all $\beta > 0$ that*

$$\mathbf{Pr}[X \geqslant (1+\beta)\mu] \leqslant \mathrm{e}^{-n\beta^2/(2(1+\beta))}.$$

Fact A.2 ([35])**.** *The spectral gap of a graph G satisfies*

$$\alpha = \inf_{\mathbf{u}\not\parallel\mathbf{1}} \frac{\mathcal{E}_{\boldsymbol{\pi},G}(\mathbf{u},\mathbf{u})}{\mathrm{Var}_{\boldsymbol{\pi}}(\mathbf{u})}$$

where π is the stationary distribution of $\mathbf{M}_G$, and the quantities

$$\mathrm{Var}_{\boldsymbol{\pi}}(\mathbf{u}) = \frac{1}{2}\sum_{u,v\in V[G]} \boldsymbol{\pi}_u\boldsymbol{\pi}_v(\mathbf{u}_u - \mathbf{u}_v)^2, \qquad \mathcal{E}_{\boldsymbol{\pi},G}(\mathbf{u},\mathbf{u}) = \frac{1}{2}\sum_{u,v\in V[G]} \boldsymbol{\pi}_u(\mathbf{M}_G)_{uv}(\mathbf{u}_u - \mathbf{u}_v)^2$$

are known as the global variance and the local variance (or Dirichlet form) of $\mathbf{u}$ respectively.

We also need an operation on graphs, called *regularization*. Formally speaking, for an undirected graph G with maximal degree Δ, let $\mathrm{Reg}(G)$ be the regular graph obtained from G by adding $\Delta - \deg(u)$ self-loops to each node $u \in V[G]$.

Lemma A.3. *Suppose graph G has spectral gap α and irregularity β. Then* $\mathrm{Reg}(G)$ *has spectral gap at least $\beta^{-2}\alpha$.*

Proof. Let $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ be the stationary distributions of $\mathbf{M}_G$ and $\mathbf{M}_{\mathrm{Reg}(G)}$ respectively, i.e. $\boldsymbol{\pi}_u = \deg(u)/(n \cdot d)$ and $\boldsymbol{\pi}'_u = 1/n$ for any node $u \in V$. Then for any $\mathbf{u} \not\parallel \mathbf{1}$, we have

$$\frac{\mathcal{E}_{\boldsymbol{\pi}',\mathrm{Reg}(G)}(\mathbf{u},\mathbf{u})}{\mathcal{E}_{\boldsymbol{\pi},G}(\mathbf{u},\mathbf{u})} \geqslant \min_{u \neq v} \frac{\boldsymbol{\pi}'_u(\mathbf{M}_{\mathrm{Reg}(G)})_{uv}}{\boldsymbol{\pi}_u(\mathbf{M}_G)_{uv}} = \min_{u \in V[G]} \frac{d \cdot \Delta^{-1}}{\deg(u) \cdot (\deg(u))^{-1}} = d/\Delta$$

and

$$\frac{\mathrm{Var}_{\boldsymbol{\pi}'}(\mathbf{u})}{\mathrm{Var}_{\boldsymbol{\pi}}(\mathbf{u})} \leqslant \min_{u \neq v} \frac{\boldsymbol{\pi}'_u \boldsymbol{\pi}'_v}{\boldsymbol{\pi}_u \boldsymbol{\pi}_v} \leqslant \min_{u \neq v} \frac{(1/n) \cdot (1/n)}{(\deg(u)/nd) \cdot (\deg(v)/nd)} = \min_{u \neq v} \frac{d^2}{\deg(u)\deg(v)} \leqslant d^2/\delta^2.$$

So

$$\frac{\mathcal{E}_{\boldsymbol{\pi}',\mathrm{Reg}(G)}(\mathbf{u},\mathbf{u})}{\mathrm{Var}_{\boldsymbol{\pi}'}(\mathbf{u})} \Big/ \frac{\mathcal{E}_{\boldsymbol{\pi},G}(\mathbf{u},\mathbf{u})}{\mathrm{Var}_{\boldsymbol{\pi}}(\mathbf{u})} \geqslant (d/\Delta) \cdot (\delta^2/d^2) \geqslant \beta^{-2}$$

and the claim follows from Fact A.2. ∎

B Existential Proof

In this section we show that $O(\log n)$ random bits are sufficient in rumor spreading for many classes of graphs (e.g. complete graphs, strong expanders, graphs with good conductance, etc.) if we do not care about the computational complexity. We will prove the following general statement:

Lemma B.1. *Let $\mathcal{C}$ be a class of graphs on n nodes with no multi-edges. Let $T' = n^{O(1)}$ be an upper bound of spreading time. Suppose the spreading time for any graph in $\mathcal{C}$ is at most T with probability p for fully-random push protocol. Then there exists a (non-explicit) function*

$$f : \{0,1\}^\ell \times [n] \times [T'] \times [\Delta] \to [\Delta]$$

such that

1. *$f(x,u,t,d) \in [d]$ for all $(x,u,t,d) \in \{0,1\}^\ell \times [n] \times [T'] \times [\Delta]$.*
2. *$\ell = \max\{\log\log|C|, \log n + \log\Delta + \log\log\Delta\} + 2\log(1/\varepsilon) + O(1)$.*
3. *for x uniformly chosen from $\{0,1\}^\ell$, the spreading time for any graph $G \in \mathcal{C}$ is at most T with probability $p - \varepsilon$ if node u uses $f(x,u,t,\deg(u)) \in [\deg(u)]$ as the index of its receiver in its adjacency list in round t.*

In particular, ℓ is bounded by $2\log n + \log\log n + 2\log(1/\varepsilon) + O(1)$ since $|\mathcal{C}| \leqslant 2^{n^2}$ and $\Delta \leqslant n$.

Proof. Choose $f(x,u,t,d) \in [d]$ independently and uniformly at random for each $(x,u,t,d) \in \{0,1\}^\ell \times [n] \times [T'] \times [\Delta]$. Fix a graph $G \in \mathcal{C}$ and an initial node in $[n]$. For each node u in the graph of degree $\deg(u)$, there are $\deg(u)!$ possible orders of neighbors of u in its adjacency list. We also fix the order for each node u. Observe that for any fixed x, the random variables $f(x,u,t,\deg(u))$ for all pairs (u,t) are independent and uniformly distributed. Let $I(x)$ be the indicator random variable that equals 1 if the spreading time of G is at most T when node u uses $f(x,u,t,\deg(u))$ to decide its receiver in round t. Then $\mathbf{Pr}_f[I(x) = 1] \geqslant p$ for any x and hence $\mathbf{E}_f[I(x)] \geqslant p$. Also note that $I(x)$'s are independent. By the Chernoff bound it holds that

$$\mathbf{Pr}_f\left[\left|2^{-\ell}\sum_x I(x) - 2^{-\ell}\sum_x \mathbf{E}_f[I(x)]\right| \geqslant \varepsilon\right] \leqslant 2\exp(-2^\ell\varepsilon^2/4).$$

So with probability at least $1-2\exp(-2^\ell\varepsilon^2/4)$, we have $\mathbf{E}_x[\,I(x)\,] \geqslant \mathbf{E}_x[\,\mathbf{E}_f[\,I(x)\,]\,]-\varepsilon \geqslant p-\varepsilon$. By the union bound, the probability that $\mathbf{E}_x[\,I(x)\,] \geqslant p-\varepsilon$ holds for all graphs in $\mathcal{C}$, arbitrary neighboring list of nodes, and all start nodes is at least

$$1 - n|\mathcal{C}| \cdot (\Delta!)^n \cdot 2\exp(-2^\ell\varepsilon^2/4),$$

which is greater than zero for sufficiently large $\ell = \max\{\log\log|C|, \log n + \log\Delta + \log\log\Delta\} + 2\log(1/\varepsilon) + O(1)$. So there exists one function f such that $\mathbf{E}_x[\,I(x)\,] \geqslant p-\varepsilon$ holds for all graphs in $\mathcal{C}$, i.e. the spreading time for any graph $G \in \mathcal{C}$ is at most T with probability $p-\varepsilon$ over the choices of x, if node u uses $f(x,u,t,\deg(u)) \in [\deg(u)]$ to choose its receiver in round t. ■

The same result also holds for pull protocols and push-pull protocols, and can be shown using similar arguments.

The following result follows from Lemma B.1 directly.

Corollary B.2 (Existential Result)**.** *Let $\mathcal{G} = \{G_n\}_{n\geqslant 1}$ be a family of graphs such that for any $G_n \in \mathcal{G}$ with n nodes the truly random protocol finishes in $T = n^{O(1)}$ rounds with high probability. Then there is a protocol which finishes in T rounds with high probability and uses $3\log n$ random bits in total.*

C Lower Bounds on Randomness Complexity

We address the randomness requirement of rumor spreading protocols. We first introduce the pull model, which is a symmetric version of the push model, and the formal description is as follows: In round $t \geqslant 0$, every node u that does not yet have the rumor selects a neighbor v uniformly at random and asks for the rumor, and gets the rumor if v received the rumor before. In the push-pull model, in every round t, every node u chooses a random neighbor to perform *push* if node u has the rumor, or perform *pull* if u has not received the rumor.

We prove the following lower bound on the number of random bits needed for any protocol in the push-pull model:

Theorem C.1. *Let G be any graph with n nodes and sufficiently large minimum degree $\delta = \Omega(\log n)$. Then any protocol in the push-pull model that is oblivious of the order of adjacency lists of G and informs at least half of the nodes of G in T rounds with nonzero probability has to use more than $\log\delta - \log T - 2$ random bits. In particular, $\Theta(\log n)$ random bits are necessary when $\delta = \Theta(n)$ and $T = O(n^{1-\varepsilon})$ for some constant $\varepsilon > 0$.*

Here we even allow the protocol access to the ID of the initial node and the structure of G, i.e., the sets of neighbors of nodes as *unordered sets*. In addition, we allow each node access to the randomness even before it obtains the rumor. All we assume is that the protocol is oblivious of the *order* of the adjacency lists.

Proof. Suppose $V[G] = [n]$. Let Δ be the maximum degree of G and s be the initial node. We first claim that there exists a subset of nodes S of size $n/2$ (for simplicity assume n is even) such that $\deg(u)/4 \leqslant |S \cap N(u)| \leqslant 3\deg(u)/4$ for all $u \in [n]$: If we pick a random subset S of size $n/2$, then for any fixed u the condition $\deg(u)/4 \leqslant |S \cap N(u)| \leqslant 3\deg(u)/4$ holds, by the Chernoff bound, with probability at least $1 - \mathrm{e}^{-\Theta(\delta)} > 1 - 1/n$ for $\delta = \Omega(\log n)$ sufficiently large. The claim then follows by taking the union bound. Pick such a subset S with the claimed property. Note that $[n] \setminus S$ has the same property. We may therefore assume $s \in S$ by swapping S and $[n] \setminus S$ if necessary.

A protocol for G using ℓ random bits in T rounds is uniquely characterized by a pair of functions

$$f_1, f_2 : \{0,1\}^\ell \times [n] \times [T] \times [\Delta] \to [\Delta]$$

satisfying $f_1(x,u,t,d), f_2(x,u,t,d) \in [d]$ for all $(x,u,t,d) \in \{0,1\}^\ell \times [n] \times [T] \times [\Delta]$, in the sense that given the random string x, node u chooses a neighbor with index $f_1(x,u,t,\deg(u))$ (resp. $f_2(x,u,t,\deg(u))$) in its adjacency list to push (resp. pull) the message in round t if it is informed (resp. uninformed). For each $u \in [n]$, define $I_u \subseteq [n]$ as

$$I_u = \begin{cases} \{f_1(x,u,t,\deg(u)) : x \in \{0,1\}^\ell, t \in [T]\} & u \in S \\ \{f_2(x,u,t,\deg(u)) : x \in \{0,1\}^\ell, t \in [T]\} & u \notin S. \end{cases}$$

Assume to the contrary that $\ell \leqslant \log\delta - \log T - 2$. Then the size of I_u is at most $2^\ell \cdot T \leqslant \delta/4 \leqslant \min\{|S \cap N(u)|, |([n] \setminus S) \cap N(u)|\}$ for each $u \in [n]$. So it is possible to order the adjacency list of each $u \in [n]$ such that the neighbors picked by u using index set I_u are all in $S \cap N(u)$ if $u \in S$, or in $([n] \setminus S) \cap N(u)$ if $u \in [n] \setminus S$. Then in the rumor spreading process, nodes in S push messages only to those also in S, and nodes in $[n] \setminus S$ pull messages only from those also in $[n] \setminus S$. As $s \in S$, the nodes in $[n] \setminus S$ never get informed. ■

For the push model and the pull model we may drop the assumption that $\delta = \Omega(\log n)$ is sufficiently large, and also simplify the proof.

Theorem C.2. *Let G be any graph with n nodes. Then any protocol in the push model that is oblivious of the order of adjacency lists of G and informs all the nodes of G in T rounds with nonzero probability has to use more than $\log(\delta-1) - \log T$ random bits.*

Proof. The protocol is now characterized by a single function f_1 describing how rumors are pushed. Define $I_u = \{f_1(x,u,t,\deg(u)) : x \in \{0,1\}^\ell, t \in [T]\}$ for each $u \in [n]$. Pick $v \in [n] \setminus \{s\}$. Assume to the contrary that $\ell \leqslant \log(\delta-1) - \log T$. Then the size of I_u is at most $2^\ell \cdot T \leqslant \delta - 1 \leqslant |N(u) \setminus \{v\}|$ for each $u \in [n]$. So it is possible to order the adjacency list of each $u \in [n]$ such that the neighbors picked by u using index set I_u are all in $N(u) \setminus \{v\}$. Then the node v never gets informed. ■

Theorem C.3. *Let G be any graph with n nodes. Then any protocol in the pull model that is oblivious of the order of adjacency lists of G and informs more than one node of G in T rounds with nonzero probability has to use more than $\log(\delta-1) - \log T$ random bits.*

Proof. The protocol is now characterized by a single function f_2 describing how rumors are pulled. Define $I_u = \{f_2(x,u,t,\deg(u)) : x \in \{0,1\}^\ell, t \in [T]\}$ for each $u \in [n]$. Assume to the contrary that $\ell \leqslant \log(\delta-1) - \log T$. Then the size of I_u is at most $2^\ell \cdot T \leqslant \delta - 1 \leqslant |N(u) \setminus \{s\}|$ for each $u \in [n]$. So it is possible to order the adjacency list of each $u \in [n]$ such that the neighbors picked by u using index set I_u are all in $N(u) \setminus \{s\}$. Then the nodes in $[n] \setminus \{s\}$ never get informed. ■

D Omitted Details in Section 2

D.1 Preliminaries

We first list definitions and results about pseudorandom generators.

Pairwise Independent Generators.

Definition D.1 (Pairwise Independent Generator)**.** *We say $X_0, \ldots, X_{d-1}$ with X_i distributed over $[m_i]$ are ε-pairwise independent if*

- $\left|\mathbf{Pr}[\, X_i = x \,] - \frac{1}{m_i}\right| \leqslant \varepsilon$ *for all $i \in [d]$ and $x \in [m_i]$, and*
- $\left|\mathbf{Pr}[\, X_i = x \wedge X_j = x' \,] - \frac{1}{m_i \cdot m_j}\right| \leqslant \varepsilon$ *for all distinct $i, j \in [d]$ and all $x \in [m_i]$, $x' \in [m_j]$.*

We say they are pairwise independent if $\varepsilon = 0$. We say $\mathcal{G} : \{0,1\}^\ell \to [m_0] \times \cdots \times [m_{d-1}]$ is an (ε-)pairwise independent generator if its outputs are (ε-)pairwise independent given a uniformly distributed seed.

Theorem D.2 ([5])**.** *There exists an explicit pairwise independent generator $\mathcal{G} : \{0,1\}^\ell \to [m]^d$ with seed length $\ell = O(\log m + \log d)$.*

Lemma D.3. *Suppose $\mathcal{G} = (\mathcal{G}_0, \ldots, \mathcal{G}_{d-1})$ is a pairwise independent generator where $\mathcal{G}_i : \{0,1\}^\ell \to [m]$. Define $\mathcal{G}' = (\mathcal{G}'_0, \ldots, \mathcal{G}'_{d-1})$ where $\mathcal{G}'_i(x) = \mathcal{G}_i(x) \bmod m_i$ for $i \in [d]$. Then $\mathcal{G}' : \{0,1\}^\ell \to [m_0] \times \cdots \times [m_{d-1}]$ is an ε-pairwise independent generator where $\varepsilon = 2/m$.*

Proof. For distinct $i, j \in [d]$ and $x \in [m_i]$, $x' \in [m_j]$, let B (resp. B') be the preimages of x (resp. x') under the map $s \mapsto s \bmod m_i$ (resp. $s \mapsto s \bmod m_j$). Then $||B| - m/m_i| \leqslant 1$ and $||B'| - m/m_j| \leqslant 1$. So $\mathbf{Pr}_s[\mathcal{G}'_i(s) = x] = |B|/m$ which differs from $1/m_i$ by at most $1/m$. Similarly $\mathbf{Pr}_s[\mathcal{G}'_i(s) = x \wedge \mathcal{G}'_j(s) = x'] = |B||B'|/m^2$ which differs from $1/(m_i m_j)$ by at most $2/m$. ∎

Lemma D.4. *Suppose $\mathcal{G} = (\mathcal{G}_0, \ldots, \mathcal{G}_{d-1})$ is an ε-PRG for CR_S where $S = [m]^d$. Define $\mathcal{G}' = (\mathcal{G}'_0, \ldots, \mathcal{G}'_{d'-1})$ where $\mathcal{G}'_j(x) = \mathcal{G}_{i_j}(x) \bmod m_j$ for $j \in [d']$ and $i_0, \ldots, i_{d'-1} \in [d]$. Then $\mathcal{G}'$ is an $(\varepsilon + \sum_{i \in [d']} m_i/m)$-PRG for $\mathsf{CR}_{S'}$ where $S' = \prod_{i \in [d']} [m_i]$.*

Proof. By definition, $\mathcal{G}' = \pi \circ \mathcal{G}$ with $\pi : (x_0, \ldots, x_{d-1}) \mapsto (x_{i_0} \bmod m_0, \ldots, x_{i_{d'-1}} \bmod m_{d'-1})$. For $A = \prod_{i \in [d']} A_i \in \mathsf{CR}_{S'}$, let $B = \pi^{-1}(A) = \prod_{i \in [d]} B_i \in \mathsf{CR}_S$. Then $\mathbf{Pr}_s[\mathcal{G}'(s) \in A] = \mathbf{Pr}_s[\mathcal{G}_i(s) \in B]$ which differs from $|B|/|S|$ by at most ε since $\mathcal{G}$ is an ε-PRG for CR_S. Note that $|B_{i_j}|/m$ differs from $|A_j|/m_j$ by at most m_j/m for $j \in [d']$, and $B_i = [m]$ for $i \in [d] \setminus \{i_0, \ldots, i_{d'-1}\}$. A simple induction shows that $|B|/|S|$ differs from $|A|/|S'|$ by at most

$$\sum_{j \in [d']} ||B_{i_j}|/m - |A_j|/m_j| \leqslant \sum_{i \in [d']} m_i/m.$$ ∎

Theorem D.5 ([26])**.** *Let $S = [m]^d$. There exists an explicit ε-PRG for CR_S with seed length $O(\log m + \log d) + O(\log(1/\varepsilon))$.* [4]

Lemma D.6. *There exists an explicit function $\mathcal{G} : \{0,1\}^\ell \to [m]^d$ that is both a pairwise independent generator and an ε-PRG for $\mathsf{CR}_{[m]^d}$ with seed length $O(\log m + \log d) + \widetilde{O}(\log(1/\varepsilon))$.*

Proof. Pick an explicit pairwise independent generator $\mathcal{G}^\flat : \{0,1\}^{\ell_1} \to [m]^d$ with see length $\ell_1 = O(\log m + \log d)$ and an explicit ε-PRG $\mathcal{G}^\sharp : \{0,1\}^{\ell_2} \to [m]^d$ for $\mathsf{CR}_{[m]^d}$ with seed length $\ell_2 = O(\log m + \log d) + \widetilde{O}(\log(1/\varepsilon))$. Identify $[m]^d$ with $\mathbb{Z}_m^d$ and define $\mathcal{G} : \{0,1\}^{\ell_1 + \ell_2} \to [m]^d$ using addition in $\mathbb{Z}_m^d$: $\mathcal{G}(x, y) = \mathcal{G}^\flat(x) + \mathcal{G}^\sharp(y)$. The definition of pairwise independent generators implies that the function $x \mapsto \mathcal{G}^\flat(x) + z$ is a pairwise independent generator for any fixed $z \in [m]^d$, i.e., the property is preserved under addition of any fixed element in $\mathbb{Z}_m^d$. Then the same is true for random $z = \mathcal{G}^\sharp(y)$. So $\mathcal{G}$ is a pairwise independent generator. A similar argument shows that it is also an ε-PRG for $\mathsf{CR}_{[m]^d}$. ∎

Definition D.7 (Branching Programs)**.** *A branching program of length L, width W and degree D, or an (L, W, D)-branching program, is a directed (multi)-graph with node set $[W] \times \{0, \ldots, L\}$. We say the nodes in $[W] \times \{i\}$ are on the ith layer for $0 \leqslant i \leqslant L$. Each node (u, i) except those on the last layer has D outgoing edges to nodes on the next layer, and these D edges are associated with D distinct labels from $[D]$.*

Theorem D.8 ([31])**.** *There exists an explicit ε-PRG for (L, W, D)-branching programs with seed length $O(\log L(\log W + \log L + \log(1/\varepsilon)) + \log D)$.*

[4] In [26] the seed length is presented as $O((\log\log m)(\log m + \log d + \log(1/\varepsilon))) + \widetilde{O}(\log(1/\varepsilon))$. But there are techniques of reducing m and d to $m' = (1/\varepsilon)^{O(1)}, d' = (1/\varepsilon)^{O(1)}$ using $O(\log m + \log d)$ randomness, cf. [2, 33].

