CSC 2421H : Graphs, Matrices, and Optimization Dec 22 2022

The Coalescing-Branching Random Walk on Expanders

Lecturer: Sushant Sachdeva *Report: Sasha Voitovych*

1 The Coalescing-Branching Random Walk

1.1 Refresher on random walks

We first give a refresher on some concepts from the study of random walks on graphs. For simplicity, let's consider an undirected unweighted graph.

Definition 1.1. *Let $G = (V, E)$ be an undirected unweighted graph. Then random walk started at $s \in V$ is a Markov process $(X_t)_{t\geq 0}$ such that $X_0 = s$ and $X_{t+1} \mid X_t$ follows a uniform distribution over the neighbourhood of X_t in G for every $t \geq 0$.*

Simply speaking, a random walk can be described by movement of a particle, which at each step moves to a uniformly at random chosen neighbour of its current position. Figure 1a gives a simple example of a random walk moving for three time steps.

We call the *cover time* $\mathbf{cov}(v)$ of the vertex $v \in V$ the minimum time the random walk visits all vertices in the graph G. We call the cover time $\mathbf{cov}(G)$ of the graph G the maximal expected cover time $\mathbf{cov}(v)$ across all $v \in V$. In other words,

$$\mathbf{cov}(G) = \max_{v \in V} \mathbb{E}\left[\mathbf{cov}(v)\right].$$

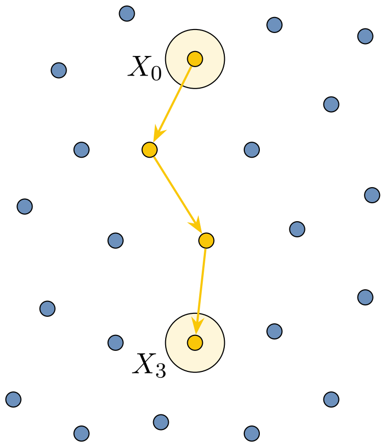
Work of [Ald89] showed that that cover time of *any* graph must be in $\Omega(n \log(n))$. On the other hand, the cover time of any (d, λ)-expander with $\lambda < 1$ is in $O(n \log(n))$ (see, e.g., Theorem 5 of [BK88]). In other words, expander graphs give an optimal cover time by a random walk.

1.2 Introducing k-COBRA walk

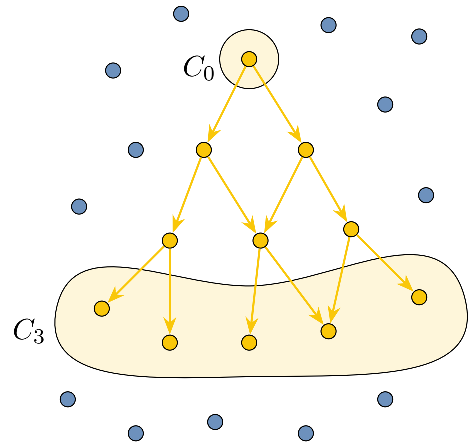
We introduce a generalized version of a simple random walk, called *a k-coalescing branching random walk*, or k-COBRA walk for short. It was first introduced in [Dut+]. This report is based on the results of [CRR16].

First, we give a verbal description. Let $k \in \mathbb{N}$ and consider a graph $G = (V, E)$ and a single particle located at a vertex $v \in V$. At each time step, each particle *branches* into k particles, which travel neighbouring chosen independently and uniformly at random. Then, at the end of the time step, all particles at the same location *coalesce* into one particle. We will use $C_t \subseteq V$ to denote the set of locations of particles at time t. We will call C_t *the active set* of the COBRA walk at time t. See Figure 1b for a simple realization of a 2-COBRA walk after three time steps.

We give a more formal definition below. For a non-empty set S, let $\mathcal{U}_k(S)$ be a distribution over the subsets of S obtained by picking k elements of S uniformly at random *with replacement.*



(a) Example of a random walk.



(b) Example of a 2-COBRA walk.

Figure 1: A comparison between a simple random walk and a 2-COBRA walk. Figure 1a and figure 1b show a realization of a random walk and a 2-COBRA walk respectively after three time steps. Here, X_t and C_t denote position of a random walk and an active set of a COBRA walk respectively at time t.

