

Rumor Spreading on the Complete Graph

Aakash Kumar, Maria Sofia Bucarelli, Emanuele Natale

October 7, 2026

Abstract

We give a complete and elementary proof that the uniform *push* protocol on the complete graph K_n , started from a single informed node, informs all n nodes within $O(\log n)$ rounds with high probability. The probability space is finite and uniform throughout; the only analytic ingredients are Bernoulli's inequality and Markov's inequality, and the only concentration tool is a one-line exponential-moment induction for adaptive Bernoulli trials. No measure theory, no PMF/**ENNReal**, no Chernoff bounds, and no martingales are used. This argument has been *fully formalized* (no **sorry**) in Lean 4 + Mathlib; the main theorem is `RumorPush.push_informs_all_whp`. With the same random calls, the *pull* and *push-pull* protocols of Karp, Schindelhauer, Shenker and Vöcking also inform all nodes within $O(\log n)$ rounds with high probability (Section 10); for pull, a second-moment bound replaces the reverse Markov inequality.

1 The model

Fix $n \geq 2$ and let $V = \{0, 1, \dots, n-1\}$ be the vertex set of the complete graph K_n .

Definition 1 (Round randomness). A *round configuration* is a function $\omega: V \rightarrow V$ with $\omega(v) \neq v$ for all v . Let R_n denote the (finite) set of round configurations; it carries the uniform probability measure, i.e. every node picks a uniformly random *target* among the other $n-1$ nodes, independently of all other nodes.

Definition 2 (Push step). For an *informed set* $I \subseteq V$ and a round configuration $\omega \in R_n$, the next informed set is

$$\text{step}(I, \omega) = I \cup \omega(I) = I \cup \{\omega(v) : v \in I\}.$$

(Only informed nodes actually transmit; letting *all* nodes draw a target and ignoring the choices of uninformed nodes is an equivalent and formalization-friendlier convention.)

Definition 3 (Process). Fix a horizon $T \in \mathbb{N}$ and a start vertex $v_0 \in V$. The sample space is $\Omega = R_n^T$ (uniform measure, i.e. i.i.d. uniform rounds). For $\omega = (\omega_0, \dots, \omega_{T-1}) \in \Omega$ define $I_0(\omega) = \{v_0\}$ and $I_{t+1}(\omega) = \text{step}(I_t(\omega), \omega_t)$ for $0 \leq t < T$. We write $m_t = |I_t|$ and $U_t = n - m_t$ for the numbers of informed and uninformed nodes. Since Ω is finite, all probabilities are finite ratios and all expectations are finite sums; “conditioning on I_t ” means partitioning Ω according to the first t coordinates. In the formalization, a trajectory is simply a `List` of round configurations and run `I l` folds `step` over the list `l`.

Two trivial but load-bearing facts, both immediate from Definition 2:

Lemma 4 (Monotonicity). $I_t(\omega) \subseteq I_{t+1}(\omega)$ for all t and all ω ; hence m_t is non-decreasing and U_t is non-increasing in t , pointwise on Ω .

Lemma 5 (At most doubling). $|I_{t+1}(\omega) \setminus I_t(\omega)| \leq |I_t(\omega)|$; equivalently $(\text{step}(I, \omega)).\text{card} \leq 2I.\text{card}$.

2 Main theorem

Throughout, $\log$ denotes the natural logarithm.

Theorem 6 (Push protocol informs K_n in $O(\log n)$ rounds w.h.p.). *Let $n \geq 2$, let $v_0 \in V$, and set*

$$T_1 = \lceil 117 \log n \rceil + 23, \quad T_2 = \lceil 6 \log n \rceil, \quad T = T_1 + T_2.$$

Then $\mathbb{P}[I_T = V] \geq 1 - \frac{2}{n}$. In particular all nodes are informed after $O(\log n)$ rounds with high probability.

Remark 7. The constants are deliberately crude: the proof is organized to minimize the analytic toolkit, not the round count. The sharp answer is $\log_2 n + \log n + o(\log n)$ (Frieze–Grimmett [1], Pittel [2]). Replacing $2/n$ by $n^{-\alpha}$ for any fixed $\alpha \geq 1$ only changes the constants in T_1, T_2 .