The following lemma about Markov chains will be used in the analysis. For an ergodic Markov chain represented by the stochastic matrix $\mathbf{M}$ and $\varepsilon > 0$, define its ℓ_2-mixing time as

$$\tau_{\mathbf{M}}(\varepsilon) = \max_{\mathbf{u}} \min\{k : \|\mathbf{u}\mathbf{M}^k - \boldsymbol{\pi}\|_2 \leqslant \varepsilon\},$$

where $\boldsymbol{\pi}$ is the stationary distribution of $\mathbf{M}$ and $\mathbf{u}$ ranges over all distributions over the state set of the chain.

Lemma D.9 ([35])**.** *Suppose $\mathbf{M} \in \mathbb{R}^{V[G]\times V[G]}$ represents a reversible Markov chain with absolute spectral gap $\alpha > 0$. Then $\tau_{\mathbf{M}}(\varepsilon) < \log_{1-\alpha} \varepsilon + 1$.*

D.2 Analysis of the Prototype Protocol

In this subsection we give the detailed analysis of Protocol 1. We start with the formal definition of reversed random walks. The basic idea is to view a push operation (or one step of a forward walk) as a pull operation (or one step of a reversed walk). However, there are several complications: (1) we let v "pull the from u" only when u is the unique node pushing to v, since v is not allowed to pull from multiple nodes at the same time; (2) we need to use auxiliary randomness $r_{i,u}$ to equalize the probabilities of successful pulls made by different nodes;[5] (3) we want the pull operations to be pairwise independent. In particular two nodes u and v pull from their common neighbor w at the same time with probability $1/\Delta^2$. To realize this, we combine two rounds into one so that w can send two messages, say to a and b at the same time. Also, note that there are two cases when w pushes to both u and v, or equivalently u and v both pull from w: $(a, b) = (u, v)$ or $(a, b) = (v, u)$. We admit only one of them, so that the event occurs with probability $1/\Delta^2$ rather than $2/\Delta^2$.

As before, for $f \sim \mathcal{D}$, we denote by $\widetilde{f}(i, u)$ the $f(i, u)$th neighbor of u in its adjacency list.

Definition D.10 (Reversed random walks)**.** *Consider a random rumor spreading process in T rounds on a graph G using Protocol 1 determined by its own randomness $f \sim \mathcal{D} = \mathcal{U}$. Pick real numbers $r_{i,u}$ independently and uniformly from $[0, 1]$ for all $i \in [T/2]$ and $u \in V[G]$.*

Fix an arbitrary total order $\preceq$ on $V[G]$. For $i \in [T/2]$ and $u \in V[G]$, define

$$N_{i,u} = \begin{cases} \left\{\widetilde{f}(T-1-2i, u), \widetilde{f}(T-2-2i, u)\right\} & \widetilde{f}(T-1-2i, u) \preceq \widetilde{f}(T-2-2i, u) \\ \emptyset & \text{otherwise} \end{cases}$$

and define $N^{\vee}_{i,u} = \{v \in V[G] : v \neq u \text{ and } u \in N_{i,v}\}$.

A reversed random walk of length $k \in [T/2]$ with pattern $S = (s_0, \ldots, s_{k-1}) \in \mathcal{C}_k$ is a sequence of $k + 1$ nodes $(p_0, \ldots, p_k)$ of G, such that for all $i \in [k]$: (i) if $s_i =$ lazy, then $p_{i+1} = p_i$; (ii) if $s_i =$ non-lazy, then $p_{i+1} = u$ if $N^{\vee}_{i,p_i} = \{u\}$ is a singleton and $r_{i,p_i} \leqslant (1 - 1/\Delta)^{\Delta - \deg(u)}$, and otherwise $p_{i+1} = p_i$.

D.2.1 Approximation via Random Walks

We elaborate the idea of bounding runtime of Protocol 1 with respect to multiple random walks. We will use three distributions in the following analysis:

- $\mathcal{D}_{\gamma,k}$ is the distribution over $\mathcal{C}_k$ where entries are independently chosen to be lazy with probability $1 - \gamma$.
- Let $r = (f, \{r_{i,u}\})$ be the whole randomness used in Definition 2.1 and Definition D.10 which determines the random walks. Let $\widetilde{\mathcal{D}}$ be the distribution of r, which is the product of $\mathcal{D}$ with copies of uniform distributions over $[0, 1]$.

[5] The auxiliary randomness only appears in the analysis.

The following lemma give a lower bound of the probability that a node w gets informed in T rounds with respect to multiple random walks.

Lemma D.11. *For Protocol 1 with $\mathcal{D} = \mathcal{U}$ and initial node s, any $0 \leqslant k \leqslant T/4$, and $\gamma \in (0,1)$, a node w is informed in T rounds with probability at least*

$$\frac{\sum_{u,v\in V[G]} \mathbf{E}_{r,S}\left[X^S_{s,u}\right]\mathbf{E}_{r,S}\left[X^S_{s,v}\right]\mathbf{E}_{r,S}\left[Y^S_{w,u}\right]\mathbf{E}_{r,S}\left[Y^S_{w,v}\right]}{\sum_{u,v\in V[G]} \mathbf{E}_{r,S,S'}\left[X^S_{s,u}X^{S'}_{s,v}\right]\mathbf{E}_{r,S,S'}\left[Y^S_{w,u}Y^{S'}_{w,v}\right]} \tag{D.1}$$

where r, S and S' are independent with distributions $\widetilde{\mathcal{D}}$, $\mathcal{D}_{\gamma,k}$ and $\mathcal{D}_{\gamma,k}$ respectively.

Proof. Define the weight of forward or reversed random walks with pattern $S = (s_0, \ldots, s_k)$ as $\mathrm{wt}(S) := (1-\gamma)^{n_1}\gamma^{n_2} > 0$ where n_1 and n_2 are the number of lazy and non-lazy s_i respectively. Let $\widetilde{X}^S_{u,v} = X^S_{u,v} \cdot \mathrm{wt}(S)$ and $\widetilde{Y}^S_{u,v} = Y^S_{u,v} \cdot \mathrm{wt}(S)$.

Suppose $s \in V[G]$ is the initial node and also fix $w \in V[G]$. If there exist a forward walk p from s to some node u and a reversed walk p' from w to u, then the rumor is sent from s to u following p and then from u to w following the reversal of p'. Also note that the two walks exist if and only if $\widetilde{X}^S_{s,u}\widetilde{Y}^{S'}_{w,u} > 0$ for some S, S' and u. Therefore,

$$\mathbf{Pr}_{f\sim\mathcal{D}}\left[\,t \text{ receives the message}\,\right] \geqslant \mathbf{Pr}_{r\sim\widetilde{\mathcal{D}}}\left[\sum_{S,S'\in\mathcal{C}_k,u\in V[G]} \widetilde{X}^S_{s,u}\widetilde{Y}^{S'}_{w,u} > 0\right]. \tag{D.2}$$

Furthermore,

$$\begin{aligned}
&\mathbf{Pr}_{r\sim\widetilde{\mathcal{D}}}\left[\sum_{S,S'\in\mathcal{C}_k,u\in V[G]} \widetilde{X}^S_{s,u}\widetilde{Y}^{S'}_{w,u} > 0\right] \\
&= \mathbf{E}_{r\sim\widetilde{\mathcal{D}}}\left[\mathbf{1}_{\sum_{S,S'\in\mathcal{C}_k,u\in V[G]}\widetilde{X}^S_{s,u}\widetilde{Y}^{S'}_{w,u}>0}\right] \\
&\geqslant \frac{\left(\mathbf{E}_{r\sim\widetilde{\mathcal{D}}}\left[\sum_{S,S'\in\mathcal{C}_k,u\in V[G]}\widetilde{X}^S_{s,u}\widetilde{Y}^{S'}_{w,u}\right]\right)^2}{\mathbf{E}_{r\sim\widetilde{\mathcal{D}}}\left[\left(\sum_{S,S'\in\mathcal{C}_k,u\in V[G]}\widetilde{X}^S_{s,u}\widetilde{Y}^{S'}_{w,u}\right)^2\right]} \\
&= \frac{\sum_{u,v\in V[G]} \mathbf{E}_{r,S}\left[X^S_{s,u}\right]\mathbf{E}_{r,S}\left[X^S_{s,v}\right]\mathbf{E}_{r,S}\left[Y^S_{w,u}\right]\mathbf{E}_{r,S}\left[Y^S_{w,v}\right]}{\sum_{u,v\in V[G]} \mathbf{E}_{r,S,S'}\left[X^S_{s,u}X^{S'}_{s,v}\right]\mathbf{E}_{r,S,S'}\left[Y^S_{w,u}Y^{S'}_{w,v}\right]}
\end{aligned} \tag{D.3}$$

where the subscripts r, S and S' are independent with distributions $\widetilde{\mathcal{D}}$, $\mathcal{D}_{\gamma,k}$ and $\mathcal{D}_{\gamma,k}$ respectively. The first inequality is an instance of the Cauchy-Schwarz inequality. The last equality uses the independence of $X^S_{s,u}X^{S'}_{s,v}$ and $Y^S_{w,u'}Y^{S'}_{w,v'}$ for any $u, v, u', v' \in V[G]$ as well as the fact that the weight $\mathrm{wt}(S)$ is just the probability of S in $\mathcal{D}_{\gamma,k}$. ∎

The following lemma characterizes the expectations in (D.1) in terms of Markov chains.

Lemma D.12. *Let r, S and S' be independent with distributions $\widetilde{\mathcal{D}}$ (induced by $\mathcal{D} = \mathcal{U}$), $\mathcal{D}_{\gamma,k}$ and $\mathcal{D}_{\gamma,k}$ respectively. Then for stochastic matrices $\mathbf{M}_1 = \mathcal{L}_\gamma\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)$, $\mathbf{M}_2 = \mathcal{L}_{\gamma,\gamma} \circ \mathcal{Q}\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)$, $\mathbf{M}_3 = \mathcal{L}_\gamma \circ \mathcal{L}_{\gamma'}\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)$, $\mathbf{M}_4 = \mathcal{L}_{\gamma,\gamma} \circ \mathcal{Q} \circ \mathcal{L}_{\gamma'}\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)$, $\gamma' \triangleq (1-1/\Delta)^{\Delta-1}$, and any $u, v, w, x \in V[G]$, the following statements hold:*

1. $\mathbf{E}_{r,S}\left[X^S_{u,v}\right] = \left\langle \mathbf{e}_u\mathbf{M}_1^k, \mathbf{e}_v\right\rangle$,
2. $\mathbf{E}_{r,S,S'}\left[X^S_{u,v}X^{S'}_{w,x}\right] = \left\langle \mathbf{e}_{(u,w)}\mathbf{M}_2^k, \mathbf{e}_{(v,x)}\right\rangle$,
3. $\mathbf{E}_{r,S}\left[Y^S_{u,v}\right] = \left\langle \mathbf{e}_u\mathbf{M}_3^k, \mathbf{e}_v\right\rangle$, *and*

4. $\mathbf{E}_{r,S,S'}\left[Y^S_{u,v}Y^{S'}_{w,x}\right] = \left\langle \mathbf{e}_{(u,w)}\mathbf{M}_4^k, \mathbf{e}_{(v,x)}\right\rangle$.

Proof. We add $\Delta - \deg(u)$ self-loops to each node u and hence a non-lazy step of a forward walk is the same as a step of the random walk over $\mathrm{Reg}(G)$. Since S has distribution $\mathcal{D}_{\gamma,k}$ where each step is chosen to be lazy with probability $1-\gamma$, the forward walk with random pattern S starting from u is just a lazy random walk from u with transition matrix $\mathcal{L}_\gamma\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)$. This proves the first claim.

For the second claim, note that two forward walks are independent in some round i if at least one is lazy, since a lazy step is deterministic. The corresponding transition matrix is $\mathbf{I}\otimes\mathbf{I}$, $\mathbf{I}\otimes\mathbf{M}_{\mathrm{Reg}(G)}$ or $\mathbf{M}_{\mathrm{Reg}(G)}\otimes\mathbf{I}$, depending on which walk is lazy. When both walks are non-lazy and are at distinct nodes u and w respectively, they are still independent and behave according to $\mathbf{M}_{\mathrm{Reg}(G)}\otimes\mathbf{M}_{\mathrm{Reg}(G)}$ by the independence of $f(i,u)$ and $f(i,w)$. If $u=w$, then the two walks move the same node according to $\mathbf{M}_{\mathrm{Reg}(G)}$. So the case for two lazy steps is exactly characterized by the Doeblin coupling $\mathcal{Q}\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)$. And when the two walks have independent random patterns $S, S' \sim \mathcal{D}_{\gamma,k}$, the corresponding transition matrix is $\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)$ by definition. The second claim follows.

For the third claim, we consider the probability that a node $u \neq v$ is included in $N_{i,v}$. We divide it into two cases: the case that $N_{i,v} = \{u\}$ (i.e. $\widetilde{f}(T-1-2i,v) = \widetilde{f}(T-2-2i,v) = u$) and the case that $N_{i,v} = \{u,u'\}$ where $u \neq u'$. The first case occurs with probability $1/\Delta^2$. For the second one, we have $\left(\widetilde{f}(T-1-2i,v), \widetilde{f}(T-2-2i,v)\right) = (u,u')$ or (u',u) for some $u' \neq u$. And exactly one of them is counted by the condition $\widetilde{f}(T-1-2i,v) \preceq \widetilde{f}(T-2-2i,v)$. As they occur with the same probability we may assume it is the first one that is counted. Summing over $u' \neq u$, we conclude that this case occurs with probability $1/\Delta\cdot(1-1/\Delta)$. So u is included in $N_{i,v}$ with probability $1/\Delta$ for any $u \neq v$. And $N^\vee_{i,u} = \{v\}$ occurs when $u \in N_{i,v}$ and $u \notin N_{i,v'}$ for all $v' \in N(u)\setminus\{v\}$, whose probability is $1/\Delta\cdot(1-1/\Delta)^{\deg(u)-1}$. Taking the condition $r_{i,u} \leqslant (1-1/\Delta)^{\Delta-\deg(u)}$ into account, we see that the reversed walk extends from u to each neighbor $v \in N(u)$ with probability $1/\Delta\cdot(1-1/\Delta)^{\Delta-1} = \left(\mathcal{L}_{\gamma'}\left(M_{\mathrm{Reg}(G)}\right)\right)_{uv}$. So a non-lazy step of a reversed walk is the same as a step of the random walk over $\mathcal{L}_{\gamma'}\left(M_{\mathrm{Reg}(G)}\right)$. And the reversed walk with a random pattern chosen from $\mathcal{D}_{\gamma,k}$ corresponds to the transition matrix $\mathcal{L}_\gamma\circ\mathcal{L}_{\gamma'}\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)$, similarly to the first claim.

The proof to the last claim is similar to the second one. The only non-trivial part is to show that when two walks are both non-lazy and are at distinct nodes u and w respectively, they behave independently according to $\mathcal{L}_{\gamma'}\left(M_{\mathrm{Reg}(G)}\right)\otimes\mathcal{L}_{\gamma'}\left(M_{\mathrm{Reg}(G)}\right)$. Note that $r_{i,u}$ and $r_{i,w}$ are independent. So it suffices to show the probability that $N^\vee_{i,u} = \{v\}$ and $N^\vee_{i,w} = \{x\}$ both occur equals the product of their individual probabilities for all $v \in N(u)$ and $x \in N(w)$. For $a \in V[G]$ and $b \in N(a)$, let $\mathcal{I}_{a,b}$ be the event that $a \in N_{i,b}$. The claim follows if the events $\mathcal{I}_{a,b}$ are independent for all $a \in \{u,w\}$ and neighbor $b \in N(a)$. Note that $\mathcal{I}_{a,b}$ depends solely on $\widetilde{f}(T-1-2i,b)$ and $\widetilde{f}(T-2-2i,b)$. And those for different b are independent. So we reduce to proving $\mathcal{I}_{u,b}$ and $\mathcal{I}_{w,b}$ are independent for fixed $b \in N(u)\cap N(w)$. Each occurs with probability $1/\Delta$, as shown in the proof to the third claim. Both occurs exactly when $\left(\widetilde{f}(T-1-2i,b), \widetilde{f}(T-2-2i,b)\right)$ equals (u,w) if $u \preceq w$, or (w,u) if $w \preceq u$. So the probability that both events occur equals $1/\Delta^2$, as desired. ∎

D.2.2 Proof of Lemma 2.5

Lemma 2.5 (from page 7)**.** *Suppose $\mathbf{M} \in \mathbb{R}^{n\times n}$ is a doubly-stochastic matrix with spectral gap $\alpha > 0$, and suppose $\mathbf{M}_{uv} \leqslant \eta$ for any distinct $u, v \in V[G]$. Then for any distribution $\mathbf{u}$ over $V[G]\times V[G]$, $k \in \mathbb{N}$, and $0 \leqslant \gamma \leqslant \min\left\{1/3, \alpha\eta^{-1/2}/9\right\}$, we have*

$$\left\|\mathbf{u}\left(\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})\right)^k - \boldsymbol{\pi}\otimes\boldsymbol{\pi}\right\|_2 \leqslant (1-\gamma\alpha/2)^k + 2\sqrt{2}\gamma\alpha^{-1}n^{-3/2},$$

where $\boldsymbol{\pi}$ denotes the uniform distribution over $V[G]$.