Definition 1.2. *Let $k \in \mathbb{N}$ and let $G = (V, E)$ be an undirected unweighted graph. Then a k-COBRA walk started at $v \in V$ is a Markov process $(C_t)_{t\geq 0}$ such that $C_0 = \{v\}$ and the distribution $C_{t+1} \mid C_t$ can be described as follows. For all $u \in C_t$, consider independent random variables $B(u) \sim \mathcal{U}_k(N(u))$. Then $C_{t+1} \mid C_t$ is the same in law as $\cup_{u \in C_t} B(u)$.*

Note that for $k = 1$, k-COBRA walk is just a simple random walk. For simplicity of exposition, we will focus in the rest of the report on analyzing a 2-COBRA walk. We aim to establish the following theorem about the covering time of 2-COBRA walk on expanders[1].

Theorem 1.3. *Let G be a (d, λ)-expander with $n - 1 \geq d \geq 3$ and constant $\lambda < 1$. Then covering time of the 2-COBRA walk on G is in $O(\log(n))$ in expectation and with high probability.*

In this report, "with high probability" means with probability at least $1 - O\left(1/n\right)$. In other words, 2-COBRA walk has an *exponential speed up* in covering time compared to a simple random walk (a.k.a., 1-COBRA walk), which has covering time in $\Omega(n \log(n))$.

2 Dual BIPS process

Covering time of a COBRA walk is challenging to analyze directly, for which we give a heuristic explanation. Consider the evolution of a 2-COBRA walk on the graph $G = (V, E)$, and let us look at the moment when the size of the active set C_t is roughly half of the size of V. Then the probability that any two particles coalesce in the following round is also roughly $1/2$. Then the size of the next active set C_{t+1} will be roughly $1/2 \cdot 2|C_t|$ (where factor 2 comes from every particle branching, and $1/2$ comes from the probability of coalescing). In other words, after time t the size of the active set *doesn't grow*, but just shuffles through V.

[1] In fact, result for 2-COBRA walk also implies the result for k-COBRA walk with $k \geq 3$.

It's very difficult to analyze how exactly this set shuffles through V due to the collisions. We circumvent this by considering a *dual* process on G called Biased Infection with Persistent Source (or BIPS), which relates to a COBRA walk, but does not not have the "shuffling" phase.

2.1 Biased Infection with Persistent Source

We first give a verbal description of k-BIPS process. Consider graph $G = (V, E)$ and a single infected node $v \in V$. We will call v the source of the infection, and v will stay permanently infected. At every time step, every node (except v) selects k of its neighbours uniformly at random with replacement. If one of its neighbours is infected, the node becomes infected. We will use A_t to denote the infected set of the BIPS process at time t. It can be defined more formally in the following way.

Definition 2.1. *Let $k \in \mathbb{N}$ and let $G = (V, E)$ be an undirected unweighted graph. Then a k-BIPS walk started at $v \in V$ is a Markov process $(A_t)_{t\geq 0}$ such that $A_0 = \{v\}$ and the distribution $A_{t+1} \mid A_t$ can be described as follows. For all $u \in V$, consider independent random variables $B'(u) \sim \mathcal{U}_k(N(u))$. Then $A_{t+1} \mid A_t$ is the same in law as $\{v\} \cup \{u \in V : A_t \cap B'(u) \neq \emptyset\}$.*

We note that if all nodes become infected (i.e., $A_t = V$), then the graph will stay infected (i.e., $A_{t+1} = V$). We will show that not only this happens with probability 1, but also that it happens quite fast, e.g., in $O(\log(n))$ time.

2.2 Duality with COBRA walk

The hitting times of a COBRA walk can be upper bounded using the BIPS process. We will denote $\mathbf{hit}(C, v)$ to be the *hitting time* of a vertex v by a 2-COBRA walk started from an active set $C_0 = C$. The following result relates hitting times of a COBRA walk to infection times of the BIPS process.

Lemma 2.2. *Let $C \subseteq V$ and $v \in V$. Let A_t be the infected set of 2-BIPS process. For every $t \geq 0$*

$$\mathbb{P}\left[\mathbf{hit}(C, v) > t\right] = \mathbb{P}\left[A_t \cap C = \emptyset \mid A_0 = \{v\}\right].$$

Proof. We will prove the statement by induction. For $t = 0$, both probabilities are 1 if $v \in C$ and both are 0 if $v \notin C$. This establishes the induction base.

