

On coalescence time in graphs: When is coalescing as fast as meeting?

Extended Abstract^{*}

Varun Kanade[†]

Frederik Mallmann-Trenn[‡]

Thomas Sauerwald[§]

Abstract

Coalescing random walks is a fundamental stochastic process, where a set of particles perform independent discrete-time random walks on an undirected graph. Whenever two or more particles meet at a given node, they merge and continue as a single random walk. The *coalescence time* is defined as the expected time until only one particle remains, starting from one particle at every node. Despite recent progress such as by Cooper, Elsässer, Ono, Radzik [13] and Cooper, Frieze and Radzik [12], the coalescence time for graphs such as binary trees, d -dimensional tori, hypercubes and more generally, vertex-transitive graphs, remains unresolved.

We provide a powerful toolkit that results in tight bounds for various topologies including the aforementioned ones. The meeting time is defined as the worst-case expected time required for two random walks to arrive at the same node at the same time. As a general result, we establish that for graphs whose meeting time is only marginally larger than the mixing time (a factor of $\log^2 n$), the coalescence time of n random walks equals the meeting time up to constant factors. This upper bound is complemented by the construction of a graph family demonstrating that this result is the best possible up to constant factors. For almost-regular graphs, we bound the coalescence time by the hitting time, resolving the discrete-time variant of a conjecture by Aldous for this class of graphs. Finally, we prove that for any graph the coalescence time is bounded by $O(n^3)$ (which is tight for the Barbell graph); surprisingly even such a basic question about the coalescing time was not answered before this work. By duality, our results give bounds on the voter model and therefore give bounds on the consensus time in arbitrary undirected graphs.

We also establish a new bound on the hitting time and cover time of regular graphs, improving and tightening previous results by Broder and Karlin [10], as well as those by Aldous and Fill [1].

1 Introduction

Coalescing random walks is a fundamental stochastic process on *connected* and *undirected* graphs. The

process begins with particles on some subset of the nodes in the graph. At discrete time-steps, every particle performs one step of an independent random walk.¹ Whenever two or more particles arrive at the same node at the same time-step, they merge into a single particle and continue as a single random walk. The *coalescence time* is defined as the first time-step when only one particle remains. The coalescence time depends on the number and starting positions of the particles.

Studying the coalescence time is of substantial importance in distributed computing: At the heart of many distributed computing applications lie consensus protocols and leader election *e.g.*, data consistency, consolidation of replicated states, synchronization of processes and devices [33, 22] and communication networks [34]). Other applications of the coalescence process appear in robotics [25]; here, robots perform random walks to gather samples from their environment and need to communicate these samples to all other robots. Studying the coalescence time also implies results for other interaction types of random walks including predator and prey particles as well as annihilating particles [17].

1.1 Relationship to consensus protocols: Arguably the simplest consensus protocol achieving consensus on any undirected graph is the voter model. Initially, every node has a distinct opinion. At every round, each node chooses synchronously one of its neighbors at random and adopts that node's opinion. The *consensus time* is defined as the time it takes until only one opinion remains. The voting process viewed backwards is exactly the same as the coalescence process starting with a random walk on every node; thus, the coalescence time and consensus time have the same distribution. Despite recent progress by Cooper *et al.* [13, 12] and Berenbrink *et al.* [9], the coalescence time and consensus time are far from being well-understood—even for certain fundamental graphs as we describe below. Recently, there have been several studies on variants of

^{*}A preprint containing details of the results presented in this extended abstract is available as [28]. The text refers to v4 of the preprint.

[†]University of Oxford. Email: varunk@cs.ox.ac.uk. This work was supported in part by the Alan Turing Institute under the EPSRC award EP/N510129/1.

[‡]MIT. Email: mallmann@mit.edu. Supported in part by NSF Award Numbers CCF-1461559, CCF-0939370, and CCF-1810758.

[§]University of Cambridge. Email: thomas.sauerwald@cl.cam.ac.uk. Supported in part by the ERC Starting Grant “Dynamic March”.