To prove Lemma 2.5, we show that $\mathcal{L}_{\gamma,\gamma} \circ \mathcal{Q}(\mathbf{M})$ behaves similarly as $\mathcal{L}_\gamma(\mathbf{M}) \otimes \mathcal{L}_\gamma(\mathbf{M})$, in the sense that it almost preserves the vector $\boldsymbol{\pi} \otimes \boldsymbol{\pi}$ and shrinks vectors orthogonal to $\boldsymbol{\pi} \otimes \boldsymbol{\pi}$. For a distribution $\mathbf{u}$ over $V[G] \times V[G]$, we have the decomposition $\mathbf{u} = \boldsymbol{\pi} \otimes \boldsymbol{\pi} + \mathbf{u}^\perp$ where $\mathbf{u}^\perp \triangleq \mathbf{u} - \boldsymbol{\pi} \otimes \boldsymbol{\pi}$ is orthogonal to $\boldsymbol{\pi} \otimes \boldsymbol{\pi}$.

Lemma D.13. *Let $\mathbf{M}$, $\boldsymbol{\pi}$ and γ be as in Lemma 2.5. Then $\left\|((\boldsymbol{\pi} \otimes \boldsymbol{\pi})\mathcal{L}_{\gamma,\gamma} \circ \mathcal{Q}(\mathbf{M}))^\perp\right\|_2 \leqslant \sqrt{2}\gamma^2 n^{-3/2}$.*

Proof. Let $\mathbf{E} = \mathcal{L}_{\gamma,\gamma} \circ \mathcal{Q}(\mathbf{M}) - \mathcal{L}_\gamma(\mathbf{M}) \otimes \mathcal{L}_\gamma(\mathbf{M})$. Note that $((\boldsymbol{\pi} \otimes \boldsymbol{\pi})\mathcal{L}_{\gamma,\gamma} \circ \mathcal{Q}(\mathbf{M}))^\perp = ((\boldsymbol{\pi} \otimes \boldsymbol{\pi})\mathbf{E})^\perp$, since $\mathcal{L}_\gamma(\mathbf{M}) \otimes \mathcal{L}_\gamma(\mathbf{M})$ fixes $\boldsymbol{\pi} \otimes \boldsymbol{\pi}$. Also note that $\|((\boldsymbol{\pi} \otimes \boldsymbol{\pi})\mathbf{E})^\perp\|_2 \leqslant \|(\boldsymbol{\pi} \otimes \boldsymbol{\pi})\mathbf{E}\|_2$. So it suffices to prove $\|(\boldsymbol{\pi} \otimes \boldsymbol{\pi})\mathbf{E}\|_2 \leqslant \sqrt{2}\gamma^2 n^{-3/2}$. By definition, we have

$$\mathbf{E}_{(u,w)(v,x)} = \begin{cases} 0 & u \neq w, \\ \gamma^2 \left(\mathbf{M}_{uv} - (\mathbf{M} \otimes \mathbf{M})_{(u,w)(v,x)}\right) & u = w, v = x, \\ -\gamma^2 (\mathbf{M} \otimes \mathbf{M})_{(u,w)(v,x)} & u = w, v \neq x. \end{cases}$$

So $\mathbf{E}_{(u,w)(v,x)} = 0$ for $u \neq w$, and $\left|\mathbf{E}_{(u,w)(v,x)}\right| \leqslant \gamma^2 \mathbf{M}_{uv}$ for $u = w$. Then for any $v, x \in V[G]$, we have

$$\left|((\boldsymbol{\pi} \otimes \boldsymbol{\pi})\mathbf{E})_{(v,x)}\right| \leqslant \sum_{u \in V[G]} (1/n^2) \cdot \gamma^2 \mathbf{M}_{uv} = \gamma^2/n^2.$$

So $\|(\boldsymbol{\pi} \otimes \boldsymbol{\pi})\mathbf{E}\|_\infty \leqslant \gamma^2/n^2$. We also have

$$\begin{aligned} \|(\boldsymbol{\pi} \otimes \boldsymbol{\pi})\mathbf{E}\|_1 &= \sum_{v,x \in V[G]} \left| \sum_{u \in V[G]} (1/n^2) \mathbf{E}_{(u,u)(v,x)} \right| \\ &\leqslant \sum_{u,v \in V[G]} (1/n^2)\gamma^2 \mathbf{M}_{uv} + \sum_{u,v,x \in V[G]} (1/n^2)\gamma^2 (\mathbf{M} \otimes \mathbf{M})_{(u,u)(v,x)} \\ &= 2\gamma^2/n. \end{aligned}$$

By Hölder's inequality, we have

$$\|(\boldsymbol{\pi} \otimes \boldsymbol{\pi})\mathbf{E}\|_2 \leqslant \sqrt{\|(\boldsymbol{\pi} \otimes \boldsymbol{\pi})\mathbf{E}\|_1 \|(\boldsymbol{\pi} \otimes \boldsymbol{\pi})\mathbf{E}\|_\infty} \leqslant \sqrt{2}\gamma^2 n^{-3/2}.$$

■

Lemma D.14. *Let $\mathbf{M}$, $\boldsymbol{\pi}$ and γ be as in Lemma 2.5. For any vector $\mathbf{u} \in \mathbb{R}^n \otimes \mathbb{R}^n$ orthogonal to $\boldsymbol{\pi} \otimes \boldsymbol{\pi}$, we have $\mathbf{u}\mathcal{L}_{\gamma,\gamma} \circ \mathcal{Q}(\mathbf{M}) \perp \boldsymbol{\pi} \otimes \boldsymbol{\pi}$ and*

$$\|\mathbf{u}\mathcal{L}_{\gamma,\gamma} \circ \mathcal{Q}(\mathbf{M})\|_2 \leqslant \left(1 - (1-\gamma)\gamma\alpha + \gamma^2\sqrt{2\eta}\right) \|\mathbf{u}\|_2.$$

Proof. Since $\mathcal{L}_{\gamma,\gamma} \circ \mathcal{Q}(\mathbf{M})$ is stochastic, we have

$$\langle \mathbf{u}\mathcal{L}_{\gamma,\gamma} \circ \mathcal{Q}(\mathbf{M}), \boldsymbol{\pi} \otimes \boldsymbol{\pi} \rangle = \langle \mathbf{u}, (\boldsymbol{\pi} \otimes \boldsymbol{\pi})(\mathcal{L}_{\gamma,\gamma} \circ \mathcal{Q}(\mathbf{M}))^\intercal \rangle = \langle \mathbf{u}, \boldsymbol{\pi} \otimes \boldsymbol{\pi} \rangle = 0.$$

To prove the second claim, we write $\mathcal{L}_{\gamma,\gamma} \circ \mathcal{Q}(\mathbf{M}) = \mathbf{R}_1 + \mathbf{R}_2$ where

$$\mathbf{R}_1 = (1-\gamma)^2(\mathbf{I} \otimes \mathbf{I}) + (1-\gamma)\gamma(\mathbf{I} \otimes \mathbf{M}) + \gamma(1-\gamma)(\mathbf{M} \otimes \mathbf{I})$$

and $\mathbf{R}_2 = \gamma^2 \mathcal{Q}(\mathbf{M})$. Then we bound $\|\mathbf{u}\mathbf{R}_1\|_2$ and $\|\mathbf{u}\mathbf{R}_2\|_2$ individually.

Observe that $\mathbf{R}_1 = (1-\gamma^2)\mathcal{L}_{\gamma_0}(\mathbf{R}_0)$ where $\mathbf{R}_0$ is the stochastic matrix $(\mathbf{I} \otimes \mathbf{M} + \mathbf{M} \otimes \mathbf{I})/2$ and $\gamma_0 = 2\gamma/(1+\gamma)$. Recall that $\mathbf{M}$ has n normalized orthogonal eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in $\mathbb{R}^n$ associated with n real eigenvalues $1 = \lambda_1 > 1 - \alpha \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n \geqslant -1$ respectively, and $\mathbf{v}_1$

is parallel to $\boldsymbol{\pi}$. Then $\mathbf{R}_0$ has n^2 normalized orthogonal eigenvectors $\mathbf{v}_i \otimes \mathbf{v}_j$ associated with eigenvalues $(\lambda_j + \lambda_i)/2$, $i, j = 1, \ldots, n$. And $\mathcal{L}_{\gamma_0}(\mathbf{R}_0)$ has the same set of eigenvectors, with the (i, j)th eigenvalue replaced by $(1 - \gamma_0) + \gamma_0(\lambda_j + \lambda_i)/2$. These eigenvalues are all non-negative, since $(\lambda_j + \lambda_i)/2 \geqslant -1$ and $\gamma_0 = 2\gamma/(1+\gamma) \leqslant 1/2$ (from the condition $\gamma \leqslant 1/3$). So the absolute spectral gap of $\mathcal{L}_{\gamma_0}(\mathbf{R}_0)$ is

$$\begin{aligned}
&1 - \max_{(i,j)\neq(1,1)} ((1 - \gamma_0) + \gamma_0(\lambda_j + \lambda_i)/2) \\
&= 1 - \max_{(i,j)\neq(1,1)} \left(1 + \left(\frac{\lambda_i + \lambda_j}{2} - 1\right)\gamma_0\right) \\
&= \min_{(i,j)\neq(1,1)} \left(\frac{1 - \lambda_i + 1 - \lambda_j}{2}\right)\gamma_0 \geqslant \gamma_0\alpha/2.
\end{aligned}$$

As $\mathbf{u}$ is parallel to $\boldsymbol{\pi} \otimes \boldsymbol{\pi}$, or equivalently $\mathbf{v}_1 \otimes \mathbf{v}_1$, we have

$$\|\mathbf{u}\mathbf{R}_1\|_2 = (1 - \gamma^2)\left\|\mathbf{u}\mathcal{L}_{\gamma_0}(\mathbf{R}_0)\right\|_2 \leqslant (1 - \gamma^2)(1 - \gamma_0\alpha/2)\|\mathbf{u}\|_2 = (1 - \gamma^2)(1 - \gamma\alpha/(1 + \gamma))\|\mathbf{u}\|_2.$$

Then we bound $\|\mathbf{u}\mathbf{R}_2\|_2 = \gamma^2\|\mathbf{u}\mathcal{Q}(\mathbf{M})\|_2$. By permutating the rows (resp. columns) of $\mathcal{Q}(\mathbf{M})$, we assume its first n rows (resp. n columns) are indexed by the diagonal elements $\{(u, u) : u \in V[G]\}$. By definition, we have

$$\mathcal{Q}(\mathbf{M}) = \begin{pmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{A} & \mathbf{B} \end{pmatrix}$$

where $\begin{pmatrix} \mathbf{A} & \mathbf{B} \end{pmatrix}$ are the last $n^2 - n$ rows of $\mathbf{M} \otimes \mathbf{M}$ (we permutate the rows and columns of $\mathbf{M} \otimes \mathbf{M}$ in the same way as we did for $\mathcal{Q}(\mathbf{M})$. Write $\mathbf{u} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{pmatrix}$ where $\mathbf{u}_1 \in \mathbb{R}^n$ and $\mathbf{u}_2 \in \mathbb{R}^{n^2-n}$, consisting of entries indexed by (u, w), $u = w$ and $u \neq w$ respectively. Then

$$\begin{aligned}
\|\mathbf{u}\mathcal{Q}(\mathbf{M})\|_2^2 &= \left\|\begin{pmatrix} \mathbf{u}_1\mathbf{M} + \mathbf{u}_2\mathbf{A} & \mathbf{u}_2\mathbf{B} \end{pmatrix}\right\|_2^2 \\
&= \|\mathbf{u}_1\mathbf{M}\|_2^2 + \|\mathbf{u}_2\mathbf{A}\|_2^2 + \|\mathbf{u}_2\mathbf{B}\|_2^2 + 2\langle \mathbf{u}_1\mathbf{M}, \mathbf{u}_2\mathbf{A}\rangle \\
&= \|\mathbf{u}_1\mathbf{M}\|_2^2 + \left\|\begin{pmatrix} \mathbf{0} & \mathbf{u}_2 \end{pmatrix}(\mathbf{M} \otimes \mathbf{M})\right\|_2^2 + 2\langle \mathbf{u}_1\mathbf{M}, \mathbf{u}_2\mathbf{A}\rangle \\
&\leqslant \|\mathbf{u}_1\|_2^2 + \|\mathbf{u}_2\|_2^2 + 2\langle \mathbf{u}_1\mathbf{M}, \mathbf{u}_2\mathbf{A}\rangle \\
&= \|\mathbf{u}\|_2^2 + 2\langle \mathbf{u}_1\mathbf{M}, \mathbf{u}_2\mathbf{A}\rangle \\
&\leqslant \|\mathbf{u}\|_2^2 + 2\|\mathbf{u}_1\mathbf{M}\|_2\|\mathbf{u}_2\mathbf{A}\|_2 \\
&\leqslant \|\mathbf{u}\|_2^2 + 2\|\mathbf{u}\|_2^2\|\mathbf{A}\|_2 \\
&\leqslant \|\mathbf{u}\|_2^2\left(1 + 2\sqrt{\|\mathbf{A}\|_1\|\mathbf{A}\|_\infty}\right)
\end{aligned} \tag{D.4}$$

The third equality uses the fact that $\begin{pmatrix} \mathbf{0} & \mathbf{u}_2 \end{pmatrix}(\mathbf{M} \otimes \mathbf{M}) = \begin{pmatrix} \mathbf{u}_2\mathbf{A} & \mathbf{u}_2\mathbf{B} \end{pmatrix}$. The first inequality uses the fact that $\|\mathbf{M}\|_2, \|\mathbf{M} \otimes \mathbf{M}\|_2 \leqslant 1$. The second inequality is an instance of the Cauchy-Schwarz inequality. The third one uses the facts that $\|\mathbf{M}\|_2 \leqslant 1$ and $\|\mathbf{u}_1\|_2, \|\mathbf{u}_2\|_2 \leqslant \|\mathbf{u}\|_2$. And the last one uses the inequality $\|\mathbf{A}\|_2 \leqslant \sqrt{\|\mathbf{A}\|_1\|\mathbf{A}\|_\infty}$.

We have $\|\mathbf{A}\|_\infty \leqslant \|\mathbf{M} \otimes \mathbf{M}\|_\infty = 1$. To bound $\|\mathbf{A}\|_1$, observe that $\|\mathbf{A}\|_1$ is by definition the maximum of the ℓ_1-norm of rows of $\mathbf{A}$. Then

$$\begin{aligned}
\|\mathbf{A}\|_1 &= \max_{\substack{u,w\in V[G] \\ u\neq w}} \sum_{v\in V[G]} \mathbf{M}_{uv}\mathbf{M}_{wv} \\
&\leqslant \max_{\substack{u,w\in V[G] \\ u\neq w}} \left(\eta\mathbf{M}_{ww} + \eta \sum_{v\in V[G]\setminus\{w\}} \mathbf{M}_{uv}\right) \leqslant 2\eta.
\end{aligned}$$

Combining it with (D.4), we obtain

$$\|\mathbf{u}\mathbf{R}_2\|_2^2 = \gamma^4\|\mathbf{u}\mathcal{Q}(\mathbf{M})\|_2^2 \leqslant \gamma^4\left(1+2\sqrt{2\eta}\right)\|\mathbf{u}\|_2^2 \leqslant \gamma^4\left(1+\sqrt{2\eta}\right)^2\|\mathbf{u}\|_2^2.$$

Therefore

$$\begin{aligned}\|\mathbf{u}\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})\|_2 &\leqslant \|\mathbf{u}\mathbf{R}_1\|_2 + \|\mathbf{u}\mathbf{R}_2\|_2\\ &\leqslant (1-\gamma^2)(1-\gamma\alpha/(1+\gamma))\|\mathbf{u}\|_2 + \gamma^2\left(1+\sqrt{2\eta}\right)\|\mathbf{u}\|_2\\ &= \left(1-(1-\gamma)\gamma\alpha+\gamma^2\sqrt{2\eta}\right)\|\mathbf{u}\|_2.\end{aligned}$$

■

Proof of Lemma 2.5. Note that we are bounding the ℓ_2-norm of $\mathbf{u}\left(\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})\right)^k - \boldsymbol{\pi}\otimes\boldsymbol{\pi} = \left(\mathbf{u}\left(\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})\right)^k\right)^\perp$. The proof is based on the induction on k. When $k=0$, we have

$$\left\|\left(\mathbf{u}\left(\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})\right)^k\right)^\perp\right\|_2 \leqslant \left\|\mathbf{u}\left(\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})\right)^k\right\|_2 \leqslant 1,$$

and hence the claim holds. For $k>0$, assume the claim holds for $k'<k$. Let $\mathbf{v} = \mathbf{u}\left(\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})\right)^{k-1}$. We have

$$\begin{aligned}\left(\mathbf{u}\left(\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})\right)^k\right)^\perp &= \left(\mathbf{v}\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})\right)^\perp\\ &= \left((\boldsymbol{\pi}\otimes\boldsymbol{\pi})\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})\right)^\perp + \left(\mathbf{v}^\perp\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})\right)^\perp.\end{aligned}$$

By Lemma D.13, we have

$$\|\left((\boldsymbol{\pi}\otimes\boldsymbol{\pi})\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})\right)^\perp\|_2 \leqslant \sqrt{2}\gamma^2 n^{-3/2}.$$

And by Lemma D.14, we have $\left(\mathbf{v}^\perp\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})\right)^\perp = \mathbf{v}^\perp\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})$ whose ℓ_2-norm is at most

$$\left(1-(1-\gamma)\gamma\alpha+\gamma^2\sqrt{2\eta}\right)\|\mathbf{v}\|_2 \leqslant (1-\gamma\alpha/2)\|\mathbf{v}\|_2$$

where we use the condition $\gamma \leqslant \left\{1/3, \alpha\eta^{-1/2}/9\right\}$. This is bounded by

$$(1-\gamma\alpha/2)\left((1-\gamma\alpha/2)^{k-1} + 2\sqrt{2}\gamma\alpha^{-1}n^{-3/2}\right) = (1-\gamma\alpha/2)^k + 2\sqrt{2}\gamma\alpha^{-1}n^{-3/2}(1-\gamma\alpha/2)$$

by the induction hypothesis. Then

$$\begin{aligned}\left\|\left(\mathbf{u}\left(\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})\right)^k\right)^\perp\right\|_2 &\leqslant \left\|\left((\boldsymbol{\pi}\otimes\boldsymbol{\pi})\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})\right)^\perp\right\|_2 + \left\|\left(\mathbf{v}^\perp\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})\right)^\perp\right\|_2\\ &\leqslant \sqrt{2}\gamma^2 n^{-3/2} + (1-\gamma\alpha/2)^k + 2\sqrt{2}\gamma\alpha^{-1}n^{-3/2}(1-\gamma\alpha/2)\\ &= (1-\gamma\alpha/2)^k + 2\sqrt{2}\gamma\alpha^{-1}n^{-3/2}\end{aligned}$$

as desired. ■

As a side product, we show that the chain $\mathcal{L}_{\gamma,\gamma}\circ\mathcal{Q}(\mathbf{M})$ behaves similarly as $\mathcal{L}_\gamma(\mathbf{M})\otimes\mathcal{L}_\gamma(\mathbf{M})$ in terms of the stationary distribution and the mixing time.