The proof has two phases, glued by Lemma 4:

- **Growth phase** (Section 7): while $m_t \leq n/2$, each round multiplies m_t by $\geq 9/8$ with probability $\geq 1/8$; an exponential-moment bound for adaptive trials shows that after T_1 rounds, $m_{T_1} > n/2$ except with probability $\leq 1/n$.
- **Saturation phase** (Section 8): once $m_t \geq n/2$, the expected number of uninformed nodes contracts by a factor $2/3$ per round; after T_2 further rounds Markov’s inequality gives $U_T = 0$ except with probability $\leq 1/n$.

3 Finite uniform probability

All randomness in this development is uniform over finite types, so instead of measure theory or PMF we work with plain finite sums. The definitions and their basic laws are those of the shared `dynamics` library (`Dynamics.Uniform`), which also identifies them with Mathlib’s finite expectation `Finset.expect` and with the independent product of uniform distributions (`Dynamics.Bridge`).

Definition 8 (Average). For a finite type α and $f: \alpha \rightarrow \mathbb{R}$, $\text{avg}(f) := (\sum_a f(a))/|\alpha|$ is the expectation of f under the uniform distribution on α .

Definition 9 (Trajectory expectation). For a finite type α , $\text{expList } \alpha \ T \ F$ is the expectation of a functional F of a trajectory of T i.i.d. uniform draws from α , defined *by recursion on T* : the head draw is averaged out first, i.e. $\text{expList } \alpha \ 0 \ F = F(\text{[]})$ and $\text{expList } \alpha \ (T+1) \ F = \text{avg}(a \mapsto \text{expList } \alpha \ T \ (l \mapsto F(a :: l)))$. This makes “conditioning on the first round” a definitional unfolding rather than a measure-theoretic construction.

Lemma 10 (Basic properties). *avg and expList are monotone, additive, and homogeneous, and constants pass through unchanged. These pointwise inequalities pushed through `avg_le_avg` / `expList_le_expList` are the only form in which Markov’s inequality and union bounds are used below (Lemmas 16 and 17); they are not separately formalized lemmas.*

Theorem 11 (Faithfulness of the model). *expList $\alpha \ T \ F$ equals the uniform average of F over all length- T sequences of draws, i.e. over the product probability space $\text{Fin } T \rightarrow \alpha$ of T i.i.d. uniform rounds. This confirms that the recursive definition of Definition 9 is the textbook object.*

4 Elementary inequalities

We write $x := \frac{1}{n-1}$, so $0 < x \leq 1$ for $n \geq 2$.

Lemma 12 (Bernoulli). *For all real $y \geq -1$ (we only use $y \in [-1, 1]$) and $m \in \mathbb{N}$: $(1 + y)^m \geq 1 + my$. Equivalently, with $y = -x$: $1 - mx \leq (1 - x)^m$.*

Lemma 13 (Second-order lower bound, “Bonferroni”). *For $0 \leq x \leq 1$ and $m \in \mathbb{N}$:*

$$1 - (1 - x)^m \geq mx - \binom{m}{2} x^2 \geq mx \left(1 - \frac{mx}{2}\right),$$

and consequently $mx/2 \leq 1 - (1 - x)^m$ whenever $mx \leq 1$.

Lemma 14 (Upper bound without e). *For $0 \leq x \leq 1$ and $m \in \mathbb{N}$: $(1 - x)^m \leq \frac{1}{1 + mx}$.*

Lemma 15 (Logarithm bounds). *For $x > 0$: $1 - \frac{1}{x} \leq \log x$. (The formalization only needs this left inequality, applied at $x = 9/8, 16/15, 3/2$, together with Mathlib’s numeric bound $\log 2 < 0.6931471808$.)*

Lemma 16 (Markov’s inequality, finite uniform version). *Let Ω be a finite nonempty set with the uniform measure, $f: \Omega \rightarrow \mathbb{R}$ with $f \geq 0$, and $a > 0$. Then $\mathbb{P}[f \geq a] \leq \mathbb{E}[f]/a$. In particular, if f takes values in $\mathbb{N}$, then $\mathbb{P}[f \neq 0] \leq \mathbb{E}[f]$.*

Lemma 17 (Reverse Markov for bounded variables). *Let $f: \Omega \rightarrow \mathbb{R}$ with $0 \leq f \leq b$ pointwise, $b > 0$, and let $0 < a < b$. Then $\mathbb{P}[f \geq a] \geq \frac{\mathbb{E}[f] - a}{b}$. (As with Lemma 16, the formalization does not isolate this as a named lemma: each use site supplies its own pointwise bound and pushes it through `avg_le_avg`, e.g. in the proof of Lemma 20.)*

5 One-round estimates

Fix an informed set I with $|I| = m \geq 1$ and let $\omega \in R_n$ be a uniform round configuration. All statements in this section concern this single round; by independence of the rounds they apply verbatim to the conditional law of round t given $I_t = I$.