¹Throughout this paper, we use random walk and particle interchangeably, assuming that every random walk has an identifier.

the voter model, most notably 2-Choices and 3-Majority which received ample attention [14, 7, 15, 6, 18, 24, 8, 26]. However, the behavior of these processes is fundamentally different and despite their efficiency in reaching consensus on expanders and cliques, they are unsuitable on more general undirected graphs as the consensus time is exponential in some graphs.

1.2 Related Work: In this paper, we follow the approach of Cooper et al. [13] and Hassin and Peleg [27] and study the consensus time through the more tangible analysis of the coalescence time. When starting with two particles, the coalescence time is referred to as the *meeting time*. Let t_{meet} denote the worst-case expected meeting time over all pairs of starting nodes and let t_{coal} denote the expected coalescence time starting from one particle on every node. It is clear that $t_{\text{meet}} \leq t_{\text{coal}}$; as for an upper bound, it can be shown that $t_{\text{coal}} = O(t_{\text{meet}} \log n)$, where n is the number of nodes in the graph. The main idea used to obtain the bound is that the number of surviving random walks halves roughly every t_{meet} steps. A proof of the result appears implicitly in the work of Hassin and Peleg [27].

Aldous [2] showed in *continuous-time* that the meeting time is bounded by the maximum hitting time, $t_{\text{hit}} := \max_{u,v} t_{\text{hit}}(u, v)$, where $t_{\text{hit}}(u, v)$ denotes the expected time required to hit v starting from vertex u . We observe that the result of Aldous also holds in discrete time. Thus, this gives a bound of $O(t_{\text{hit}} \log n)$ for the coalescing time; however, in general $O(t_{\text{hit}})$ may be a loose upper bound on t_{meet} . In recent work, Cooper et al. [13] provide results that are better than $O(t_{\text{meet}} \log n)$ for several interesting graph classes, notably expanders and power-law graphs. They show that $t_{\text{coal}} = O((\log^4 n + \|\pi\|_2^{-2}) \cdot (1 - \lambda_2)^{-1})$, where λ_2 is the second largest eigenvalue of the transition matrix of the random walk and π is the stationary distribution. Berenbrink et al. [9] show that $t_{\text{coal}} = O(m/(d_{\min} \cdot \Phi))$, where m is the number of edges, d_{\min} is the minimum degree and Φ is the conductance. Their result improves on that of Cooper et al. for certain graph classes, e.g., cycles.

As mentioned before, despite the recent progress due to Cooper et al. [13] and Berenbrink et al. [9], for many fundamental graphs such as the binary tree, hypercube and the (d -dimensional) torus, the coalescing time in the discrete setting remains unsettled. We provide a rich toolkit allowing us to derive tight bounds for many graphs including all of the aforementioned ones. One of our main results establishes a relationship between the ratios $t_{\text{coal}}/t_{\text{meet}}$ and $t_{\text{meet}}/t_{\text{mix}}$, where

$t_{\text{mix}} = t_{\text{mix}}(1/e)$ denotes the mixing time.² In particular, the result shows that if $t_{\text{meet}}/t_{\text{mix}} = \Omega(\log^2 n)$, then $t_{\text{coal}} = O(t_{\text{meet}})$; however, we also provide a more fine-grained tradeoff. For almost-regular graphs,³ we bound the coalescence time by the hitting time. For vertex-transitive graphs we show that the coalescence time, the meeting time, and the hitting time are equal up to constant factors. Finally, we prove that for any graph the coalescence time is bounded by $O(n^3)$; it can be easily verified that this is tight by considering the barbell graph. Surprisingly, the right bound on this fundamental quantity was not known prior to this work. Unlike in the analogous case of the cover time [1] where such a bound can be easily derived, the argument in the case of coalescence time appears significantly involved.⁴ Prior to this work, Hassin and Peleg [27] had shown a worst-case upper bound of $O(n^3 \log n)$. We also give worst-case upper and lower bounds on the meeting time and coalescence time that are tight for general graphs and regular (or nearly-regular) graphs.