Corollary D.15. *Let $\mathbf{M}$, γ and α be as in Lemma 2.5. Let $\boldsymbol{\pi}'$ be the stationary distribution*[6] *of $\mathcal{L}_{\gamma,\gamma} \circ \mathcal{Q}(\mathbf{M})$. Then*

$$\left\|\boldsymbol{\pi}' - \boldsymbol{\pi} \otimes \boldsymbol{\pi}\right\|_2 \leqslant (1-\gamma\alpha/2)^k + 2\sqrt{2}\gamma\alpha^{-1}n^{-3/2}.$$

Define the ℓ_1-mixing time $\bar{\tau}(\varepsilon) := \max_{\mathbf{u}} \min\{k : \|\mathbf{u}\left(\mathcal{L}_{\gamma,\gamma} \circ \mathcal{Q}(\mathbf{M})\right)^k - \boldsymbol{\pi} \otimes \boldsymbol{\pi}\|_1 \leqslant \varepsilon\}$ where $\mathbf{u}$ ranges over all distributions over $V[G] \times V[G]$. Assuming $\gamma\alpha^{-1} = O(n^{1/2-c})$ for some constant $c > 0$, we have $\bar{\tau}(\varepsilon) = O(\gamma^{-1}\alpha^{-1}(\log n + \log \varepsilon^{-1}))$.

Proof. The first claim follows directly from Lemma 2.5. We also have

$$\left\|\mathbf{u}\left(\mathcal{L}_{\gamma,\gamma} \circ \mathcal{Q}(\mathbf{M})\right)^k - \boldsymbol{\pi} \otimes \boldsymbol{\pi}\right\|_1 \leqslant n\left\|\mathbf{u}\left(\mathcal{L}_{\gamma,\gamma} \circ \mathcal{Q}(\mathbf{M})\right)^k - \boldsymbol{\pi} \otimes \boldsymbol{\pi}\right\|_2 \leqslant n^{-c}$$

for sufficiently large $k = O(\gamma^{-1}\alpha^{-1}\log n)$, again by Lemma 2.5. So $\bar{\tau}(n^{-c}) = O(\gamma^{-1}\alpha^{-1}\log n)$. The second claim then follows from the well-known fact that $\bar{\tau}(\varepsilon) \leqslant \bar{\tau}(\delta)\lceil\log_\delta \varepsilon\rceil$ for $\varepsilon, \delta > 0$. ■

We know that the stationary distribution of $\mathcal{Q}(\mathbf{M})$ is the uniform distribution over the set of diagonal entries $\{(u,u) : u \in V[G]\}$. So is the stationary distribution of the lazy chain $\mathcal{L}_\gamma \circ \mathcal{Q}(\mathbf{M})$ for any $\gamma \in (0,1]$. Interestingly, Corollary D.15 tells us that the "bi-lazy" chain $\mathcal{L}_{\gamma,\gamma} \circ \mathcal{Q}(\mathbf{M})$ behaves very differently, as its stationary distribution is close to $\boldsymbol{\pi} \otimes \boldsymbol{\pi}$ instead.

D.2.3 Proof of Theorem 2.6

We are now ready to derive a bound on the runtime of Protocol 1.

Lemma D.16. *Suppose G has spectral gap α and irregularity β. Using Protocol 1 with distribution $\mathcal{D} = \mathcal{U}$, any node gets the rumor in $T = O(C\log n)$ rounds with probability at least $1 - O(n^{-2c})$ where $C = (1/\alpha)\cdot\beta^2\max\{1, 1/(\alpha\cdot\Delta^{0.5-c})\}$ and $c > 0$ is an arbitrary small constant.*

Proof. Let $s \in V[G]$ be the initial node and fix a target node $w \in V[G]$. Let $c > 0$ be any constant. Choose $\gamma = \min\left\{1/3, \Delta^{0.5-c}\alpha/9\right\} \leqslant n^{0.5-c}\alpha/9$. Choose $k = (\gamma\gamma'\alpha)^{-1}\beta^2\log n + 1$ and let $T = 4k$. So $T = O(C\log n)$. Define the distributions $\mathbf{u} = \mathbf{e}_s\mathbf{M}_1^k$, $\mathbf{v} = \mathbf{e}_{(s,s)}\mathbf{M}_2^k$, $\mathbf{u}' = \mathbf{e}_w\mathbf{M}_3^k$, and $\mathbf{v}' = \mathbf{e}_{(w,w)}\mathbf{M}_4^k$, where $\mathbf{M}_1, \ldots, \mathbf{M}_4$ are as in Lemma D.12. Let $\boldsymbol{\pi}$ be the uniform distribution over $V[G]$. As before, let $\mathbf{u}^\perp = \mathbf{u} - \boldsymbol{\pi}$ and $\mathbf{v}^\perp = \mathbf{v} - \boldsymbol{\pi}\otimes\boldsymbol{\pi}$, and similarly for $\mathbf{u}'$ and $\mathbf{v}'$. By Lemma D.11 and Lemma D.12, the probability that w gets the rumor in k rounds is lower bounded by

$$\begin{aligned}
&\frac{\sum_{u,v\in V[G]}\langle\mathbf{u},\mathbf{e}_u\rangle\langle\mathbf{u},\mathbf{e}_v\rangle\langle\mathbf{u}',\mathbf{e}_u\rangle\langle\mathbf{u}',\mathbf{e}_v\rangle}{\sum_{u,v\in V[G]}\left\langle\mathbf{v},\mathbf{e}_{(u,v)}\right\rangle\left\langle\mathbf{v}',\mathbf{e}_{(u,v)}\right\rangle} = \frac{\langle\mathbf{u},\mathbf{u}'\rangle^2}{\langle\mathbf{v},\mathbf{v}'\rangle} \\
&= \frac{\left(\langle\boldsymbol{\pi},\boldsymbol{\pi}\rangle + \langle\mathbf{u}^\perp,\boldsymbol{\pi}\rangle + \langle\boldsymbol{\pi},\mathbf{u}'^\perp\rangle + \langle\mathbf{u}^\perp,\mathbf{u}'^\perp\rangle\right)^2}{\langle\boldsymbol{\pi}\otimes\boldsymbol{\pi},\boldsymbol{\pi}\otimes\boldsymbol{\pi}\rangle + \langle\mathbf{v}^\perp,\boldsymbol{\pi}\otimes\boldsymbol{\pi}\rangle + \langle\boldsymbol{\pi}\otimes\boldsymbol{\pi},\mathbf{v}'^\perp\rangle + \langle\mathbf{v}^\perp,\mathbf{v}'^\perp\rangle} \\
&= \frac{\left(1/n + \langle\mathbf{u}^\perp,\mathbf{u}'^\perp\rangle\right)^2}{1/n^2 + \langle\mathbf{v}^\perp,\mathbf{v}'^\perp\rangle}.
\end{aligned} \tag{D.5}$$

Note that $\mathbf{M}_1 = \mathcal{L}_\gamma\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)$ and $\mathbf{M}_3 = \left(\mathcal{L}_\gamma \circ \mathcal{L}_{\gamma'}\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)\right)$ have absolute spectral gaps $\gamma\alpha\beta^{-2}$ and $\gamma\gamma'\alpha\beta^{-2}$ respectively. This follows from Lemma A.3 and the definition of lazy Markov chains (Also, the lazyness guarantees that the eigenvalues are all non-negative, and hence the bounds are about absolute spectral gaps, not just spectral gaps). By Lemma D.9 and the fact that $k \geqslant (\gamma\gamma'\alpha)^{-1}\beta^2\log n + 1 \geqslant \log_{1-\gamma\gamma'\alpha\beta^{-2}}(1/n) + 1$, we have $|\langle\mathbf{u}^\perp,\mathbf{u}'^\perp\rangle| \leqslant \left\|\mathbf{u}^\perp\right\|_2\left\|\mathbf{u}'^\perp\right\|_2 \leqslant 1/n^2$. By Lemma 2.5 (with $\eta = 1/\Delta$), we have

$$|\langle\mathbf{v}^\perp,\mathbf{v}'^\perp\rangle| \leqslant \left\|\mathbf{v}^\perp\right\|_2\left\|\mathbf{v}'^\perp\right\|_2 \leqslant \left((1-\gamma\alpha/2)^k + 2\sqrt{2}\gamma\alpha^{-1}n^{-3/2}\right)^2 \leqslant 1/n^{2+2c}.$$

So (D.5) is lower bounded by $\frac{(1/n-1/n^2)^2}{1/n^2-1/n^{2+2c}} = 1 - O(n^{-2c})$. ■

[6]The lazyness and $\alpha > 0$ guarantees that $\mathcal{L}_{\gamma,\gamma} \circ \mathcal{Q}(\mathbf{M})$ is ergodic and has a unique stationary distribution.

Theorem 2.6 is obtained by repeating the protocol $O(1)$ times and apply the union bound.

D.3 Analysis of Protocol 2

Let $\mathcal{P}$ be the distribution over the set of functions $f : [T] \times V[G] \to [\Delta]$ associated with Protocol 2. The values $f(i,u)$ in the ith round are generated using the PRG $\mathcal{G}$, and the seeds of $\mathcal{G}$ in different rounds are generated by the PRG $\mathcal{G}'$. In this section we show that Protocol 1 with distribution $\mathcal{D} = \mathcal{P}$ has almost the same performance as the one with $\mathcal{D} = \mathcal{U}$. As an intermediate step, we consider the distribution $\mathcal{P}'$ defined as follows: the values of f in each round are determined by the PRG $\mathcal{G}$ in the same way as for $\mathcal{P}$ but the seeds of $\mathcal{G}$ in different rounds are now independent and random, instead of being generated by $\mathcal{G}'$. With $\mathcal{D} = \mathcal{P}'$, Definition 2.1 are still valid, and Lemma D.11 still holds by exactly the same proof. Moreover, Lemma D.12 "almost holds" in the following sense.

Lemma D.17. *Let r, S and S' be independent with distributions $\widetilde{\mathcal{D}}$ (induced by $\mathcal{D} = \mathcal{P}'$), $\mathcal{D}_{\gamma,k}$ and $\mathcal{D}_{\gamma,k}$ respectively. Then there exist stochastic matrices $\mathbf{M}'_1, \mathbf{M}'_3 \in \mathbb{R}^{n\times n}$, $\mathbf{M}'_2, \mathbf{M}'_4 \in \mathbb{R}^{n\times n} \otimes \mathbb{R}^{n\times n}$ such that $\|\mathbf{M}'_i - \mathbf{M}_i\|_1 \leqslant 12\gamma\Delta^2(\varepsilon + 2\Delta^3/m)$ for $1 \leqslant i \leqslant 4$, where $\mathbf{M}_i$ are as in Lemma D.12 and ε, m are as in Protocol 2. Moreover, for any $u, v, w, x \in V[G]$, the following statements hold:*

1. $\mathbf{E}_{r,S}\left[\, X^S_{u,v} \,\right] = \left\langle \mathbf{e}_u \mathbf{M}'^k_1, \mathbf{e}_v \right\rangle$,
2. $\mathbf{E}_{r,S,S'}\left[X^S_{u,v} X^{S'}_{w,x} \right] = \left\langle \mathbf{e}_{(u,w)} \mathbf{M}'^k_2, \mathbf{e}_{(v,x)} \right\rangle$,
3. $\mathbf{E}_{r,S}\left[\, Y^S_{u,v} \,\right] = \left\langle \mathbf{e}_u \mathbf{M}'^k_3, \mathbf{e}_v \right\rangle$,
4. $\mathbf{E}_{r,S,S'}\left[Y^S_{u,v} Y^{S'}_{w,x} \right] = \left\langle \mathbf{e}_{(u,w)} \mathbf{M}'^k_4, \mathbf{e}_{(v,x)} \right\rangle$.

Proof. Let $\mathbf{M}'_1$ (resp. $\mathbf{M}'_3$) be the transition matrix of a forward (reversed) random walk with random pattern $S \sim \mathcal{D}_{\gamma,k}$. Let $\mathbf{M}'_2$ (resp. $\mathbf{M}'_4$) be the joint transition matrix of two forward (reversed) random walks with random patterns $S, S' \sim \mathcal{D}_{\gamma,k}$. This is exactly the same setting as in Lemma D.12, except that now $\mathcal{D} = \mathcal{P}'$. Since the randomness $f(i,u)$ and $r_{i,u}$ in different rounds are independent, Items 1 – 4 clearly hold. It remains to show that $\|\mathbf{M}'_i - \mathbf{M}_i\|_1 \leqslant 12\gamma\Delta^2(\varepsilon + 2\Delta^3/m)$ for $1 \leqslant i \leqslant 4$.

Recall that the ℓ_1-norm of a matrix equals the maximal sum of absolute values of entries in a row. So we may fix the row index u (or (u,v)) maximizing the sum. Also fix the auxiliary randomness $\{r_{i,u}\}$ and since if we have a bound for all fixed $\{r_{i,u}\}$, the same bound applies when they are random.

Consider the ith step of a forward walk with random pattern $S \sim \mathcal{D}_{\gamma,k}$ from node u. The walk stays at u if that step is lazy for $\mathcal{D} = \mathcal{P}'$ and also for $\mathcal{D} = \mathcal{U}$. So we may assume the step is non-lazy which occurs with probability γ. The event that the walk moves to v is determined solely by $f(i,u)$ and hence characterized by a combinatorial rectangle of dimension one. By Lemma D.4, we have $|(\mathbf{M}'_1 - \mathbf{M}_1)_{uv}| \leqslant \gamma(\varepsilon + \Delta/m)$ (note that the difference is counted only when the step is non-lazy). Note that the walk always moves to a node in $N(u) \cup \{u\}$. Taking the sum of differences, we have $\|\mathbf{M}'_1 - \mathbf{M}_1\|_1 \leqslant \gamma(\Delta+1)(\varepsilon + \Delta/m)$.

Now consider the ith step of two forward walks from u and w respectively. We may assume at least one of them has a non-lazy step which occurs with probability $2(1-\gamma)\gamma + \gamma^2 \leqslant 2\gamma$. The event that the first walk moves to some node v is determined by $f(i,u)$ whereas the event that the second walk moves to some x is determined by $f(i,w)$. Each is characterized by a combinatorial rectangle in $\prod_{a\in\{u,w\}}[\Delta]$ of dimension one (if $u = w$) or two (if $u \neq w$). The conjunction of these two events is characterized by the intersection of the two combinatorial rectangles, which is again a combinatorial rectangle in $\prod_{a\in\{u,w\}}[\Delta]$. By Lemma D.4, we have $\left|(\mathbf{M}'_2 - \mathbf{M}_2)_{(u,w)(v,x)}\right| \leqslant$

$2\gamma(\varepsilon + 2\Delta/m)$. Also the only possible (v, x) are in $(N(u) \cup \{u\}) \times (N(w) \cup \{w\})$. Taking the sum of differences, we have $\|\mathbf{M}_2' - \mathbf{M}_2\|_1 \leqslant 2\gamma(\Delta+1)^2(\varepsilon + 2\Delta/m)$.

Now consider the ith step of a reversed walk with random pattern $S \sim \mathcal{D}_{\gamma,k}$ from a node u. Again assume the step is non-lazy which occurs with probability γ. Let $i_0 = T-2i-1$ and $i_1 = T-2i-2$. The event $u \in N_{i,v}$ for $v \in N(u)$ is determined by whether $(f(i_0, v), f(i_1, v)) \in S_{v,u}$ for some $S_{v,u} \subseteq [\Delta^2]$. Then the event whether $N_{i,u}^{\vee} = \{v\}$ for $v \in N(u)$ is characterized by the combinatorial rectangle $\prod_{w\in N(u)} S_w \subseteq \prod_{w\in N(u)}[\Delta^2]$ of dimension $\deg(u) \leqslant \Delta$ where S_w equals $S_{w,u}$ if $w = v$, and equals $[\Delta^2]\setminus S_{w,u}$ if $w \neq v$. By Lemma D.4, we have $|(\mathbf{M}_3' - \mathbf{M}_3)_{uv}| \leqslant \gamma(\varepsilon+\Delta^3/m)$ for $v \in N(u)$. When $v = u$, we have $|(\mathbf{M}_3' - \mathbf{M}_3)_{uv}| \leqslant \sum_{w\in N(u)} |(\mathbf{M}_3' - \mathbf{M}_3)_{uw}| \leqslant \gamma\Delta(\varepsilon + \Delta^3/m)$ since $\sum_{w\in N(u)\cup\{u\}} (\mathbf{M}_3' - \mathbf{M}_3)_{uw} = \sum_{w\in N(u)\cup\{u\}} (\mathbf{M}_3')_{uw} - \sum_{w\in N(u)\cup\{u\}} (\mathbf{M}_3)_{uw} = 1 - 1 = 0$. Taking the sum of differences, we have $\|\mathbf{M}_3' - \mathbf{M}_3\|_1 \leqslant 2\gamma\Delta(\varepsilon + \Delta^3/m)$.