Suppose the lemma statement holds for t. We will show it also holds for $t + 1$. First, consider $\mathbb{P}\left[A_{t+1} \cap C = \emptyset\right]$. Consider the k-BIPS process at time step $t+1$ and consider k random neighbours each node from C selects. For $u \in C$, let $B'(u) \sim \mathcal{U}_k(N(u))$ be such random subset of neighbours. Node u does not become infected at step $t + 1$ if and only if $B'(u) \cap A_t = \emptyset$ (i.e., node u didn't sample an infected node among it's neighbours). Denote $B'(C) = \cup_{u\in C} B'(u)$. Then none of the nodes in C are infected at round $t + 1$ iff $B'(C) \cap A_t = \emptyset$. In other words,

$$\mathbb{P}\left[A_{t+1} \cap C = \emptyset\right] = \sum_{B\subseteq V} \mathbb{P}\left[A_t \cap B = \emptyset\right] \mathbb{P}[B'(C) = B] = \sum_{B\subseteq V} \mathbb{P}[\mathbf{hit}(B, v) > t]\mathbb{P}[B'(C) = B], \quad (1)$$

where second transition is by induction hypothesis. On the other hand, consider the COBRA walk starting with $C_0 = C$ at step 0. For $u \in C$, let $B(u) \sim \mathcal{U}_k(u)$ to be the random set of neighbours to which particles travel from u in the next round. Then $B(C) = \cup_{u\in C} B(u)$ will be the active set of the COBRA walk at time 1. Then

$$\mathbb{P}\left[\mathbf{hit}(C, v) > t + 1\right] = \sum_{B\subseteq V} \mathbb{P}\left[\mathbf{hit}(B, v) > t\right] \mathbb{P}[B(C) = B]. \quad (2)$$

On the other hand, $B'(u)$ is the same in law as $B(u)$. Then $B'(C) = \cup_{u \in C} B'(u)$ is the same in law as $B(C) = \cup_{u \in C} B(u)$. Then, for any B, we have $\mathbb{P}[B(C) = B] = \mathbb{P}[B'(C) = B]$. Then, from (1) and (2) we may conclude

$$\mathbb{P}\left[A_{t+1} \cap C = \emptyset \mid A_0 = \{v\}\right] = \mathbb{P}\left[\mathbf{hit}(v) > t+1 \mid C_0 = C\right].$$

□

Let $\mathbf{infec}(v)$ be the time it takes for the 2-BIPS process to infect the whole graph. Letting $C = \{u\}$ in Lemma 2.2, we obtain

$$\mathbb{P}[\mathbf{infec}(v) > t] \geq \mathbb{P}\left[u \notin A_t \mid A_0 = \{v\}\right] = \mathbb{P}\left[\mathbf{hit}(u, v) > t\right]. \tag{3}$$

In the rest of the report we will aim to show that $\mathbf{infec}(v) \in O(\log(n))$ with probability at least $1 - O\left(\frac{1}{n^3}\right)$. This will imply that for a fixed u, $\mathbf{cov}(u) \in O(\log(n))$ with probability $1 - O\left(\frac{1}{n^2}\right)$, which also implies the result in expectation[2].

3 Infection time of 2-BIPS process

We now give a result for the expected growth rate of the 2-BIPS process, which is essential to the rest of the proof. Unlike 2-COBRA walk, infected set of the 2-BIPS set grows (in expectation) at every step.

Lemma 3.1. *Let A_t be the infected set of the 2-BIPS process at step t. Then for any $A \subseteq V$*

$$\mathbb{E}\left[|A_{t+1}| \mid A_t = A\right] \geq |A| \left(1 + (1 - \lambda^2)\left(1 - \frac{|A|}{n}\right)\right).$$

Proof. Suppose $A_t = A$. Let $\mathbb{I}(u \in A_{t+1})$ be the indicator random variable of the event $u \in A_{t+1}$

$$|A_{t+1}| = \sum_{u \in V} \mathbb{I}(u \in A_{t+1}).$$

Note that $\mathbb{I}(v \in A_{t+1})$ is always 1, and for $u \neq v$ we have

$$\mathbb{E}[\mathbb{I}(u \in A_{t+1}) \mid A_t = A] = \left(1 - \left(1 - \frac{\deg_A(u)}{d}\right)^2\right) = \frac{2\deg_A(u)}{d} - \frac{\deg_A(u)^2}{d^2}.$$