In the process of establishing bounds on the coalescence time, we develop techniques to give tight bounds on the meeting time. We apply these to various topologies such as the binary tree, torus and hypercube. We believe that these techniques might be of more general interest.

The process of coalescing random walks was first studied in *continuous time*; in this case, particles jump to a random neighboring node when activated according to a Poisson clock with mean 1. As Cooper and Rivera [19] recently pointed out “*It is however, not clear whether the continuous-time results apply to the discrete-time setting*”, and to the best of our knowledge, there is no general way in which results in *continuous time* can be transferred to *discrete time* or vice versa, even when the random walks in discrete-time are *lazy*. In the continuous time setting, Cox [20] show that the coalescence time is bounded by $\Theta(t_{\text{hit}})$ for tori. Oliveira [31] showed that the coalescence time is $O(t_{\text{hit}})$ in general. In a different work, Oliveira [32] derived so-called mean field conditions, which are sufficient conditions for the coalescing process on a graph to behave similarly to that on the complete graph up to scaling by the expected meeting time. His main result (for non vertex-transitive graphs) in [32, Theorem 1.2], implies that $t_{\text{coal}} = O(t_{\text{meet}})$

²The *mixing time* is the first time-step at which the distribution of a random walk starting from an arbitrary node is close to the stationary distribution.

³We call a graph *almost-regular* if $\deg(u) = \Theta(\deg(v))$ for all $u, v \in V$.

⁴Cooper et al. [13] mistakenly stated, as a side remark, that this last result was a simple consequence of their main result.

whenever $t_{\text{mix}} \cdot \pi_{\text{max}} = O(1/\log^4 n)$. One of our main results, Theorem 2.1, implies $t_{\text{coal}} = O(t_{\text{meet}})$ whenever $t_{\text{mix}}/t_{\text{meet}} = O(1/\log^2 n)$. Notice that since $t_{\text{meet}} \geq 1/(\|\pi\|_2^2) \geq 1/\pi_{\text{max}}$, our condition is considerably more general—however, the results in [32] also establish mean-field behavior (that is, when suitably scaled, the distribution of the coalescence time is similar to that on a complete graph), while ours are only concerned with the expected coalescence time, t_{coal} . On the other hand, our result also applies to graphs where $t_{\text{coal}} \gg t_{\text{meet}}$ such as the star graph, and together with Theorem 2.2, demonstrate that the trade-off between meeting and mixing time is the best possible.

2 Contributions

In this work, we provide several results relating the coalescence and meeting times to each other and to other fundamental quantities of random walks on undirected graphs. In particular, our focus is on understanding for which graphs the coalescence time is the same as the meeting time, as we know that t_{coal} is always in the rather narrow interval of $[t_{\text{meet}}, O(t_{\text{meet}} \cdot \log n)]$. As a consequence of our results, we derive new and re-derive existing bounds on the meeting and coalescence times for several graph families of interest. These results are summarized in Table 1 and discussed in greater detail in Appendix C of [28]. Formal definitions of all quantities used below appear in Section 2 of [28]. Throughout this paper, we assume that random walks are *lazy* meaning that w.p. $1/2$ the walk stays put.

Our first main result relates t_{coal} to t_{meet} and t_{mix} . As already mentioned in the introduction, the crude bound $t_{\text{coal}} = O(t_{\text{meet}} \log n)$ is well-known. However, this bound is not in general tight, as demonstrated by our result below.

THEOREM 2.1. *For any graph G , we have*

$$t_{\text{coal}} = O\left(t_{\text{meet}} \left(1 + \sqrt{\frac{t_{\text{mix}}}{t_{\text{meet}}}} \cdot \log n\right)\right),$$

Consequently, when $t_{\text{meet}} \geq t_{\text{mix}} \log^2 n$, $t_{\text{coal}} = O(t_{\text{meet}})$.

The proof of Theorem 2.1 appears in Section 3 of [28]. One interesting aspect about this bound is that it can be used to establish $t_{\text{coal}} = \Theta(t_{\text{meet}})$ even without having to know the quantities t_{meet} or t_{mix} . This flexibility turns out to be particularly useful when dealing with random graph models for “real world” networks, where we establish (nearly-)tight and sublinear bounds (w.r.t. to the number of vertices) in Appendix C.5 of [28].