Finally consider the ith step of two reversed walks from u and w respectively. We may assume at least one of them has a non-lazy step which occurs with probability $2(1-\gamma)\gamma+\gamma^2 \leqslant 2\gamma$. Similar to the case of two forward walks, using the fact that the family of combinatorial rectangles is closed under intersection, we know the event that the two walks move to some nodes $v \in N(u)$ and $x \in N(u)$ respectively is characterized by a combinatorial rectangle in $\prod_{a\in N(u)\cup N(w)}[\Delta^2]$ of dimension at most 2Δ. By Lemma D.4, we have $\left|(\mathbf{M}_4' - \mathbf{M}_4)_{(u,w)(v,x)}\right| \leqslant 2\gamma(\varepsilon + 2\Delta^3/m)$ for $v \in N(u)$ and $x \in N(w)$. When $u \neq v$ and $w = x$, using the fact that $\mathbf{M}_4$ (resp. $\mathbf{M}_4'$) is a coupling of two copies of $\mathbf{M}_3$ (resp. $\mathbf{M}_3'$), we have $\sum_{x'\in N(w)} (\mathbf{M}_4' - \mathbf{M}_4)_{(u,w)(v,x')} = (\mathbf{M}_3' - \mathbf{M}_3)_{uv}$ and hence

$$\begin{aligned}\left|\left(\mathbf{M}_4' - \mathbf{M}_4\right)_{(u,w)(v,x)}\right| &\leqslant \left|\left(\mathbf{M}_3' - \mathbf{M}_3\right)_{uv}\right| + \sum_{x'\in N(w)\setminus\{w\}} \left|\left(\mathbf{M}_4' - \mathbf{M}_4\right)_{(u,w)(v,x')}\right| \\ &\leqslant \gamma(\varepsilon + \Delta^3/m) + 2\gamma\Delta(\varepsilon + 2\Delta^3/m).\end{aligned} \tag{D.6}$$

The case that $u = v$ and $w \neq x$ is symmetric. When $u = v$ and $w = x$, the first inequality of (D.6) still holds, yet the RHS of the second one becomes $\gamma\Delta(\varepsilon + \Delta^3/m) + \Delta(\gamma(\varepsilon + \Delta^3/m) + 2\gamma\Delta(\varepsilon + 2\Delta^3/m))$. Taking the sum of differences, we have $\|\mathbf{M}_4' - \mathbf{M}_4\|_1 \leqslant 12\gamma\Delta^2(\varepsilon + 2\Delta^3/m)$. ■

Next we consider the case $\mathcal{D} = \mathcal{P}$. Again Definition 2.1 is still valid and Lemma D.11 still holds by the same proof. Furthermore we show that the expectations are almost the same as in $\mathcal{D} = \mathcal{P}'$ since they can be computed by small-width branching programs:

Lemma D.18. *For any $u, w \in V[G]$, the quantities*

$$\sum_{v\in V[G]} \left|\mathbf{E}_{r\sim\widetilde{\mathcal{P}}',S}\left[X_{u,v}^S\right] - \mathbf{E}_{r\sim\widetilde{\mathcal{P}},S}\left[X_{u,v}^S\right]\right| \tag{D.7}$$

and

$$\sum_{v,x\in V[G]} \left|\mathbf{E}_{r\sim\widetilde{\mathcal{P}}',S,S'}\left[X_{u,v}^S X_{w,x}^{S'}\right] - \mathbf{E}_{r\sim\widetilde{\mathcal{P}},S,S'}\left[X_{u,v}^S X_{w,x}^{S'}\right]\right| \tag{D.8}$$

are bounded by ε', where $\widetilde{\mathcal{P}}$ (resp. $\widetilde{\mathcal{P}}'$) is the distribution of r induced by $\mathcal{P}$ (resp. $\mathcal{P}'$), S, S' in the subscripts are independent and have distribution $\mathcal{D}_{\gamma,k}$, and ε' is as in Protocol 2. The same statement holds with $X_{u,v}^S$ and $X_{w,x}^S$ replaced by $Y_{u,v}^S$ and $Y_{w,x}^S$ respectively.

Proof. It suffices to bound the quantities with S, S' and the auxiliary randomness $\{r_{i,u}\}$ fixed. Then (D.7) becomes $\sum_{v\in V[G]} \left|\mathbf{E}_{f\sim\mathcal{P}'}\left[X_{u,v}^S\right] - \mathbf{E}_{f\sim\mathcal{P}}\left[X_{u,v}^S\right]\right|$. Note that for both cases $f \sim \mathcal{P}$ and $f \sim \mathcal{P}'$ we can view f as a random variable determined by a sequence of seeds $y = (y_0, \ldots, y_{k-1}) \in \left(\{0,1\}^\ell\right)^k$. In the former case y is truly random whereas in the latter case it is generated by the PRG $\mathcal{G}'$. So we may rewrite (D.7) as

$$\sum_{v\in V[G]} \left|\mathbf{E}_{y\in\{0,1\}^{\ell'}}\left[X_{u,v}^S(\mathcal{G}'(y))\right] - \mathbf{E}_{y\in\left(\{0,1\}^\ell\right)^k}\left[X_{u,v}^S(y)\right]\right|,$$

where $X^S_{u,v}(y)$ denotes the value of $X^S_{u,v}$ determined by the sequence of seeds y. We claim that $X^S_{u,v}(y)$ is computed by a $(k, n, 2^\ell)$-branching program $\mathcal{B}$. More specifically, it holds that $X^S_{u,v}(y) = 1$ iff $\mathcal{B}(u, y) = v$. The branching program $\mathcal{B}$ is easy to construct: we use the set of nodes $[n] = V[G]$ in the ith level to keep track of the where the random walk is at the ith step. This location together with the seed y_i (which is used as the label of the outgoing edge in $\mathcal{B}$) uniquely determines the next node. Then the fact that y is generated by an ε'-PRG for $(T/2, n^2, 2^\ell)$-branching program $\mathcal{B}$ easily implies the bound. The bound for (D.8) is derived in the same way, except that we use a $(k, n^2, 2^\ell)$-branching program to keep track of two random walks simultaneously. The cases for $Y^S_{u,v}$ and $Y^S_{u,v}Y^S_{w,x}$ are the same, except that the time is reversed. ■

Now we are ready to prove a derandomized version of Lemma D.16.

Theorem D.19. *Suppose G has spectral gap α and irregularity β. Using Protocol 1 with distribution $\mathcal{D} = \mathcal{P}$, any node gets the rumor in $T = O(C \log n)$ rounds with probability at least $1 - n^{-2c}$, where $C = (1/\alpha) \cdot \beta^2 \max\{1, 1/(\alpha \cdot \Delta^{0.5-c})\}$ and $c > 0$ is an arbitrary small constant.*

Proof. Let $s \in V[G]$ be the initial node and fix a target node $w \in V[G]$. Let $c, \gamma, k, T, \boldsymbol{\pi}, \mathbf{u}, \mathbf{u}', \mathbf{v}, \mathbf{v}'$ be as in the proof of Lemma D.16 and $T = O(C \log n)$. Define $\bar{\mathbf{u}} = \mathbf{e}_s \mathbf{M}_1'^k$, $\bar{\mathbf{v}} = \mathbf{e}_{(s,s)} \mathbf{M}_2'^k$, $\bar{\mathbf{u}}' = \mathbf{e}_w \mathbf{M}_3'^k$, and $\bar{\mathbf{v}}' = \mathbf{e}_{(w,w)} \mathbf{M}_4'^k$, where $\mathbf{M}'_1, \ldots, \mathbf{M}'_4$ are as in Lemma D.17. Then

$$\|\bar{\mathbf{u}} - \mathbf{u}\|_1 = \left\|\mathbf{e}_s \left(\mathbf{M}_1'^k - \mathbf{M}_1^k\right)\right\|_1 \leqslant \left\|\mathbf{M}_1'^k - \mathbf{M}_1^k\right\|_1 \leqslant k \left\|\mathbf{M}'_1 - \mathbf{M}_1\right\|_1 \leqslant k\varepsilon_0$$

where $\varepsilon_0 = 12\gamma\Delta^2(\varepsilon + 2\Delta^3/m)$ (c.f. Lemma D.17). Here the second inequality holds by a simple induction on k. Similarly $\|\bar{\mathbf{u}}' - \mathbf{u}'\|_1, \|\bar{\mathbf{v}} - \mathbf{v}\|_1, \|\bar{\mathbf{v}}' - \mathbf{v}'\|_1 \leqslant k\varepsilon_0$. Define $\widetilde{\mathbf{u}}, \widetilde{\mathbf{u}}' \in \mathbb{R}^n$ and $\widetilde{\mathbf{v}}, \widetilde{\mathbf{v}}' \in \mathbb{R}^n \otimes \mathbb{R}^n$ such that $\widetilde{\mathbf{u}}_u = \mathbf{E}_{r,S}\left[X^S_{s,u}\right]$, $\widetilde{\mathbf{u}}'_u = \mathbf{E}_{r,S}\left[Y^S_{w,u}\right]$, $\widetilde{\mathbf{v}}_{u,v} = \mathbf{E}_{r,S,S'}\left[X^S_{s,u}X^S_{s,v}\right]$ and $\widetilde{\mathbf{v}}'_{u,v} = \mathbf{E}_{r,S,S'}\left[Y^S_{w,u}Y^S_{w,v}\right]$ where r, S and S' are independent with distributions $\widetilde{\mathcal{P}}$ (induced by $\mathcal{P}$), $\mathcal{D}_{\gamma,k}$ and $\mathcal{D}_{\gamma,k}$ respectively. Then Lemma D.17 and Lemma D.18 altogether imply that $\|\widetilde{\mathbf{u}} - \bar{\mathbf{u}}\|_1 \leqslant \varepsilon'$ and hence $\|\widetilde{\mathbf{u}} - \mathbf{u}\|_1 \leqslant k\varepsilon_0 + \varepsilon'$. Obviously we have $\|\widetilde{\mathbf{u}} - \mathbf{u}\|_\infty \leqslant 1$. Therefore by Hölder's inequality, we have $\|\widetilde{\mathbf{u}} - \mathbf{u}\|_2 \leqslant \sqrt{k\varepsilon_0 + \varepsilon'}$. Similarly,

$$\|\widetilde{\mathbf{u}}' - \mathbf{u}'\|_2 \leqslant \sqrt{k\varepsilon_0 + \varepsilon'}, \|\widetilde{\mathbf{v}} - \mathbf{v}\|_2 \leqslant \sqrt{k\varepsilon_0 + \varepsilon'}, \|\widetilde{\mathbf{v}}' - \mathbf{v}'\|_2 \leqslant \sqrt{k\varepsilon_0 + \varepsilon'}.$$

As shown in the proof of Lemma D.16, we have $\left\|\mathbf{u}^\perp\right\|_2, \left\|\mathbf{u}'^\perp\right\|_2 \leqslant n^{-1}$, and $\left\|\mathbf{v}^\perp\right\|_2, \left\|\mathbf{v}'^\perp\right\|_2 \leqslant n^{-(1+c)}$. Note that

$$\widetilde{\mathbf{u}}^\perp = \widetilde{\mathbf{u}} - \boldsymbol{\pi} = (\widetilde{\mathbf{u}} - \mathbf{u}) + (\mathbf{u} - \boldsymbol{\pi}) = (\widetilde{\mathbf{u}} - \mathbf{u}) + \mathbf{u}^\perp.$$

So we have $\left\|\widetilde{\mathbf{u}}^\perp\right\|_2 \leqslant \sqrt{k\varepsilon_0 + \varepsilon'} + n^{-1}$ and similarly $\left\|\widetilde{\mathbf{u}}'^\perp\right\|_2 \leqslant \sqrt{k\varepsilon_0 + \varepsilon'} + n^{-1}$, and $\left\|\widetilde{\mathbf{v}}^\perp\right\|_2, \left\|\widetilde{\mathbf{v}}'^\perp\right\|_2 \leqslant \sqrt{k\varepsilon_0 + \varepsilon'} + n^{-(1+c)}$.

By Lemma D.11, the probability that t gets the rumor in k rounds is lower bounded by

$$\begin{aligned}
&\frac{\sum_{u,v \in V[G]} \langle \widetilde{\mathbf{u}}, \mathbf{e}_u\rangle \langle \widetilde{\mathbf{u}}, \mathbf{e}_v\rangle \langle \widetilde{\mathbf{u}}', \mathbf{e}_u\rangle \langle \widetilde{\mathbf{u}}', \mathbf{e}_v\rangle}{\sum_{u,v \in V[G]} \left\langle \widetilde{\mathbf{v}}, \mathbf{e}_{(u,v)}\right\rangle \left\langle \widetilde{\mathbf{v}}', \mathbf{e}_{(u,v)}\right\rangle} = \frac{\left\langle \widetilde{\mathbf{u}}, \widetilde{\mathbf{u}}'\right\rangle^2}{\left\langle \widetilde{\mathbf{v}}, \widetilde{\mathbf{v}}'\right\rangle} \\
&= \frac{\left(\langle \boldsymbol{\pi}, \boldsymbol{\pi}\rangle + \left\langle \widetilde{\mathbf{u}}^\perp, \boldsymbol{\pi}\right\rangle + \left\langle \boldsymbol{\pi}, \widetilde{\mathbf{u}}'^\perp\right\rangle + \left\langle \widetilde{\mathbf{u}}^\perp, \widetilde{\mathbf{u}}'^\perp\right\rangle\right)^2}{\langle \boldsymbol{\pi} \otimes \boldsymbol{\pi}, \boldsymbol{\pi} \otimes \boldsymbol{\pi}\rangle + \left\langle \widetilde{\mathbf{v}}^\perp, \boldsymbol{\pi} \otimes \boldsymbol{\pi}\right\rangle + \left\langle \boldsymbol{\pi} \otimes \boldsymbol{\pi}, \widetilde{\mathbf{v}}'^\perp\right\rangle + \left\langle \widetilde{\mathbf{v}}^\perp, \widetilde{\mathbf{v}}'^\perp\right\rangle} \\
&= \frac{\left(\langle \boldsymbol{\pi}, \boldsymbol{\pi}\rangle + \left\langle \widetilde{\mathbf{u}}^\perp, \widetilde{\mathbf{u}}'^\perp\right\rangle\right)^2}{\langle \boldsymbol{\pi} \otimes \boldsymbol{\pi}, \boldsymbol{\pi} \otimes \boldsymbol{\pi}\rangle + \left\langle \widetilde{\mathbf{v}}^\perp, \widetilde{\mathbf{v}}'^\perp\right\rangle} \\
&= \frac{\left(1/n + \left\langle \widetilde{\mathbf{u}}^\perp, \widetilde{\mathbf{u}}'^\perp\right\rangle\right)^2}{1/n^2 + \left\langle \widetilde{\mathbf{v}}^\perp, \widetilde{\mathbf{v}}'^\perp\right\rangle}.
\end{aligned} \tag{D.9}$$

We have

$$\left|\left\langle \widetilde{\mathbf{u}}^{\perp}, \widetilde{\mathbf{u}}'^{\perp} \right\rangle\right| \leqslant \left\|\widetilde{\mathbf{u}}^{\perp}\right\|_2 \left\|\widetilde{\mathbf{u}}'^{\perp}\right\|_2 = O\left(k\varepsilon_0 + \varepsilon' + n^{-2}\right),$$
$$\left|\left\langle \widetilde{\mathbf{v}}^{\perp}, \widetilde{\mathbf{v}}'^{\perp} \right\rangle\right| \leqslant \left\|\widetilde{\mathbf{v}}^{\perp}\right\|_2 \left\|\widetilde{\mathbf{v}}'^{\perp}\right\|_2 = O\left(k\varepsilon_0 + \varepsilon' + n^{-(2+2c)}\right).$$

So (D.9) is lower bounded by $1 - O(n^2(k\varepsilon_0 + \varepsilon') + n^{-2c})$ where $\varepsilon_0 = 12\gamma\Delta^2(\varepsilon + 2\Delta^3/m)$. The claim follows since we pick $\varepsilon^{-1}, \varepsilon'^{-1}, m = n^{\Theta(1)}$ sufficiently large in Protocol 2. ■

By repeating the protocol $O(1)$ times and apply the union bound, we obtain Theorem 2.7.

E Simplified Protocol with $O(\Delta)$ Preprocessing Time

E.1 Description of the Protocol

Protocol 5. *Let m be a prime power. Pick the following objects:*

- *an explicit pairwise independent generator $\mathcal{G} = (\mathcal{G}_0, \ldots, \mathcal{G}_{n-1}) : \{0,1\}^{\ell} \to [m]^n$ with seed length ℓ, and*
- *an explicit ε-PRG $\mathcal{G}' = (\mathcal{G}'_0, \ldots, \mathcal{G}'_{T-1}) : \{0,1\}^{\ell'} \to \left(\{0,1\}^{\ell}\right)^T$ for $(T, n^2, 2^{\ell})$-branching programs with seed length ℓ'*

where $\varepsilon^{-1}, m = n^{\Theta(1)}$ are sufficiently large.

The initial node having the rumor independently chooses a random string $x \in \{0,1\}^{\ell'}$ which is appended with the rumor and sent to other nodes. Once one node gets the rumor, it gets the ID u. Let $y = (y_0, \ldots, y_{T-1})$ be the sequence of seeds generated by $\mathcal{G}'$, i.e., $y_i = \mathcal{G}'_i(x)$. For $i \in [T]$ and $u \in V[G]$, define $(w_{u,i}, z_{u,i}) = \mathcal{G}_u(y_i) \bmod 4\Delta \in [2\Delta] \times \{\text{active}, \text{inactive}\}$. We say u is active in the ith round if $z_{u,i}$ is active, and otherwise inactive. We say u selects *v if v is the $w_{u,i}$th neighbor of u. In the ith round, an informed node u sends the rumor to the unique neighbor v (if exist) if $\{u, v\}$ is a good pair, where we call $\{u, v\}$ is a good pair if (i) u is active, v is inactive, and u is the unique node selecting v, or (ii) the same holds with u and v swapped.*

Checking the conditions requires u and v knowing its index in the lists of its neighbors as well as the IDs of its neighbors. One can deterministically use $O(\Delta)$ preprocessing time to guarantee this assumption. Then Condition (ii) can be checked directly by u. For Condition (i), note that an active node u can send the rumor and the seed to its unique inactive neighbor v specified by $w_{i,u}$ and then v can check if the condition is met, i.e., if u is the unique node selecting v. [7]

Theorem E.1. *Let G be any graph with spectral gap α and irregularity β. Then Protocol 5 uses 2ℓ random bits, and with high probability informs all nodes of G in $T = O(\beta^2\alpha^{-1}\log n)$ rounds.*

As a consequence, we obtain the following reduction:

Corollary E.2. *Assume each node knows its index in the lists of its neighbors as well as the IDs of its neighbors. Then the following statements hold:*

1. *Given an explicit ε-PRG for $(T/2, n^2, 2^{\ell})$-branching programs with seed length ℓ', where $\varepsilon^{-1} = n^{\Theta(1)}$ and $\ell = O(\log n)$ are sufficiently large, there exists an explicit protocol using $2\ell'$ random bits, and with high probability informs all nodes in $T = O((1/\alpha) \cdot \beta^2 \log n)$ rounds.*

[7] The uniqueness requirement in Condition (i) is necessary only for analyzing the associated averaging algorithm. For the sake of rumor spreading, dropping the requirement only make the rumor spread faster.

2. *In particular, given an explicit ε-PRG for $(T/2, n^2, \varepsilon)$-branching programs with seed length $O(\log n)$ where $\varepsilon^{-1} = n^{\Theta(1)}$ is sufficiently large, there exists an explicit protocol using $O(\log n)$ random bits, and with high probability informs all nodes in $T = O((1/\alpha)\cdot\beta^2 \log n)$ rounds.*

Combining the reduction above with known explicit constructions of PRGs (Theorem D.8), we obtain Theorem 1.2.

We study Protocol 5 by analyzing the following associated averaging protocol, which is closely related to other gossip processes, e.g. random-matching model of load balancing processes. In the following, let $\mathbf{v}(k) \in \mathbb{R}^{V[G]}$ denote the values of nodes after k rounds.

Protocol 6 (Averaging Protocol). *Each node u has a value $\mathbf{v}(0)_u$ specified by the distribution $\mathbf{v}(0) = \mathbf{e}_s$ where s is the initial node. Proceed as in Protocol 5. When node u sends the rumor to node v, set the both values of u and v as the average of their original values.*

We define the *averaging time* $\tau_{\mathsf{avg}}(\delta)$ of the protocol as the smallest $k \in \mathbb{N}$ such that $\mathbf{Pr}\left[\, \|\mathbf{v}(k)^{\perp}\|_2 < \delta \,\right] > 1 - \delta$ for any distribution $\mathbf{v}$, or ∞ if there is no such k.