Therefore,

$$\begin{aligned}\mathbb{E}[|A_{t+1}| \mid A_t = A] &= 1 + \sum_{u \in V \setminus \{v\}} \left(\frac{2\deg_A(u)}{d} - \frac{\deg_A(u)^2}{d^2}\right) \geq \sum_{u \in V} \left(\frac{2\deg_A(u)}{d} - \frac{\deg_A(u)^2}{d^2}\right) \\ &= 2|A| - \sum_{u \in V} \left(\frac{\deg_A(u)^2}{d^2}\right).\end{aligned}$$

[2]Here and in the rest of the report $\mathbf{cov}(u)$ refers to the cover time of a 2-COBRA walk started from vertex v

Let $\mathbf{A}$ be the adjacency matrix of G and let $\hat{\mathbf{A}} = \frac{1}{d}\mathbf{A}$ be its normalized version. Consider the vector $\hat{\mathbf{A}}\mathbb{1}_A$. It's u^{th} coordinate will be equal to

$$\left(\hat{\mathbf{A}}\mathbb{1}_A\right)_u = \frac{\deg_A(u)}{d}.$$

Then

$$\sum_{u \in V} \left(\frac{\deg_A(u)^2}{d^2}\right) = \left(\hat{\mathbf{A}}\mathbb{1}_A\right)^\top \left(\hat{\mathbf{A}}\mathbb{1}_A\right) = \mathbb{1}_A^\top \mathbf{A}^{\mathbf{2}} \mathbb{1}_{\mathbf{A}}.$$

Next, we express $\mathbb{1}_A$ as a sum of (1) it's projection onto $\mathbb{1}$ and (2) component orthogonal to $\mathbb{1}$. More precicely

$$\mathbb{1}_A = \mathbb{1}_A^{\|} + \mathbb{1}_A^{\perp},$$

where $\mathbb{1}_A^{\|} = \frac{\langle \mathbb{1}_A, \mathbb{1} \rangle}{\|\mathbb{1}\|_2^2}\mathbb{1} = \frac{|A|}{n}\mathbb{1}$ and $\langle \mathbb{1}_A^{\perp}, \mathbb{1} \rangle = 0$. Then

$$\sum_{u \in V} \left(\frac{\deg_A(u)^2}{d^2}\right) = \mathbb{1}_A^\top \mathbf{A}^{\mathbf{2}} \mathbb{1}_{\mathbf{A}}$$

since $\hat{\mathbf{A}}\mathbb{1}_A^{\|} = \mathbb{1}_A^{\|}$ and $\mathbb{1}_A^{\perp}$ are orthogonal

$$\begin{aligned}
&= (\mathbb{1}_A^{\|})^\top \hat{\mathbf{A}}^2 (\mathbb{1}_A^{\|}) + (\mathbb{1}_A^{\perp})^\top \hat{\mathbf{A}}^2 (\mathbb{1}_A^{\perp}) \\
&= \frac{|A|^2}{n^2} \mathbb{1}^\top \hat{\mathbf{A}}^2 \mathbb{1} + (\mathbb{1}_A^{\perp})^\top \hat{\mathbf{A}}^2 (\mathbb{1}_A^{\perp})
\end{aligned}$$

since $\mathbb{1}$ is an eigenvector of $\hat{\mathbf{A}}$

$$= \frac{|A|^2}{n} + (\mathbb{1}_A^{\perp})^\top \hat{\mathbf{A}}^2 (\mathbb{1}_A^{\perp})$$

since all eigenvalues of $\mathbf{A}$ are upper bounded by λ in absolute value

$$\begin{aligned}
&\leq \frac{|A|^2}{n} + \lambda^2 (\mathbb{1}_A^{\perp})^\top (\mathbb{1}_A^{\perp}) \\
&= \frac{|A|^2}{n} + \lambda^2 \left\|\mathbb{1}_A^{\perp}\right\|_2^2.
\end{aligned}$$

By Pythagorean theorem,

$$\|\mathbb{1}_A\|^2 = \left\|\mathbb{1}_A^{\|}\right\|^2 + \left\|\mathbb{1}_A^{\perp}\right\|^2 \Rightarrow \left\|\mathbb{1}_A^{\perp}\right\|^2 = \|\mathbb{1}_A\|^2 - \left\|\mathbb{1}_A^{\|}\right\|^2 = |A| - \frac{|A|^2}{n}.$$

Hence,

$$\sum_{u \in V} \left(\frac{\deg_A(u)^2}{d^2}\right) \leq \frac{|A|^2}{n} + \lambda^2 \left(|A| - \frac{|A|^2}{n}\right) = \left(1 - \lambda^2\right) \frac{|A|^2}{n} + \lambda^2 |A|.$$