Another interesting feature of our theorem is that the main result of Cooper *et al.* [13, Theorem 1]

can be reproven by combining [13, Theorem 2] with Theorem 2.1 (see Prop. B.2 in [28]).

Our next main result shows that the bound in Theorem 2.1 is tight up to a constant factor, which we establish by constructing an explicit family of graphs. Interestingly, for this family of almost-regular graphs we also have $t_{\text{hit}} \gg t_{\text{meet}}$, thus showing that t_{hit} may be a rather loose upper bound for t_{coal} in some cases.⁵

THEOREM 2.2. *For any sequence $(\alpha_n)_{n \geq 0}$, $\alpha_n \in [1, \log^2 n]$ there exists a family of almost-regular graphs (G_n) , with G_n having $\Theta(n)$ nodes and satisfying $\frac{t_{\text{meet}}}{t_{\text{mix}}} = \Theta(\alpha_n)$ such that*

$$t_{\text{coal}} = \Omega\left(t_{\text{meet}} \cdot \left(1 + \sqrt{\frac{t_{\text{mix}}}{t_{\text{meet}}}} \cdot \log n\right)\right).$$

The above two results show that that $t_{\text{meet}}/t_{\text{mix}}$ should be $\Omega(\log^2 n)$ to guarantee that $t_{\text{coal}} = O(t_{\text{meet}})$.

A natural question is therefore whether in the case of structured sub-classes such as regular graphs, or vertex-transitive graphs, or special graphs such as grids, tori, binary trees, cycles, real-world (power-law) graphs, *etc.*, better bounds can be obtained through other methods. We provide results that are tight or nearly tight in several of these cases; some of these results were previously known using other methods, some are novel to the best of our knowledge.

THEOREM 2.3. *The following hold for graphs of the stated kind*

(i) *For any graph G ,*

$$t_{\text{coal}} = O(t_{\text{hit}} \cdot \log \log n).$$

(ii) *For any graph G with maximum degree Δ and average degree d ,*

$$t_{\text{coal}} = O(t_{\text{hit}} + t_{\text{meet}} \cdot \log(\Delta/d)).$$

Hence for any almost-regular graph G , $t_{\text{coal}} = O(t_{\text{hit}})$.

(iii) *For any vertex-transitive G ,*

$$t_{\text{coal}} = \Theta(t_{\text{meet}}) = \Theta(t_{\text{hit}}).$$

(iv) *In the case of binary trees, d -dimensional tori/grids, paths/cycles, expanders, hypercubes, random power law graphs,⁶ we have $t_{\text{coal}} = \Theta(t_{\text{meet}})$.*

⁵Note that the star also exhibits $t_{\text{hit}} \gg t_{\text{meet}}$. However, the star is not almost-regular.

⁶The exact model is specified in Appendix C.5 of [28].

Graph	t_{mix}	t_{meet}		t_{coal}		t_{hit}
Binary tree	$\Theta(n)$	$\Theta(n \log n)$	Thm. 2.3 & Thm. C.6[28]	$\Theta(n \log n)$	Thm. 2.3 & Thm. C.6[28]	$\Theta(n \log n)$
Clique	$\Theta(1)$	$\Theta(n)$	[13, 9] & Thm. 2.1	$\Theta(n)$	[13, 9] & Thm. 2.1	$\Theta(n)$
Cycle	$\Theta(n^2)$	$\Theta(n^2)$	[9] & Thm. 2.3	$\Theta(n^2)$	[9] & Thm. 2.3	$\Theta(n^2)$
Rand. r -reg.	$\Theta(\log n)$	$\Theta(n)$	[12, 13, 9] & Thm. 2.1	$\Theta(n)$	[12, 13, 9] & Thm. 2.1	$\Theta(n)$
Hypercube	$\Theta(\log n \log \log n)$	$\Theta(n)$	Thm. 2.3	$\Theta(n)$	Thm. 2.1	$\Theta(n)$
Path	$\Theta(n^2)$	$\Theta(n^2)$	[9] & Thm. 2.3	$\Theta(n^2)$	[9] & Thm. 2.3	$\Theta(n^2)$
Star	$\Theta(1)$	$\Theta(1)$	folklore	$\Theta(\log n)$	[27] , Prop. 3.4[28] & Thm. 2.5	$\Theta(n)$
Torus ($d = 2$)	$\Theta(n)$	$\Theta(n \log n)$	Thm. 2.3	$\Theta(n \log n)$	Thm. 2.3	$\Theta(n \log n)$
Torus ($d > 2$)	$\Theta(n^{2/d})$	$\Theta(n)$	Thm. 2.1	$\Theta(n)$	Thm. 2.1	$\Theta(n)$