Theorem E.3. *For $\delta > 0$, assume $2\varepsilon < \delta^2$ where ε is as in Protocol 5. Then Protocol 6 uses $2\ell'$ random bits with $\tau_{\mathsf{avg}}(\delta) = O((1/\alpha) \cdot \beta^2 \log(1/\delta))$.*

Theorem E.1 is simple corollary of Theorem E.3 with $\delta = 1/n$, since when $\|\mathbf{v}(k)^{\perp}\|_2 < 1/n$ then all $\mathbf{v}(k)_u$ must be nonzero, and $\mathbf{v}(k)_u \neq 0$ implies that u is informed in k rounds.

In Theorem E.3 we only consider initial values specified by $\mathbf{v}(0) = \mathbf{e}_s$. Assuming $\varepsilon/\delta^2 = n^{-\Theta(1)}$ is sufficiently small, it is easy to establish a upper bound $O(1/\alpha \cdot \beta^2(\log n + \log(1/\delta)))$ on the averaging time regarding a general distribution $\mathbf{v}(0)$: first use $T = O(1/\alpha \cdot \beta^2 \log(1/\delta))$ rounds to inform all the nodes with high probability. Then set the new initial values $\mathbf{v}'(0) = \mathbf{v}(T)$, and run the averaging protocol for another $O(1/\alpha \cdot \beta^2(\log n + \log(1/\delta)))$ rounds. The process with initial value distribution $\mathbf{v}'(0)$ can be viewed as a convex combination of those with initial value distribution $\mathbf{e}_u$, $u \in V[G]$ (note that each node u is already informed). With high probability, for all initial value distributions $\mathbf{e}_u$, the values converge to the average up to ℓ_2-distance δ. So the same is true for $\mathbf{v}'(0)$.

E.2 Analysis of the Protocol

For $x \in \{0,1\}^{\ell}$, define the following matrix

$$\mathbf{M}(x)_{uv} = \begin{cases} 1/2 & u \neq v \text{ and } \{u,v\} \text{ is a good pair,} \\ 1/2 & u = v \text{ and } \{u,v'\} \text{ is a good pair for some } v' \in V[G], \\ 1 & u = v \text{ and } \{u,v'\} \text{ is not a good pair for any } v' \in V[G], \\ 0 & u \neq v \text{ and } \{u,v\} \text{ is not a good pair} \end{cases}$$

where the set of good pairs are determined by the seed $y_i = x$ (see Protocol 5, where the definition of good pairs are the same for all round number i). It is easy to check that $\mathbf{M}(x)$ is doubly stochastic, symmetric and $\mathbf{M}(x)^2 = \mathbf{M}(x)$ for all $x \in \{0,1\}^{\ell}$. Moreover it characterizes the averaging operations using the seed $y_i = x$.

Lemma E.4. *It holds that $\mathbf{v}(i+1) = \mathbf{v}(i)\mathbf{M}(y_i)$ for any $i \in [T]$.*

Proof. By definition, $\mathbf{M}(y_i)$ acts on $\mathbb{R}^{V[G]}$ by averaging the values of u and v for each good pair $\{u,v\}$. Protocol 6 guarantees that averaging operations are performed for each good pair $\{u,v\}$, where u or v are already informed. If neither u nor v is informed, their values are both zero (by induction with the base case $\mathbf{v}(0) = \mathbf{e}_s$) and hence the averaging operation between them can be safely ignored. ∎

Let $\mathbf{M} = \mathbf{E}_{x\in\{0,1\}^\ell}[\,\mathbf{M}(x)\,]$. Then $\mathbf{M}$ is doubly-stochastic. We have the following lemma:

Lemma E.5. $\mathbf{M}_{uv} \geqslant c \cdot \mathcal{L}_{1/2}\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)_{uv}$ *for some constant* $c \in (0,1)$.

Proof. Each edge $\{u,v\}$ with $u \neq v$ is a good pair if either of the two mutually exclusive conditions (c.f. Protocol 5) is met. The first one holds with probability at least

$$\mathbf{Pr}_{x\in\{0,1\}^\ell}[\,u \text{ is active and selects } v\,] - \sum_{u'\in N(v)\setminus\{u\}} \mathbf{Pr}_{x\in\{0,1\}^\ell}\left[\,u \text{ is active and both } u,u' \text{ select } v\,\right]$$

taken over the seed $y_i = x$. As $\mathcal{G}$ is a pairwise independent generator, by Lemma D.3, this probability is lower bounded by $\left(\frac{1}{4\Delta} - \frac{2}{m}\right) - \Delta \cdot \left(\frac{1}{4\Delta}\cdot\frac{1}{2\Delta} + \frac{2}{m}\right) \geqslant \frac{c}{2\Delta}$ for some $c > 0$ and $m = \Omega(\Delta^2)$. The case for the second condition is the same. So $\{u,v\}$ is a good pair with probability at least $\frac{c}{\Delta}$. Note that $\mathbf{M}(x)_{uv} = 1/2$ whenever $\{u,v\}$ is a good pair. Therefore

$$\mathbf{M}_{uv} = \mathbf{E}_{x\in\{0,1\}^\ell}[\,\mathbf{M}(x)_{uv}\,] \geqslant \frac{c}{2\Delta} = c\mathcal{L}_{1/2}\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)_{uv}.$$

For $u = v$, note that $\mathbf{M}_{uv} \geqslant 1/2$ by definition and $\mathcal{L}_{1/2}\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)_{uv} \leqslant 1$. ∎

Again let $\boldsymbol{\pi} \in \mathbb{R}^{V[G]}$ denote the uniform distribution over $V[G]$.

Lemma E.6. *For any* $\mathbf{v} \in \mathbb{R}^{V[G]}$ *orthogonal to* $\boldsymbol{\pi}$*, it holds that* $0 \leqslant \mathbf{E}_{x\in\{0,1\}^\ell}[\,\|\mathbf{v}\mathbf{M}(x)\|_2\,] \leqslant (1 - c\beta^{-2}\alpha)\|\mathbf{v}\|_2$ *for some constant* $c \in (0,1)$.

Proof. The non-negativity is obvious. For the upper bound, we have

$$\begin{aligned}
\mathbf{E}_{x\in\{0,1\}^\ell}[\,\|\mathbf{v}\mathbf{M}(x)\|_2\,] &= \mathbf{E}_{x\in\{0,1\}^\ell}[\,\mathbf{v}\mathbf{M}(x)\mathbf{M}(x)^\intercal\mathbf{v}^\intercal\,] \\
&= \mathbf{v}\mathbf{E}_{x\in\{0,1\}^\ell}[\,\mathbf{M}(x)\mathbf{M}(x)^\intercal\,]\,\mathbf{v}^\intercal \\
&= \mathbf{v}\mathbf{E}_{x\in\{0,1\}^\ell}[\,\mathbf{M}(x)\,]\,\mathbf{v}^\intercal \\
&= \mathbf{v}\mathbf{M}\mathbf{v}^\intercal.
\end{aligned}$$

Let $\mathbf{M}' = \mathbf{M} - c\cdot\mathcal{L}_{1/2}(\mathbf{M}_{\mathrm{Reg}(G)})$ where c is as in Lemma E.5. Then $\mathbf{M}'$ is a non-negative matrix by Lemma E.5. As both $\mathbf{M}$ and $\mathcal{L}_{1/2}\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)$ are doubly-stochastic, so is $\mathbf{M}'/(1-c)$. Then $\lambda_{\max}(\mathbf{M}') \leqslant \|\mathbf{M}'\|_2 \leqslant 1 - c$. Note that $\lambda_{\max}\left(\mathcal{L}_{1/2}\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)\right) \leqslant 1 - \beta^{-2}\alpha/2$. Therefore

$$\lambda_{\max}(\mathbf{M}) \leqslant \lambda_{\max}(\mathbf{M}') + c\cdot\lambda_{\max}\left(\mathcal{L}_{1/2}\left(\mathbf{M}_{\mathrm{Reg}(G)}\right)\right) \leqslant 1 - (c/2)\beta^{-2}\alpha$$

and the claim follows. ∎

Lemma E.7. *For any* $\mathbf{v} \in \mathbb{R}^n$ *orthogonal to* $\boldsymbol{\pi}$ *and* $k \in [T]$*, it holds that*

$$\mathbf{E}_{y_0,\ldots,y_{k-1}\in\{0,1\}^\ell}\left[\,\left\|\mathbf{v}\prod_{i=0}^{k-1}\mathbf{M}(y_i)\right\|_2\,\right] \leqslant (1 - c\beta^{-2}\alpha)^k\|\mathbf{v}\|_2$$

for some constant $c \in (0,1)$.

Proof. Induct on k. The claim is trivial for $k = 0$. For $k > 0$, assume the claim holds for $k' < k$. Let $\mathbf{v} \in \mathbb{R}^n$ be a vector orthogonal to $\boldsymbol{\pi}$, and define $\mathbf{v}' = \mathbf{v}\prod_{i=0}^{k-2}\mathbf{M}(y_i)$. Then $\mathbf{v}'$ is also orthogonal to $\boldsymbol{\pi}$. So

$$\begin{aligned}
\mathbf{E}_{y_0,\ldots,y_{k-1}\in\{0,1\}^\ell}\left[\,\left\|\mathbf{v}\prod_{i=0}^{k-1}\mathbf{M}(y_i)\right\|_2\,\right] &= \mathbf{E}_{y_0,\ldots,y_{k-2}\in\{0,1\}^\ell}\left[\,\mathbf{E}_{y_{k-1}\in\{0,1\}^\ell}\left[\,\left\|\mathbf{v}'\mathbf{M}(y_{k-1})\right\|_2\,\right]\,\right] \\
&\leqslant \mathbf{E}_{y_0,\ldots,y_{k-2}\in\{0,1\}^\ell}\left[\,(1 - c\beta^{-2}\alpha)\left\|\mathbf{v}'\right\|_2\,\right] \\
&\leqslant (1 - c\beta^{-2}\alpha)^k\|\mathbf{v}\|_2.
\end{aligned}$$

The first inequality uses Lemma E.6 and the second one uses the induction hypothesis. ∎

Let $\mathcal{P}$ be the distribution of $y = (y_0, \ldots, y_{T-1})$ in Protocol 5. Then we have

Lemma E.8. *For any $u \in V[G]$,*

$$\left| \mathbf{E}_{y\sim\mathcal{P}} \left[\left\| \mathbf{e}_u \prod_{i=0}^{T-1} \mathbf{M}(y_i) \right\|_2 \right] - \mathbf{E}_{y\in\left(\{0,1\}^\ell\right)^T} \left[\left\| \mathbf{e}_u \prod_{i=0}^{T-1} \mathbf{M}(y_i) \right\|_2 \right] \right| \leqslant \varepsilon$$

where ε is as in Protocol 5.

Proof. For $x \in \{0,1\}^\ell$, write $\mathbf{M}(x) = \frac{1}{2}\mathbf{M}_{\text{lazy}}(x) + \frac{1}{2}\mathbf{M}_{\text{non-lazy}}(x)$ where $\mathbf{M}_{\text{lazy}}(x)$ is simply the identity matrix $\mathbf{I}$, and $\mathbf{M}_{\text{non-lazy}}(x)$ is the following permutation matrix:

$$\left(\mathbf{M}_{\text{non-lazy}}(x)\right)_{uv} = \begin{cases} 1 & u \neq v \text{ and } \{u,v\} \text{ is a good pair} \\ 0 & u = v \text{ and } \{u,v'\} \text{ is a good pair for some } v' \in V[G], \\ 1 & u = v \text{ and } \{u,v'\} \text{ is not a good pair for any } v' \in V[G], \\ 0 & u \neq v \text{ and } \{u,v\} \text{ is not a good pair.} \end{cases}$$

As before, let $\mathcal{C}_T = \{\text{lazy}, \text{non-lazy}\}^T$. Note that for any $y = (y_0, \ldots, y_{T-1}) \in \left(\{0,1\}^\ell\right)^T$, we have

$$\begin{aligned} \left\| \mathbf{e}_u \prod_{i=0}^{T-1} \mathbf{M}(y_i) \right\|_2 &= \mathbf{e}_u \left(\prod_{i=0}^{T-1} \mathbf{M}(y_i) \right) \left(\prod_{i=0}^{T-1} \mathbf{M}(y_i) \right)^\intercal \mathbf{e}_u^\intercal \\ &= 2^{-2T} \sum_{c,c'\in\mathcal{C}_T} \mathbf{e}_u \left(\prod_{i=0}^{T-1} \mathbf{M}_{c_i}(y_i) \right) \left(\prod_{i=0}^{T-1} \mathbf{M}_{c'_i}(y_i) \right)^\intercal \mathbf{e}_u^\intercal \\ &= 2^{-2T} \sum_{c,c'\in\mathcal{C}_T, v\in V[G]} (\mathbf{e}_u \otimes \mathbf{e}_u) \prod_{i=0}^{T-1} \left(\mathbf{M}_{c_i}(y_i) \otimes \mathbf{M}_{c'_i}(y_i) \right) (\mathbf{e}_v \otimes \mathbf{e}_v)^\intercal. \end{aligned}$$

For any $c, c' \in \mathcal{C}_T$, it is easy to construct a $(T, n^2, 2^\ell)$-branching program $\mathcal{B}_{c,c'}$ that has state set $V[G] \times V[G]$, such that for any node $v \in V[G]$ and input $y = (y_0, \ldots, y_{T-1})$, it holds that $\mathcal{B}_{c,c'}((u,u), y) = (v,v)$ (resp. $\mathcal{B}_{c,c'}((u,u), y) \neq (v,v)$) iff

$$(\mathbf{e}_u \otimes \mathbf{e}_u) \prod_{i=0}^{T-1} \left(\mathbf{M}_{c_i}(y_i) \otimes \mathbf{M}_{c'_i}(y_i) \right) (\mathbf{e}_v \otimes \mathbf{e}_v)^\intercal.$$

equals 1 (resp. 0). More specifically, The transition matrix between the ith and the $(i+1)$st layer of $\mathcal{B}_{c,c'}$ with edge label y_i is just $\mathbf{M}_{c_i}(y_i) \otimes \mathbf{M}_{c'_i}(y_i)$. Then the absolute difference between $\mathbf{E}_{y\sim\mathcal{P}} \left[\left\| \mathbf{v} \prod_{i\in[k]} \mathbf{M}(y_i) \right\|_2 \right]$ and $\mathbf{E}_{y\in\left(\{0,1\}^\ell\right)^T} \left[\left\| \mathbf{v} \prod_{i\in[k]} \mathbf{M}(y_i) \right\|_2 \right]$ is bounded by

$$2^{-2T} \sum_{c,c'\in\mathcal{C}_T, v\in V[G]} \left| \mathbf{Pr}_{y\sim\mathcal{P}} \left[\mathcal{B}_{c,c'}((u,u), y) = (v,v) \right] - \mathbf{Pr}_{y\in\left(\{0,1\}^\ell\right)^T} \left[\mathcal{B}_{c,c'}((u,u), y) = (v,v) \right] \right|$$

which is bounded by ε since $\mathcal{G}$ is an ε-PRG for $(T, n^2, 2^\ell)$-branching programs. ∎

Proof of Theorem E.3. By Lemma E.7, we have

$$\mathbf{E}_{y\in\left(\{0,1\}^\ell\right)^T} \left[\left\| \mathbf{e}_s^\perp \prod_{i=0}^{T-1} \mathbf{M}(y_i) \right\|_2 \right] \leqslant (1 - c\beta^{-2}\alpha)^T.$$

Combining this with Lemma E.8 and using the fact that $\|\mathbf{v}\|_2 = \|\mathbf{v}^\perp\|_2 + \|\boldsymbol{\pi}\|_2$ for any distribution $\mathbf{v}$, we obtain

$$\mathbf{E}_{y\sim\mathcal{P}}\left[\left\|\left(\mathbf{e}_s\prod_{i=0}^{T-1}\mathbf{M}(y_i)\right)^\perp\right\|_2\right] = \mathbf{E}_{y\sim\mathcal{P}}\left[\left\|\mathbf{e}_s^\perp\prod_{i=0}^{T-1}\mathbf{M}(y_i)\right\|_2\right] \leqslant (1-c\beta^{-2}\alpha)^T + \varepsilon < \delta^2$$

for sufficiently large $T = O(\beta^2\alpha^{-1}\log\delta^{-1})$. The claim then follows from Lemma E.4 and the Markov's inequality. ■

F Omitted Details in Section 3

F.1 Preliminaries

In this subsection we list all necessary definitions and results that are used to construct the protocols in Section 3.

Unbalanced Expanders with Near-Optimal Expansion We consider the following kind of left-regular bipartite graphs.

Definition F.1. *Let $\Gamma : [N]\times[D] \to \bigsqcup_{i\in[D]}[M_i]$ be a function where $\Gamma(x,y)\in[M_y]$ for any $x\in[N]$, $y\in[D]$. Function Γ specifies a left-degree D bipartite graph with left vertex set $[N]$ and right vertex set $\bigsqcup_{i\in[D]}[M_i]$ in the following way: for $x\in[N]$ and $y\in[D]$, the yth neighbor of x is given by $\Gamma(x,y)$.*

We are interested in graphs Γ exhibiting excellent expansion properties. This leads to the notion of unbalanced expanders [27, 37].

Definition F.2 (Unbalanced expanders [27, 37])**.** *Let $\Gamma : [N]\times[D] \to \bigsqcup_{i\in[D]}[M_i]$ be as in Definition F.1. We call Γ a (K,A)-expander if for any set $S\subseteq[N]$ of size K, it holds that $|N(S)|\geqslant AK$. We call Γ a $(\leqslant K,A)$-expander if it is a (K',A)-expander for all $K'\leqslant K$.* [8]

In particular we are interested in (K,A)-expanders, where the parameter $A=(1-\varepsilon)D$ for small ε, i.e. for any subset S of size K from the left set $[N]$, there is almost no collision among the neighbors of nodes in S. Explicit constructions of such unbalanced expanders with near-optimal expansion are known.

Theorem F.3 ([27])**.** *For any $N\in\mathbb{N}$, $K\leqslant N$, and $\varepsilon>0$, there is an explicit $(K,(1-\varepsilon)D)$-expander $\Gamma : [N]\times[D]\to\bigsqcup_{i\in[D]}[M_i]$ with $D=\left(\frac{\log N}{\varepsilon}\right)^{O(1)}$ and $M_0=\cdots=M_{D-1}\leqslant\max\left\{D,K^{O(1)}\right\}$.*

Assume that $\Gamma : [N]\times[D]\to\bigsqcup_{i\in[D]}[M_i]$ is a $(K,(1-\varepsilon)D)$-expander. We consider the map $\Gamma(\cdot,U)$ applied on any K elements of $[N]$ where U is uniformly distributed over $[D]$. The following lemma states that with high probability these K elements are mapped into $\bigsqcup_{i\in[D]}[M_i]$ with almost no collision.