Therefore,

$$\mathbb{E}[|A_{t+1}| \mid A_t = A] \geq 2|A| - \left(1 - \lambda^2\right) \frac{|A|^2}{n} - \lambda^2 |A| = |A| \left(1 + (1 - \lambda^2)\left(1 - \frac{|A|}{n}\right)\right).$$

$\square$

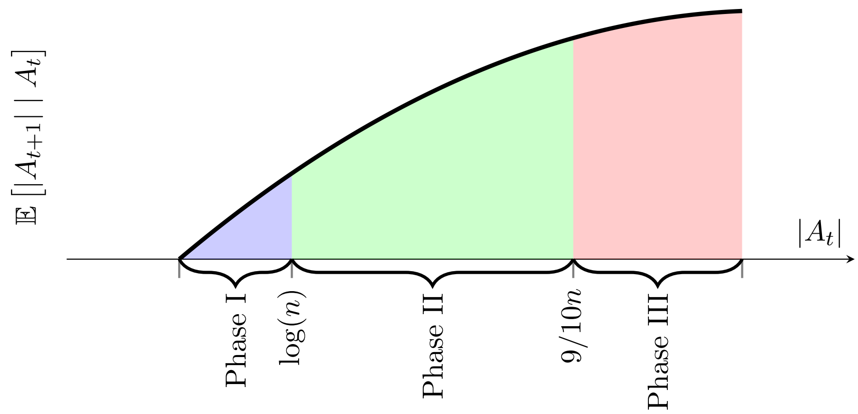


Figure 2: Growth of BIPS, as given by the lower bound in Lemma 3.1. In phase I, the active set is sublogarithmic. In phase II the active set is at least logarithmic and at most $9/10n$. In phase III, and the active set is at least $9/10n$ and reaches size n at the end. Each phase takes $O(\log(n))$ rounds.

Due to space limitations, we only give the skeleton for the rest of the proof. We first observe that if $|A_t|/n \in 1 - \Omega_n(1)$, then $|A_{t+1}|$ is larger in expectation than $|A_t|$ by at least a constant factor. In other words, $|A_t|$ grows exponentially in expectation as long as $|A_t|/n$ is bounded away from 1 by a constant. However, we need to prove a stronger statement than that, namely that A_t grows *with high probability*.

Notice that we can write $|A_{t+1}|$ as a sum of indicator random variables $\mathbb{I}(u \in A_{t+1})$ as follows

$$|A_{t+1}| = \sum_{u \in V} \mathbb{I}(u \in A_{t+1}). \tag{4}$$

Moreover, it follows from Definition 2.1 that if we condition on A_t, the indicators $\mathbb{I}(u \in A_{t+1})$ are *independent* Bernoulli variables. Then, if $\mathbb{E}[|A_{t+1}| \mid A_t]$ is sufficiently large (but still far from n), we can show that $|A_{t+1}|$ is larger than $|A_t|$ by at least a constant fraction via *Chernoff bounds*. This observation lays the base for the following proof structure (it's illustrated in Figure 2).

Phase I: Growth on small sets First, we show that the infected set A_t grows to size $\Omega(\log(n))$ in at most $O(\log(n))$ time steps. The proof is similar in spirit to Chernoff bounds. We define E_t to be the event that all $A_0, A_1, \ldots, A_t$ are less than some $K\log(n)$. By applying an exponential Markov inequality to $|A_t|$ and using indicator representation as in (4) we obtain $\mathbb{P}[E_t]$ is less than $\mathbb{P}[E_{t-1}]$ by at least a constant factor for every t. Then, by induction, probability $\mathbb{P}[E_t]$ is in $O(1/n^3)$ (for a suitable choice of t). Formal statement is given below.

Lemma 3.2. *Let $(A_t)_{t\geq 0}$ be 2-BIPS process started from an arbitrary vertex $v \in V$. Then for any constant $K > 0$, there exist constant $c_1 = c_1(K)$ and $t_1 \leq c_1 \log(n)$ such that $|A_{t_1}| \geq K\log(n)$ with probability at least $1 - O\left(\frac{1}{n^3}\right)$.*

Phase II: Steady growth Now, suppose $|A_t| \geq K\log(n)$ for sufficiently large K. Then we can apply Chernoff bounds to (4) and obtain that with probability $1 - O(1/n^4)$, the active set grows by at least a constant factor in the next step. Then, in $O(\log(n))$ rounds, the active set grows to the size $\geq 9/10n$ with probability $1 - O(1/n^3)$. The formal statement is given below.