Lemma 18 (Contact probability). *For any fixed $u \notin I$,*

$$\mathbb{P}_\omega[u \in \text{step}(I, \omega)] = 1 - \left(1 - \frac{1}{n-1}\right)^m.$$

The Lean proof is the only place where a probability is computed rather than estimated: it counts round configurations avoiding u on I via an explicit equivalence to a product of subtypes and `Fintype.card_pi`.

Lemma 19 (Expected growth). *Let $N(\omega) = |\text{step}(I, \omega)| - m$ be the number of newly informed nodes. The formalization gives the exact formula*

$$\mathbb{E}_\omega[|\text{step}(I, \omega)|] = m + (n - m)(1 - (1 - x)^m),$$

from which, if $1 \leq m \leq n/2$: $\mathbb{E}_\omega[N] \geq m/4$ (using Lemma 13 and $n - m \geq n/2, mx \leq 1$).

Lemma 20 (A good round has constant probability). *If $1 \leq m \leq n/2$, then $\mathbb{P}_\omega[|\text{step}(I, \omega)| \geq \frac{9}{8}m] \geq \frac{1}{8}$.*

Lemma 21 (Contraction above half). *If $m \geq n/2$, then for every fixed $u \notin I$, $\mathbb{P}_\omega[u \notin \text{step}(I, \omega)] = (1 - x)^m \leq \frac{2}{3}$, and consequently, with $U' := n - |\text{step}(I, \omega)|$, $\mathbb{E}_\omega[U'] \leq \frac{2}{3}(n - m)$.*

6 Adaptive Bernoulli trials

Definition 22 (Good rounds). A round is *good* if it multiplies the informed set by $\geq 9/8$, or if the informed set already exceeds $n/2$:

$$X_t = \mathbf{1}\left[m_{t+1} \geq \frac{9}{8} m_t \text{ or } m_t > \frac{n}{2}\right] \in \{0, 1\}.$$

goodCount counts the good rounds along a trajectory. By Lemma 20 (applied when $m_t \leq n/2$; trivial otherwise), $\mathbb{P}[X_t = 1 \mid \omega_0, \dots, \omega_{t-1}] \geq \frac{1}{8}$ for every prefix, since ω_t is independent of the prefix and expList conditions on it definitionally.

Lemma 23 (Exponential moment for adaptive trials). *For every T and every nonempty I ,*

$$\mathbb{E}\left[\left(\frac{1}{2}\right)^{\text{goodCount}(I, \cdot)}\right] \leq \left(\frac{15}{16}\right)^T$$

over T rounds started at I . (This specializes the textbook “ $\mathbb{E}[s^G] \leq (1 - p(1 - s))^T$ for adaptive $\{0, 1\}$ -trials with per-step success probability $\geq p$ ” at $p = 1/8$, $s = 1/2$; the formalization proves the specialized statement directly by induction on T rather than the general one, since only this instance is needed. The ensuing tail bound $\mathbb{P}[G < L] \leq 2^L(15/16)^T$ — via $s^L \mathbb{P}[G < L] \leq \mathbb{E}[s^G]$ — is likewise not isolated as a separate lemma; it is proved inline inside Lemma 25 (phase1).

7 Growth phase

Lemma 24 (Enough good rounds force half-saturation). *Let $L \in \mathbb{N}$ satisfy $(9/8)^L > n/2$. If the informed set never exceeds $n/2$ up to time T_1 , then along that trajectory $m_{T_1} \geq (9/8)^{\text{goodCount} \cdot m_0}$: every good round multiplies m_t by $\geq 9/8$ and every round is non-decreasing (Lemma 4), so $\text{goodCount} \geq L$ would force $m_{T_1} > n/2$, a contradiction — hence $\text{goodCount} < L \Rightarrow m_{T_1} \leq n/2$ is impossible to rule out only via the deterministic growth bound stated here.*

Lemma 25 (Phase 1). *If $(9/8)^L \geq n$, then after T_1 rounds the informed set is still $\leq n/2$ with probability at most $2^L(15/16)^{T_1}$.*