Lemma F.4. *Let $\Gamma : [N]\times[D]\to\bigsqcup_{i\in[D]}[M_i]$ be a $(K,(1-\varepsilon)D)$-expander. Let S be a subset of $[N]$ of size K. Then for at least $(1-\sqrt{\varepsilon})$-fraction of $y\in[D]$, it holds that $|\{\Gamma(x,y) : x\in S\}|\geqslant(1-\sqrt{\varepsilon})K$.*

Proof. The size of $N(S)=\bigsqcup_{y\in[D]}\{\Gamma(x,y) : x\in S\}$ is at least $(1-\varepsilon)DK$ as Γ is a $(K,(1-\varepsilon)D)$-expander. So $\mathbf{E}_y\left[|\{\Gamma(x,y) : x\in S\}|\right]\geqslant(1-\varepsilon)K$ with y uniformly distributed over $[D]$. Also note that $|\{\Gamma(x,y) : x\in S\}|\leqslant|S|=K$ for any $y\in[D]$. Applying Markov's inequality on $K-|\{\Gamma(x,y) : x\in S\}|$, we have $\mathbf{Pr}_y[|\{\Gamma(x,y) : x\in S\}|<(1-\sqrt{\varepsilon})K]\leqslant\sqrt{\varepsilon}$. ■

[8] The definition here is slightly different from [27, 37] as we require $\Gamma(x,y)\in[M_y]$. This is analogous to the difference between standard and strong condensers.

F.2 Analysis of Protocol 3

We start by analyzing a single round t and see the properties of our protocol. Let I_t be the set of informed nodes after round t, and U_t the set of uninformed nodes after round t. Remember that all the random choices in round t are determined by (x_t, y_t).

We need the following lemma:

Lemma F.5. *Fix any round $0 \leqslant t < T$. For any $u \in U_t$, $v \in I_t$, let $X_{v\to u}$ be the boolean random variable whose value is 1 iff v informs u in round $t+1$. Then it holds that*

1. $|\mathbf{E}[X_{v\to u}] - 1/\Delta| \leqslant \varepsilon$ *for any $u \in U_t$, $v \in I_t$;*
2. $\mathbf{Cov}[X_{v\to u}, X_{v'\to u'}] \leqslant \varepsilon$ *for any $u, u' \in U_t$, $v, v' \in I_t$ satisfying $(u, v) \neq (u', v')$.*

Proof. For any $u \in U_t$ and $v \in I_t$, suppose the index of u in the adjacency list of v is z. By construction, $X_{v\to u}$ equals 1 iff $\mathcal{G}_{\Gamma(v,x_t)}(y_t) \bmod \Delta = z$. Fix x_t. The fact that $\mathcal{G}$ is a pairwise independent generator together with Lemma D.3 shows that $|\mathbf{E}[X_{v\to u}] - 1/\Delta| \leqslant 2/m \leqslant \varepsilon$.

For any $u, u' \in U_t$ and $v, v' \in I_t$, first assume $v \neq v'$. Suppose the index of u (resp. u') in the adjacency list of v (resp. v') is z (resp z'). By construction, $X_{v\to u}$ equals 1 iff $\mathcal{G}_{\Gamma(v,x_t)}(y_t) \bmod \Delta = z$, and similarly for $X_{v'\to u'}$. By Lemma F.4 and the fact that Γ is a $(K, (1-\varepsilon^2/4)D)$-expander, the event $|\{\Gamma(v, x_t), \Gamma(v', x_t)\}| \geqslant (1-\varepsilon/2)\cdot 2 > 1$ occurs with probability at least $1-\varepsilon/2$ over the choices of x_t. Condition on any x_t such that this event occurs. We have $\Gamma(v, x_t) \neq \Gamma(v', x_t)$. Using the fact that $\mathcal{G}$ is pairwise independent together with Lemma D.3, we have $\mathbf{Cov}[X_{v\to u}, X_{v'\to u'}] \leqslant 2/m$. For the other choices of x_t, we have $\mathbf{Cov}[X_{v\to u}, X_{v'\to u'}] \leqslant 1$ since $X_{v\to u}, X_{v'\to u'}$ are boolean. Therefore $\mathbf{Cov}[X_{v\to u}, X_{v'\to u'}] \leqslant (1-\varepsilon/2)(2/m) + (\varepsilon/2) \leqslant \varepsilon$ for random x_t.

Now assume $v = v'$ and hence $u \neq u'$. We have

$$\begin{aligned}\mathbf{Cov}[X_{v\to u}, X_{v\to u'}] &= \mathbf{E}[X_{v\to u}\cdot X_{v\to u'}] - \mathbf{E}[X_{v\to u}]\cdot\mathbf{E}[X_{v\to u'}] \\ &= 0 - \mathbf{E}[X_{v\to u}]\cdot\mathbf{E}[X_{v\to u'}] \leqslant 0.\end{aligned}$$

■

Next we prove the following lemma:

Lemma F.6. *Fix a round $0 \leqslant t < T$ and the set I_t of informed nodes before round $t+1$. Fix also an arbitrary set of edges $F \subseteq E(I_t, U_t)$. Let J be the set of nodes that become informed in round $t+1$ if we consider only transmissions of the rumor along the edges in F.*

1. $\mathbf{Pr}[J \neq \emptyset] \geqslant c_1 \min\{|F|/\Delta, 1\}$ *for some constant $c_1 > 0$.*
2. *If $|F| = \Omega(\Delta)$ then $\mathbf{Pr}[|J| \geqslant c_2|F|/\Delta] \geqslant c_3$ for some constant $c_2, c_3 > 0$.*

Proof. Let $X_{v\to u}$ be the boolean random variable whose value is 1 iff v informs u in round $t+1$.

We first prove (1). Let $k = |F|$ and suppose $F = \{(v_0, u_0), \ldots, (v_{k-1}, u_{k-1})\}$. Let $X = \sum_{i\in[k]} X_{v_i\to u_i}$. Then by Cauchy-Schwarz inequality, $\mathbf{E}[\mathbf{1}_{X>0}] \geqslant (\mathbf{E}[X])^2/\mathbf{E}[X^2]$. By Lemma F.5, it holds that

$$\mathbf{E}[X] = \sum_{i\in[k]} \mathbf{E}[X_{v_i\to u_i}] \geqslant k(1/\Delta - \varepsilon) = \Omega(|F|/\Delta)$$

and

$$\begin{aligned}\mathbf{E}[X^2] &= \sum_{i,j\in[k]} \mathbf{E}[X_{v_i\to u_i} X_{v_j\to u_j}] \\ &= \sum_{i\in[k]} \mathbf{E}[X_{v_i\to u_i}] + \sum_{\substack{i,j\in[k]\\ i\neq j}} (\mathbf{E}[X_{v_i\to u_i}]\,\mathbf{E}[X_{v_j\to u_j}] + \mathbf{Cov}[X_{v_i\to u_i}, X_{v_j\to u_j}]) \\ &\leqslant k(1/\Delta + \varepsilon) + (k^2-k)((1/\Delta+\varepsilon)^2 + \varepsilon) = O(|F|/\Delta + |F|^2/\Delta^2)\end{aligned}$$

where we use the condition that $\varepsilon = \Delta^{-\Theta(1)}$ is sufficiently small. So

$$\mathbf{Pr}[J \neq \emptyset] = \mathbf{E}[\mathbf{1}_{X>0}] \geqslant (\mathbf{E}[X])^2/\mathbf{E}\left[X^2\right] = \Omega(\min\{|F|/\Delta, 1\}),$$

and the first statement follows.

Next we prove the second statement. For $u \in U_t$, let F_u be the set of edges in F incident to u, Z_u be the boolean random variable whose value is 1 iff u is informed in round $t+1$ via edges in F_u, and $X_u = \sum_{(v,u)\in F_u} X_{v\to u}$. So $Z_u = \mathbf{1}_{X_u>0}$ and $|J| = \sum_{u\in U_t} Z_u$. For $u \in U_t$, $\mathbf{E}[Z_u] = \mathbf{E}[\mathbf{1}_{X_u>0}] \geqslant (\mathbf{E}[X_u])^2/\mathbf{E}\left[X_u^2\right] = \Omega(|F_u|/\Delta)$ by a similar argument as above. So $\mathbf{E}[|J|] = \Omega(\sum_{u\in U_t}|F_u|/\Delta) = \Omega(|F|/\Delta)$. Suppose $\mathbf{E}[|J|] \geqslant c|F|/\Delta$ for constant $c > 0$.

On the other hand, for any $c_2 \geqslant 0$, we have

$$\begin{aligned}\mathbf{E}[|J|] &= \mathbf{E}\left[\mathbf{1}_{|J|\geqslant c_2|F|/\Delta}\cdot|J|\right] + \mathbf{E}\left[\mathbf{1}_{|J|<c_2|F|/\Delta}\cdot|J|\right]\\ &\leqslant \mathbf{E}\left[\mathbf{1}_{|J|\geqslant c_2|F|/\Delta}\cdot|J|\right] + \mathbf{E}\left[\mathbf{1}_{|J|<c_2|F|/\Delta}\right]\cdot c_2|F|/\Delta\end{aligned}$$

and hence $\mathbf{E}\left[\mathbf{1}_{|J|\geqslant c_2|F|/\Delta}\cdot|J|\right] \geqslant \mathbf{E}[|J|] - \mathbf{E}\left[\mathbf{1}_{|J|<c_2|F|/\Delta}\right]\cdot c_2|F|/\Delta \geqslant (c-c_2)|F|/\Delta$. Pick $c_2 = c/2$. By Cauchy-Schwarz inequalty, we have

$$\mathbf{Pr}[|J| \geqslant c_2|F|/\Delta] = \mathbf{E}\left[\mathbf{1}_{|J|\geqslant c_2|F|/\Delta}\right] \geqslant \frac{\left(\mathbf{E}\left[\mathbf{1}_{|J|\geqslant c_2|F|/\Delta}\cdot|J|\right]\right)^2}{\mathbf{E}[|J|^2]} \geqslant \frac{((c-c_2)|F|/\Delta)^2}{\mathbf{E}[|J|^2]}. \quad \text{(F.1)}$$

Note that

$$\begin{aligned}\mathbf{E}\left[|J|^2\right] &= \sum_{u\in U_t}\mathbf{E}[Z_u] + \sum_{\substack{u,u'\in U_t\\ u\neq u'}}\mathbf{E}[Z_uZ_{u'}]\\ &\leqslant \mathbf{E}[|J|] + \sum_{\substack{u,u'\in U_t\\ u\neq u'}}\mathbf{E}[X_uX_{u'}]\\ &= \mathbf{E}[|J|] + \sum_{\substack{u,u'\in U_t\\ u\neq u'}}\sum_{\substack{(v,u)\in F_u\\ (v',u')\in F_{u'}}}\left(\mathbf{E}[X_{v\to u}]\,\mathbf{E}[X_{v'\to u'}] + \mathbf{Cov}[X_{v\to u}, X_{v'\to u'}]\right)\\ &\leqslant \mathbf{E}[|J|] + \sum_{\substack{u,u'\in U_t\\ u\neq u'}}\left(\left(\sum_{(v,u)\in F_u}\mathbf{E}[X_{v\to u}]\right)\left(\sum_{(v',u')\in F_{u'}}\mathbf{E}[X_{v'\to u'}]\right) + |F_u||F_{u'}|\varepsilon\right)\\ &= \mathbf{E}[|J|] + O\left(\sum_{\substack{u,u'\in U_t\\ u\neq u'}}|F_u||F_{u'}|/\Delta^2\right)\\ &= \mathbf{E}[|J|] + O\left(\left(\sum_{u\in U_t}|F_u|\right)^2/\Delta^2\right)\\ &= \mathbf{E}[|J|] + O\left(|F|^2/\Delta^2\right).\end{aligned}$$

Here $\mathbf{E}[|J|] = \sum_{u\in U_t}\mathbf{E}[Z_u] \leqslant \sum_{u\in U_t}\mathbf{E}[X_u] = \sum_{u\in U_t}O(|F_u|/\Delta) = O(|F|/\Delta)$. Using the condition $|F| = \Omega(\Delta)$, we have $\mathbf{E}\left[|J|^2\right] = O\left(|F|^2/\Delta^2\right)$. Substitute it in (F.1), and then the second statement follows. ■

Now we prove Theorem 1.3. We first define a matrix $\mathcal{M} \in \mathbb{R}^{n\times n}$ that is associated with graph G. For any $u, v \in V[G]$, let $\mathcal{M}_{u,v} = 1/\Delta$ if $\{u,v\} \in E[G]$, $\mathcal{M}_{u,v} = 1 - \deg(u)/\Delta$ if $u = v$,

and $\mathcal{M}_{u,v} = 0$ otherwise. Notice that matrix $\mathcal{M}$ is doubly stochastic. We further define the conductance of matrix $\mathcal{M}$ by

$$\Phi(\mathcal{M}) \triangleq \min_{\substack{A \subset V \\ |A| \leqslant n/2}} \frac{e(A, \overline{A})}{\Delta \cdot |A|}.$$

Notice that $\Phi(\mathcal{M}) \leqslant \phi(G) \leqslant \Phi(\mathcal{M}) \cdot \beta$, where $\beta \triangleq \Delta/\delta$. Hence it suffices to work with $\Phi(\mathcal{M})$ in the following.

Proof of Theorem 1.3. The proof is divided into four phases, depending on the number of informed nodes $|I_t|$ after round t.

Phase 1: $1 \leqslant |I_t| \leqslant 1/\Phi$. This phase is divided into several subphases. For every $1 \leqslant i \leqslant \log(1/\phi)$, subphase i begins when the number of informed nodes is at least 2^{i-1} and ends when this number is at least 2^i. Assume that we are at the beginning of the ith subphase. Fix an arbitrary round t of the ith subphase and the set of informed nodes I_t; thus, $2^{i-1} \leqslant |I_t| < 2^i$. We consider the number of nodes that become informed in round $t+1$. Applying Lemma F.6(1) with $F = E(I_t, U_t)$ gives

$$\mathbf{Pr}\left[\, |I_{t+1} \setminus I_t| \geqslant 1 \,\right] \geqslant c_1 \min\{e(I_t, U_t)/\Delta, 1\} \geqslant c_1 \min\{\Phi \cdot |I_t|/\beta, 1\},$$

Let $p \triangleq c_1 \min\{\Phi \cdot |I_t|/\beta, 1\}$, and hence $p = O(\Phi \cdot |I_t|)$ since $|I_t| \leqslant 1/\Phi$ and $\beta \geqslant 1$. Therefore, the expected time to increase $|I_t|$ from 2^{i-1} to 2^i is at most $2^{i-1}/p = O(1/\Phi)$. By Markov's inequality,

$$\mathbf{Pr}\left[\, |I_{t+\tau}| \leqslant 2^i \;\middle|\; |I_t| \geqslant 2^{i-1} \,\right] \leqslant 1/2$$

for some $\tau = O(\Phi^{-1})$. Hence the time to complete Phase 1 can be upper bounded by $\tau = O((1/\Phi))$ multiplied with the sum of $\log(1/\Phi) = O(\log n)$ independent geometric random variables each with parameter $1/2$. Applying a Chernoff bound for the sum of independent geometric random variables yields that the number of rounds required for Phase 1 is at most $O((1/\Phi) \cdot \log n) = O((1/\phi) \cdot \beta \cdot \log n)$ with high probability.

Phase 2: $1/\Phi \leqslant |I_t| \leqslant n/2$**.** Fix a round t and the set of informed nodes I_t. We apply Lemma F.6(2), with $F = E(I_t, U_t)$. Note that the precondition $|F| = \Omega(\Delta)$ is satisfied, as

$$|F| = e(I_t, U_t) \geqslant \Phi \cdot \Delta \cdot |I_t| \geqslant \Phi \cdot \Delta \cdot (1/\Phi) = \Omega(\Delta).$$

Hence we conclude from Lemma F.6(2) that

$$\mathbf{Pr}\left[\, |I_{t+1} \setminus I_t| \geqslant c_2 \cdot \phi \cdot \delta \cdot |I_t|/\Delta \,\right] \geqslant c_3,$$

for some constant $c_2, c_3 > 0$. When this event occurs, we have $|I_{t+1}| \geqslant (1 + c_2 \cdot \phi/\beta)|I_t|$. So, the number of rounds until we have $|I_t| \leqslant n/2$ can be upper bounded by the sum of $\log_{1+c_2 \cdot \phi/\beta}(n/2) = O((1/\phi) \cdot \beta \cdot \log n)$ independent geometric random variables with parameters c_3. Using again the Chernoff bound we obtain that Phase 2 is completed within at most $O((1/\phi) \cdot \beta \cdot \log n)$ rounds with high probability.

Phase 3: $n/2 \leqslant |I_t| \leqslant n - 1/\Phi$**.** The analysis is the same as in Phase 2 with the roles of I_t and U_t switched.

Phase 4: $n - 1/\Phi \leqslant |I_t| \leqslant n$**.** Again, the analysis is the same as in Phase 1 with the roles of I_t and U_t switched.

Since each of the four phases requires only $O((1/\phi) \cdot \beta \cdot \log n)$ rounds with high probability, the result follows by applying the union bound. ∎

F.3 Analysis of Protocol 4

We first remark that the condition $\alpha = 1 - o(1)$ is equivalent to $\lambda \triangleq \lambda_2 = o(1)$, which will be used in the following.

To relate the spectral expansion of G with the expansion property, we use the following expander mixing lemma for general graphs.