Lemma 3.3. *Let $(A_t)_{t\geq 0}$ be 2-BIPS process. Suppose $t_1 > 0$ is such that $A_{t_1} \geq K\log(n)$ for large enough constant K. Then there exist a constant $c_2 = c_2(K)$ such that for $t_2 = t_1 + c_2\log(n)$, we have $|A_{t_2}| \geq 9/10n$ with probability $1 - O\left(\frac{1}{n^3}\right)$.*

Phase III: Infecting everyone Now, suppose $|A_t| \geq 9/10n$. We look at $V \setminus A_t$ (i.e. set of uninfected nodes). We can show that with very high probability it shrinks by a constant factor in expectation at each time step. Then, in $O(\log(n))$ rounds, we will have $\mathbb{E}[|V \setminus A_t|] \leq 1/n^3$, which by Markov inequality implies $\mathbb{P}[|V \setminus A_t| \geq 1] \leq 1/n^3$.

Lemma 3.4. *Let $(A_t)_{t\geq 0}$ be 2-BIPS process. Suppose $t_2 > 0$ is such that $A_{t_2} \geq 9/10n$. Then there exist a constant c_3 such that for $t_3 = t_2 + c_3\log(n)$, we have $A_{t_3} = V$ with probability $1 - O\left(\frac{1}{n^3}\right)$.*

Proof of Theorem 1.3 By gluing Lemmas 3.2, 3.3, 3.4 together with Lemma 2.2 we obtain the proof of Theorem 1.3.

Proof of Theorem 1.3. Lemmas 3.2, 3.3, 3.4 together imply that there exist $c > 0$ such that for any $v \in V$ we have $\mathbf{infec}(v) \leq t = c\log(n)$ with probability $1 - O\left(\frac{1}{n^3}\right)$.

Then, by Lemma 2.2 for $C = \{u\}$, we have

$$\mathbb{P}\left[\mathbf{hit}(u,v) > t\right] = \mathbb{P}\left[u \notin A_t \mid A_0 = \{v\}\right] \leq \mathbb{P}\left[\mathbf{infec}(v) > t\right] \in O\left(\frac{1}{n^3}\right).$$

Then

$$\mathbb{P}[\mathbf{cov}(u) > t] \leq \sum_{v\in V} \mathbb{P}\left[\mathbf{hit}(u,v) > t\right] \in O\left(\frac{1}{n^2}\right).$$

Then $\mathbf{cov}(u) < t$ with high probability.

Now we establish a bound in expectation. Consider the cobra walk at time t. Suppose it didn't cover G yet, which happens with probability at most $O\left(\frac{1}{n^2}\right)$. Then we will follows a single branch of such a COBRA walk starting step t. It follows in distribution a simple random walk, which covers the graph in $O(n\log(n))$ expected steps. Then

$$\mathbb{E}\left[\mathbf{cov}(u)\right] \in t + O\left(\frac{1}{n^2}\right) \cdot O(t + O(n\log(n))) \subseteq O(\log(n)).$$

□

References

[Ald89] David J. Aldous. "Lower bounds for covering times for reversible Markov chains and random walks on graphs". In: *Journal of Theoretical Probability* 2 (1989), pp. 91–100 (cit. on p. 1).

[BK88] Andrei Z. Broder and Anna R. Karlin. "Bounds on the cover time". In: *Journal of Theoretical Probability* 2 (1988), pp. 101–120 (cit. on p. 1).

[CRR16] Colin Cooper, Tomasz Radzik, and Nicolas Rivera. "The Coalescing-Branching Random Walk on Expanders and the Dual Epidemic Process". In: *Proceedings of the 2016 ACM Symposium on Principles of Distributed Computing*. PODC '16. Chicago, Illinois, USA: Association for Computing Machinery, 2016, pp. 461–467 (cit. on p. 1).

[Dut+] Chinmoy Dutta, Gopal Pandurangan, Rajmohan Rajaraman, and Scott Roche. "Coalescing-Branching Random Walks on Graphs". In: *Proceedings of the Twenty-Fifth Annual ACM Symposium on Parallelism in Algorithms and Architectures (SPAA 2013)*. Montréal, Québec, Canada (cit. on p. 1).