Lemma 26 (Admissible constants). *With $L = \lceil 9 \log n \rceil + 1$ we have $(9/8)^L \geq n$ (**numeric_A**), and with $T_1 = \lceil 117 \log n \rceil + 23$ we have $2^L(15/16)^{T_1} \leq 1/n$ (**numeric_B**), using $\log \frac{9}{8} \geq \frac{1}{9}$, $\log \frac{16}{15} \geq \frac{1}{16}$ and $\log 2 < 0.6932$. Combined with Lemma 25: $\mathbb{P}[m_{T_1} \leq n/2] \leq 1/n$. The clean statement to formalize is that any L, T_1 satisfying these two numeric inequalities work; the values above merely exhibit admissible ones.*

8 Saturation phase

Lemma 27 (Phase 2). *Let $A = \{m_{T_1} > n/2\}$ (an event determined by the first T_1 coordinates). Then for every $s \geq 0$, $\mathbb{E}[U_{T_1+s} \mathbf{1}_A] \leq (2/3)^s n$, by induction on s using Lemma 21 at each step (monotonicity keeps $m_{T_1+s} > n/2$ on A).*

Corollary 28. *With $T_2 = \lceil 6 \log n \rceil$: $\mathbb{P}[U_T \geq 1 \text{ and } A] \leq n(2/3)^{T_2} \leq 1/n$. (U_T is $\mathbb{N}$ -valued, so Markov applied to $U_T \mathbf{1}_A$ and Lemma 27 give the first inequality; $(2/3)^{T_2} n \leq 1/n$ — **numeric_C** — follows from $\log \frac{3}{2} \geq \frac{1}{3}$ and $T_2 \geq 6 \log n$.)*

9 Proof of the main theorem

Lemma 29 (Gluing the two phases). *Let $\text{prNotAllInformed}(n, v_0, T)$ be the probability that some node is still uninformed after T rounds started at v_0 . For all L, T_1, T_2 with $(9/8)^L \leq n$:*

$$\text{prNotAllInformed}(n, v_0, T_1 + T_2) \leq 2^L \left(\frac{15}{16}\right)^{T_1} + \left(\frac{2}{3}\right)^{T_2} n.$$

Failure requires either phase 1 to fail (informed set still $\leq n/2$ after T_1 rounds, Lemma 25) or phase 2 to leave an uninformed node on the complementary event, controlled by Lemma 27 via Markov.

With $T_1 = \lceil 117 \log n \rceil + 23$, $T_2 = \lceil 6 \log n \rceil$, $T = T_1 + T_2$, Theorem 6 follows from Lemma 29, Lemma 26 and Corollary 28:

$$\mathbb{P}[I_T \neq V] = \mathbb{P}[U_T \geq 1] \leq \mathbb{P}[\bar{A}] + \mathbb{P}[U_T \geq 1 \text{ and } A] \leq \frac{1}{n} + \frac{1}{n} = \frac{2}{n}.$$

10 PULL and PUSH-PULL

We now treat the two other protocols of Karp, Schindelhauer, Shenker and Vöcking [3], driven by the *same* random calls: in a round configuration $\omega \in R_n$ every node v calls $\omega(v)$. In PULL the caller learns the rumor from its callee; in PUSH-PULL the rumor travels in both directions along every call. As for PUSH, partners are chosen among the other $n - 1$ nodes (KSSV allow self-calls; only constants change), and the round count is a crude $O(\log n)$ instead of the sharp $\log_3 n + O(\log \log n)$ of [3, Thm. 2.1].

Definition 30 (PULL and PUSH-PULL rounds). $\text{pullStep}(I, \omega) = I \cup \{v : \omega(v) \in I\}$ and $\text{pushPullStep}(I, \omega) = \text{step}(I, \omega) \cup \text{pullStep}(I, \omega)$; trajectories fold these steps over a list of i.i.d. uniform rounds, and the failure probabilities are the probabilities that some node is still uninformed after T rounds started from $\{v_0\}$.