Lemma F.7 (Expander Mixing Lemma for General Graphs [9])**.** *Let G be a general graph. Then for any subset X and Y it holds that*

$$\left| e(X,Y) - \frac{\mathrm{vol}(X)\cdot\mathrm{vol}(Y)}{\mathrm{vol}(G)} \right| \leqslant \lambda \cdot \frac{\sqrt{\mathrm{vol}(X)\cdot\mathrm{vol}(Y)\cdot\mathrm{vol}(\overline{X})\cdot\mathrm{vol}(\overline{Y})}}{\mathrm{vol}(G)}.$$

In order to prove Theorem 1.4, it suffices to show the following lemma:

Lemma F.8. *Let G be a graph that satisfies the preconditions of Theorem 1.4. Then with high probability all the following statements hold:*

- ***Phase I*** *Suppose $1 \leqslant |I_t| \leqslant n/\log n$. Then there is $\tau = \log n + o(\log n)$ such that $|I_{t+\tau}| > n/\log n$.*
- ***Phase II*** *Suppose $n/\log n \leqslant |I_t| \leqslant n - n/\log n$. Then there is $\tau = o(\log n)$ such that $|I_{t+\tau}| > n - n/\log n$.*
- ***Phase III*** *Suppose $|I_t| \geqslant n - n/\log n$. Then there is $\tau = \ln n + o(\log n)$ such that $|I_{t+\tau}| = n$.*

Proof. For any round t and $u \in U_t$, $v \in I_t$, let $X_{v\to u}$ be the boolean random variable whose value is 1 iff v informs u in round $t+1$. Note that Γ is a $(\leqslant K, (1-\varepsilon^2/4)D)$-expander and hence a $(2, (1-\varepsilon^2/4)D)$-expander. And $\mathcal{G}$ is a pairwise independent generator. Then we observe that the statements in Lemma F.5 hold here as well by the same proof. Notice that it holds by Lemma F.7 that

$$\begin{aligned} e(I_t, U_t) &\geqslant \frac{\mathrm{vol}(I_t)\cdot\mathrm{vol}(U_t)}{\mathrm{vol}(G)} - \lambda\cdot\frac{\mathrm{vol}(I_t)\cdot\mathrm{vol}(U_t)}{\mathrm{vol}(G)} \\ &\geqslant (1-\lambda)\cdot\frac{\mathrm{vol}(I_t)\cdot(\mathrm{vol}(G)-\mathrm{vol}(I_t))}{\mathrm{vol}(G)} \end{aligned} \tag{F.2}$$

Phase I. By (F.2) we have

$$e(I_t, U_t) \geqslant (1-\lambda)\cdot\delta\cdot|I_t|\left(1 - \frac{\Delta\cdot|I_t|}{nd}\right).$$

Since $\lambda = o(1)$ and $|I_t| \leqslant n/\log n$, we have

$$e(I_t, U_t) \geqslant (1-o(1))\cdot\Delta\cdot|I_t|\left(\frac{\delta}{\Delta} - \frac{\delta}{d\cdot\log n}\right) \geqslant \left(1 - \frac{1}{\log n} - o(1)\right)\cdot\Delta\cdot|I_t|. \tag{F.3}$$

Hence

$$|N(I_t)\setminus I_t| \geqslant \frac{e(I_t,U_t)}{\Delta} \geqslant \left(1 - \frac{1}{\log n} - o(1)\right)\cdot|I_t|.$$

Define $\gamma \triangleq \lambda + \frac{1}{\log n}$, and $A \triangleq \{u \in N(I_t)\setminus I_t : |N(u)\cap I_t| \geqslant 2d\sqrt{\gamma}\}$. Then $e(A, I_t) \geqslant |A|\cdot 2d\cdot\sqrt{\gamma}$. On the other hand by Lemma F.7 it holds that

$$\begin{aligned} e(A, I_t) &\leqslant \frac{\mathrm{vol}(A)\cdot\mathrm{vol}(I_t)}{\mathrm{vol}(G)} + \lambda\sqrt{\mathrm{vol}(A)\cdot\mathrm{vol}(I_t)} \\ &\leqslant \frac{\Delta^2\cdot|A|\cdot|I_t|}{nd} + \gamma\Delta\cdot\sqrt{|A|\cdot|I_t|}. \end{aligned}$$

By the definition of set A we have $e(A, I_t) \geqslant 2d\sqrt{\gamma} \cdot |A|$, and hence

$$\begin{aligned} |A| \cdot 2d \cdot \sqrt{\gamma} &\leqslant \frac{\Delta^2 \cdot |A| \cdot |I_t|}{nd} + \gamma\Delta \cdot \sqrt{|A| \cdot |I_t|} \\ &\leqslant (1 + o(1)) \cdot \frac{\Delta \cdot |A|}{\log n} + \gamma\Delta \cdot \sqrt{|A| \cdot |I_t|}, \end{aligned}$$

which implies $|A| \leqslant \gamma \cdot |I_t|$.

Now define $B \triangleq N(I_t) \setminus I_t \setminus A$. We have

$$e(B, I_t) = e(N(I_t), I_t) - e(A, I_t) \geqslant \left(1 - \frac{1}{\log n} - o(1) - \gamma\right) \Delta \cdot |I_t|.$$

With the above estimate at hand, we compute the expected value of $|I_t \cap B|$. Note that for any $u \in B$, the chance that it gets informed in round $t + 1$ is

$$p_{t+1}(u) \triangleq \mathbf{Pr}\left[\bigvee_{v \in N(u) \cap I_t} (X_{v \to u} = 1)\right],$$

which is lower bounded by

$$\sum_{v \in N(u) \cap I_t} \mathbf{Pr}\left[\, X_{v \to u} = 1 \,\right] - \sum_{\substack{v_1, v_2 \in N(u) \cap I_t \\ v_1 < v_2}} \mathbf{Pr}\left[\bigwedge_{i=1,2} (X_{v_i \to u} = 1)\right]$$

by Bonferroni inequalities. Hence

$$\begin{aligned} p_{t+1}(u) &\geqslant |N(u) \cap I_t| \left(\frac{1}{\Delta} - \varepsilon\right) - \binom{|N(u) \cap I_t|}{2} \left(\frac{1}{\delta^2} + \varepsilon\right) \\ &\geqslant (1 - o(1)) \cdot \frac{|N(u) \cap I_t|}{\Delta} - (1 + o(1)) \cdot \binom{|N(u) \cap I_t|}{2} \cdot \frac{1}{\Delta^2} \\ &\geqslant (1 - o(1)) \cdot \frac{|N(u) \cap I_t|}{\Delta} \left(1 - \frac{(1 + o(1)) \cdot |N(u) \cap I_t|}{2\Delta}\right) \\ &\geqslant (1 - o(1)) \cdot \frac{|N(u) \cap I_t|}{\Delta}, \end{aligned} \tag{F.4}$$

where the first inequality follows from Lemma F.5 and the fact that $\varepsilon = (1/\Delta)^{\Theta(1)}$ is sufficiently small, and the last step uses the condition that $|N(u) \cap I_t| \leqslant 2d\sqrt{\gamma} = o(\Delta)$. Hence we have

$$\begin{aligned} \mathbf{E}\left[\, |I_{t+1} \setminus I_t| \,\right] \geqslant \mathbf{E}\left[\, |I_{t+1} \cap B| \,\right] &= \sum_{u \in B} p_{t+1}(u) \geqslant \sum_{u \in B} (1 - o(1)) \cdot \frac{|N(u) \cap I_t|}{\Delta} \\ &= (1 - o(1)) \cdot \frac{e(B, I_t)}{\Delta} \geqslant (1 - o(1)) \cdot |I_t|. \end{aligned}$$

Since $|I_{t+1} \setminus I_t| \leqslant |I_t|$, it follows by using Markov's inequality (applied to $|I_t| - |I_{t+1} \setminus I_t|$) that $\mathbf{Pr}\left[\, |I_{t+1}| \geqslant (2 - f(n))|I_t| \,\right] \geqslant 1 - g(n)$, where $f(n)$ and $g(n)$ are both functions that tend to zero. Hence the time to reach $|I_t| \geqslant n/\log n$ can be upper bounded by the sum of $\log_{2-f(n)} n$ independent, identically distributed geometric random variables with expectation at most $1 - o(1)$ each. Using the Chernoff bound from Lemma A.1 yields for $\tau \triangleq \log_2 n + o(\log n)$ that $\mathbf{Pr}\left[\, |I_{t+\tau}| > n/\log n \,\right] = 1 - o(1)$.

Phase II $|I_t| \in [n/\log n, n - n/\log n]$. We further divide this phase into the two cases $|I_t| \in [n/\log n, n/2]$ and $|I_t| \in [n/2, n - n/\log n]$. We start with the first case $|I_t| \in [n/\log n, n/2]$.

For any $u \in N(I_t) \setminus I_t$, the probability $p_{t+1}(u)$ that u gets informed in round $t+1$ is lowered bounded by

$$(1 - o(1)) \cdot \frac{|N(u) \cap I_t|}{\Delta} \left(1 - \frac{(1 + o(1)) \cdot |N(u) \cap I_t|}{2\Delta}\right)$$

by the same argument as in (F.4). This is then lower bounded by

$$(1 - o(1)) \cdot \frac{|N(u) \cap I_t|}{2\Delta},$$

since we have $|N(u) \cap I_t| \leqslant \Delta$.

By (F.2), we have

$$e(I_t, U_t) = (1 - o(1)) \cdot \frac{\delta}{2}|I_t|.$$

Similar to the analysis of Phase I, we can lower bound the expected number of nodes that become informed in round $t + 1$:

$$\begin{aligned}\mathbf{E}\left[\,|I_{t+1} \setminus I_t|\,\right] &\geqslant \sum_{u \in N(I_t) \setminus I_t} p_{t+1}(u) \geqslant (1 - o(1)) \sum_{u \in N(I_t) \setminus I_t} \frac{|N(u) \cap I_t|}{2\Delta} \\ &= (1 - o(1)) \frac{e(I_t, U_t)}{2\Delta} \geqslant \frac{\delta}{8\Delta}|I_t|.\end{aligned}$$

Since $|I_{t+1}| \leqslant 2|I_t|$, we obtain that as long as $|I_t| \leqslant n/2$ there are constants $\alpha, \beta > 0$ so that $\mathbf{Pr}\left[\,|I_{t+1}| \geqslant (1 + \alpha)|I_t|\,\right] \geqslant \beta$. Hence the time to reach $|I_t| \geqslant n/2$ can be upper bounded by the sum of $\log_{1+\alpha}(\log n)$ independent, identically distributed geometric random variables with expectation at most $1/\beta$ each. Using the Chernoff bound for the sum of geometric random variables (see Lemma A.1) yields that with probability $1 - o(1)$, we reach $|I_t| \geqslant n/2$ within at most $o(\log n)$ additional rounds.

Consider now the case $|I_t| \in [n/2, n - n/\log n]$. To analyze this case, we examine the shrinking of $U_t = V \setminus I_t$. Note that for any $u \in U_t$, the probability $p_{t+1}(u)$ that u gets informed in round $t + 1$ is lowered bounded by

$$(1 - o(1)) \cdot \frac{|N(u) \cap I_t|}{\Delta} \left(1 - \frac{(1 + o(1)) \cdot |N(u) \cap I_t|}{2\Delta}\right)$$

by the same argument as in (F.4). This is then lower bounded by

$$(1 - o(1)) \cdot \frac{|N(u) \cap I_t|}{2\Delta}$$

since we have $|N(u) \cap I_t| \leqslant \Delta$.

Again, as $|U_t| \leqslant n/2$, by (F.2) we have

$$e(I_t, U_t) \geqslant (1 - o(1)) \cdot \frac{\delta}{2}|U_t|.$$

Let us now compute the expected number of uninformed nodes after one additional round:

$$\begin{aligned}\mathbf{E}\left[\,|U_{t+1}|\,\right] &= \sum_{u \in U_t} (1 - p_{t+1}(u)) \leqslant |U_t| - (1 - o(1)) \sum_{u \in U_t} \left(\frac{|N(u) \cap I_t|}{2\Delta}\right) \\ &= |U_t| - (1 - o(1)) \frac{e(I_t, U_t)}{2\Delta} \leqslant \left(1 - \frac{\delta}{8\Delta}\right)|U_t|.\end{aligned}$$

A simple inductive argument yields for any integer τ that,

$$\mathbf{E}\left[\,|U_{t+\tau}|\,\right] \leqslant \left(1 - \frac{\delta}{8\Delta}\right)^{\tau} |U_t|,$$

so for $\tau \triangleq \log\log n/\log(1/(1-\frac{\delta}{8\Delta}))+\omega(1)$, where $\omega(1)$ is an arbitrarily slow growing function, we have $\mathbf{E}\left[\,|U_{t+\tau}|\,\right] = o(n/\log n)$. Hence by Markov's inequality, $\mathbf{Pr}\left[\,|U_{t+\tau}| \geqslant n/\log n\,\right] = o(1)$.

Phase III $|I_t| \in [n - n/\log n, n]$. Again, we analyze the shrinking of the set U_t. By Lemma F.4, for at least $(1-\varepsilon/2)$-fraction of the choices of x_t, it holds that the size of $\{\Gamma(v,x_t) : v \in N(u)\cap I_t\}$ is at least $(1-\varepsilon/2)|N(u)\cap I_t|$. From now on fix x_t such that this event occurs.

For any $u \in U_t$, we have

$$\mathbf{Pr}\left[\,u \notin I_{t+1}\,\right] = \mathbf{Pr}\left[\bigwedge_{v\in N(u)\cap I_t}(X_{v\to u}=0)\right].$$

Let F be a subset of $N(u)\cap I_t$ of size $(1-\varepsilon/2)|N(u)\cap I_t|$ such that the map $\Gamma(\cdot,x_t)$ is injective when restricted to F. By Lemma D.4, the function $y \mapsto \left(\mathcal{G}_{\Gamma(v,x_t)}(y) \bmod \deg(v)\right)_{v\in F}$ is an $(\varepsilon' + |F|\Delta/m)$-$\mathsf{PRG}$ for CR_S where $S = \prod_{v\in F}[\deg(v)]$.

Then we have

$$\begin{aligned}
\mathbf{Pr}\left[\,u \notin I_{t+1}\,\right] &\leqslant \mathbf{Pr}\left[\bigwedge_{v\in F}(X_{v\to u}=0)\right] \leqslant \prod_{v\in F}\mathbf{Pr}\left[\,X_{v\to u}=0\,\right] + \varepsilon' + |F|\Delta/m\\
&\leqslant \prod_{v\in F}\left(1-\frac{1}{\deg(v)}+\varepsilon\right)+\varepsilon'+\Delta^2/m\\
&\leqslant \left(1-\frac{1}{\Delta}+\varepsilon\right)^{(1-\varepsilon/2)|N(u)\cap I_t|}+\varepsilon'+\Delta^2/m,
\end{aligned}$$

where the second inequality follows from the properties of PRGs for combinatorial rectangles, and the third inequality follows from using pairwise independent generators. Since $\varepsilon \leqslant \frac{1}{\Delta}$, a simple induction shows that

$$\left(1-\frac{1}{\Delta}+\varepsilon\right)^k \leqslant \left(1-\frac{1}{\Delta}\right)^k + k\varepsilon$$

for any $k \geqslant 0$. So we have

$$\begin{aligned}
\mathbf{Pr}\left[\,u \notin I_{t+1}\,\right] &\leqslant \left(1-\frac{1}{\Delta}\right)^{(1-\varepsilon/2)|N(u)\cap I_t|} + (1-\varepsilon/2)\cdot|N(u)\cap I_t|\cdot\varepsilon+\varepsilon'+\Delta^2/m\\
&\leqslant \left(1-\frac{1}{\Delta}\right)^{(1-\varepsilon/2)|N(u)\cap I_t|} + (1-\varepsilon/2)\cdot\Delta\cdot\varepsilon+\varepsilon'+\Delta^2/m.
\end{aligned}$$

The bound above applies for any choice of x_t such that the size of $\{\Gamma(v,x_t) : v \in N(u)\cap I_t\}$ is at least $(1-\varepsilon/2)|N(u)\cap I_t|$. And the probability of choosing such x_t is at least $1-\varepsilon/2$. So for random x_t, we have

$$\begin{aligned}
\mathbf{Pr}\left[\,u \notin I_{t+1}\,\right] &\leqslant \left(1-\frac{1}{\Delta}\right)^{(1-\varepsilon/2)\cdot|N(u)\cap I_t|} + (1-\varepsilon/2)\cdot\Delta\cdot\varepsilon+\varepsilon'+\Delta^2/m+\varepsilon/2\\
&\leqslant \left(1-\frac{1}{\Delta}\right)^{(1-\varepsilon/2)\cdot|N(u)\cap I_t|} + o(1),
\end{aligned}$$

where we use the fact that $\varepsilon = (1/\Delta)^{\Theta}$ is sufficiently small, and $m = \Theta((\log n)/\varepsilon)$.

By (F.3) it holds that $e(I_t,U_t) \geqslant (1-\frac{1}{\log n}-o(1))\cdot\Delta|U_t|$. Let $A \subseteq U_t$ be the set of nodes v for which $|N(v)\cap I_t| \leqslant (1-\sqrt{\gamma}/2)\cdot\Delta$, where $\gamma \triangleq \frac{1}{\log n}+o(1)$. We assume for a contradiction that $|A| > 2\sqrt{\gamma}\cdot|U_t|$. Hence,

$$\begin{aligned}
e(I_t,U_t) &= \sum_{v\in A}|N(v)\cap I_t| + \sum_{v\in U_t\setminus A}|N(v)\cap I_t| \leqslant |A|\cdot(1-\sqrt{\gamma}/2)\Delta + |U_t\setminus A|\Delta\\
&= |U_t|\Delta - |A|\sqrt{\gamma}\Delta/2 < \left(1-\frac{1}{\log n}-o(1)\right)\cdot\Delta|U_t|,
\end{aligned}$$

which yields the desired contradiction. Hence $|A| \leqslant 2\sqrt{\gamma}|U_t|$. Now define $B \triangleq U_t \setminus A$ so that for each $u \in B$, $|N(v) \cap I_t| > (1-\sqrt{\gamma}/2)\Delta$ and $|B| \geqslant (1-2\sqrt{\gamma})|U_t|$. Using linearity of expectation,

$$\begin{aligned}
\mathbf{E}\left[\,|U_{t+1}|\,\right] &\leqslant \sum_{u\in B} \mathbf{Pr}\left[\,u \notin I_{t+1}\,\right] + \sum_{u\in A} \mathbf{Pr}\left[\,u \notin I_{t+1}\,\right] \\
&\leqslant \sum_{u\in B} \left(\left(1-\frac{1}{\Delta}\right)^{(1-\varepsilon/2)|N(u)\cap I_t|} + o(1)\right) + \sum_{u\in A} 1 \\
&\leqslant \sum_{u\in B} \left(1-\frac{1}{\Delta}\right)^{(1-\varepsilon/2)|N(u)\cap I_t|} + o(|U_t|) + |A| \\
&= \sum_{u\in B} \left(1-\frac{1}{\Delta}\right)^{(1-\varepsilon/2)|N(u)\cap I_t|} + o(|U_t|).
\end{aligned}$$

Using the inequalities that $(1-1/k) \leqslant \mathrm{e}^{-1/k}$ for $k \geqslant 1$, $\mathrm{e}^x \leqslant 1+2x$ for sufficiently small constant $x > 0$, and the condition that $|N(u) \cap I_t| \geqslant (1-\sqrt{\gamma}/2)\cdot\Delta$ for $u \in B$, we get

$$\begin{aligned}
\mathbf{E}\left[\,|U_{t+1}|\,\right] &\leqslant \sum_{u\in B} \mathrm{e}^{-(1-\varepsilon/2)|N(u)\cap I_t|/\Delta} + o(|U_t|) \leqslant \sum_{u\in B} \mathrm{e}^{-(1-\sqrt{\gamma}/2-o(1))} + o(|U_t|) \\
&= \sum_{u\in B} \mathrm{e}^{-1}\cdot \mathrm{e}^{\sqrt{\gamma}/2+o(1)} + o(|U_t|) \leqslant \sum_{u\in B} \mathrm{e}^{-1}\cdot(1+\sqrt{\gamma}+o(1)) + o(|U_t|) \\
&= (1+o(1))\cdot \mathrm{e}^{-1}\cdot|U_t|.
\end{aligned}$$

By induction, it follows that for any step $\tau > 0$, $\mathbf{E}\left[\,|U_{t+\tau}|\,\right] \leqslant ((1+o(1))\cdot\mathrm{e}^{-1})^\tau \cdot |U_t|$. We choose $\tau \triangleq -\log_{(1+o(1))\cdot\mathrm{e}^{-1}}(n) = \ln n + o(\log n)$ and obtain that $\mathbf{E}\left[\,|U_{t+\tau}|\,\right] \leqslant (1/\log n)$. So $\mathbf{Pr}\left[\,|U_{t+\tau}| \geqslant 1\,\right] \leqslant \mathbf{E}\left[\,|U_{t+\tau}|\,\right] \leqslant 1/\log n$. ■