Lemma 31 (Pathwise domination). *With the same calls, the PUSH-PULL informed set contains the PUSH and the PULL informed sets after every number of rounds.*

Lemma 32 (Independent calls). *The calls $\omega(v)$, $v \in V$, of a uniform round are independent, each uniform among the other $n - 1$ nodes; in particular a caller $v \notin I$ calls into I with probability $|I|/(n - 1)$.*

Lemma 33 (One PULL or PUSH-PULL round, exactly). *Let $m = |I|$ and $u = n - m$. From a single informed node a PULL round informs nobody with probability $(1 - \frac{1}{n-1})^{n-1}$. After one PULL round, $\mathbb{E}|I'| = m + u \frac{m}{n-1}$ and $\mathbb{E}[n - |I'|] = \frac{u(u-1)}{n-1} \leq \frac{u^2}{n}$ (quadratic shrinking). After one PUSH-PULL round, $\mathbb{E}[n - |I'|] = \frac{u(u-1)}{n-1}(1 - \frac{1}{n-1})^m$.*

Lemma 34 (The two-phase argument for any inflationary step). *Let $f(I, a) \supseteq I$ be a spreading step driven by i.i.d. uniform rounds a . Suppose that from every nonempty I a round is good ($|f(I, a)| \geq \frac{9}{8}|I|$, or $|I| > n/2$) with probability at least $1/8$, and that for $|I| \geq n/2$ the expected number of uninformed nodes contracts by $2/3$. Then after $\lceil 160 \log n \rceil$ rounds from $\{v_0\}$, some node is uninformed with probability at most $2/n$. The proof is the PUSH proof of Sections 6–9, which uses no other property of step.*

Lemma 35 (The two inputs for PULL). *A PULL round is good with probability at least $1/8$, and for $|I| \geq n/2$ it contracts the expected uninformed count by $2/3$.*

Proof. The contraction follows from $\frac{u(u-1)}{n-1} \leq \frac{2}{3}u$ for $u \leq n/2$. For good rounds, let X be the number of uninformed callers that call into I , so $|I'| = |I| + X$. By Lemma 32, $\mu = \mathbb{E}X = \frac{(n-m)m}{n-1} \geq \max(m/2, 1/2)$ for $m \leq n/2$ and, the calls being pairwise independent, $\mathbb{E}X^2 \leq \mu + \mu^2$. A round that is not good has $X < m/8 \leq \mu/4$, so pointwise $X \leq \mu/4 + X^2/(2t) + (t/2) \mathbf{1}[\text{good}]$ with $t = 4(1 + \mu)/3$; taking expectations gives $\mathbb{P}[\text{good}] \geq \frac{3\mu}{4t} \geq 3/16$ (a Paley–Zygmund bound). Unlike PUSH, the reverse Markov argument of Lemma 20 does not apply, since a PULL round can inform many more than $|I|$ new nodes. $\square$

Theorem 36 (PULL and PUSH–PULL inform K_n in $O(\log n)$ rounds w.h.p.). *Let $n \geq 2$ and $v_0 \in V$. After $\lceil 160 \log n \rceil$ rounds of PULL, or of PUSH–PULL, started from $\{v_0\}$, all nodes are informed with probability at least $1 - 2/n$.*

Proof. For PULL, apply Lemma 34 with the inputs of Lemma 35. For PUSH–PULL, the failure event is contained in the PULL failure event by Lemma 31. $\square$

11 The Lean formalization

The argument above is fully formalized in the accompanying repository (library `RumorSpread`, namespace `RumorPush`); the build contains no `sorry` and the main theorem depends only on the three standard axioms (`propext`, `Classical.choice`, `Quot.sound`). File structure (dependency order): `Bounds.lean` (Section 4), `Dynamics.Uniform` of the shared `dynamics` library (Section 3), `Model.lean` (Section 1), `OneRound.lean` (Section 5), `Growth.lean` (Sections 6–7), `Saturation.lean` (Section 8), `Main.lean` (main theorem and numeric lemmas), `Equivalence.lean` (Theorem 11). PULL and PUSH–PULL (Section 10): `PullFold1.lean` and `PullModel.lean` (Definition 30, Lemma 31), `PullIndep.lean` (Lemma 32), `PullOneRound.lean` (Lemma 33), `PullPhases.lean` (Lemma 34), `PullGood.lean` (Lemma 35), `PullMain.lean` and `PullPushPull.lean` (Theorem 36).

References

- [1] A. Frieze, G. Grimmett. *The shortest-path problem for graphs with random arc-lengths*. Discrete Applied Mathematics 10 (1985) 57–77.
- [2] B. Pittel. *On spreading a rumor*. SIAM J. Appl. Math. 47 (1987) 213–223.
- [3] R. Karp, C. Schindelhauer, S. Shenker, B. Vöcking. *Randomized rumor spreading*. FOCS 2000.