Table 1: A summary of bounds on the mixing, meeting, coalescence and hitting times for fundamental topologies for discrete-time random walks. All bounds on the mixing and hitting times appear directly or implicitly in [1]. The cited results may refer to those appearing in [28].

The proof of the first three statements of Theorem 2.3 appear in Section 4 and the last statement follows from the results in Appendix C of [28]. We point out that since $t_{\text{meet}} = O(t_{\text{hit}})$ for any graph,⁷ Theorem 2.3 implies the bound $t_{\text{coal}} = O(t_{\text{hit}})$ not only for almost-regular graphs, but also for dense graphs where $|E| = \Theta(n^2)$. This settles the discrete-time analogue of a conjecture by Aldous [1, Open Problem 14.13] for these graph classes. In very recent work, Oliveira and Peres improve on these results and establish that $t_{\text{coal}} = O(t_{\text{hit}})$ holds for all undirected graphs [30].

Another natural question is to express t_{meet} or t_{coal} solely in terms of t_{mix} , the spectral gap $1 - \lambda_2$ or other connectivity properties of G . We derive several such bounds on t_{meet} , t_{hit} and t_{coal} .

As a by-product of our techniques, we also derive new bounds on t_{hit} and t_{cov} , the cover-time. The detailed results are given in Appendix B of [28], but we highlight the results for regular graphs here:

THEOREM 2.4. *Let G be any graph with $\Gamma = \Delta/\delta$, where Δ is the maximum degree and δ the minimum degree. It holds that*

$$t_{\text{hit}} = O(\Gamma n / \sqrt{1 - \lambda_2}) = O(\Gamma n / \Phi),$$

where Φ is the conductance of the graph and λ_2 is the second largest eigenvalue of the transition matrix P of a lazy random walk. Consequently, $t_{\text{meet}} \leq t_{\text{coal}} = O(\Gamma n \log(\Gamma) / \sqrt{1 - \lambda_2}) = O(\Gamma n \log(\Gamma) / \Phi)$ and $t_{\text{cov}} = O(\Gamma n \log n / \Phi)$.

We point out that so far the best possible bound on t_{coal} for regular graphs has been $t_{\text{coal}} = O(n/(1 - \lambda_2))$

from [13].⁸ The best possible bound on t_{hit} (and t_{cov}) in terms of $1 - \lambda_2$, was $t_{\text{hit}} = O(n/(1 - \lambda_2))$ and $t_{\text{cov}} = O(n \log n / (1 - \lambda_2))$ due to Broder and Karlin [10] from 1989. In all four cases, t_{meet} , t_{coal} , t_{hit} , and t_{cov} , Theorem 2.4 improves the dependency on $1/(1 - \lambda_2)$ (or, equivalently t_{mix}), by almost a square-root (we refer the reader to Theorem B.6 and Lemma B.3 of [28] for further details). As a result of this improvement, we get a bound of $O(n/\Phi)$ on the hitting time which is the best known bound on the hitting time (and cover time) in terms of the conductance and improves the bound of [1, Corollary 6.2.1] by a factor of $1/\Phi$.

We also derive a general lower bound on t_{meet} that combines the trivial bound, $1/\|\pi\|_2^2$, with the minimum number of collisions (see Theorem B.1(iii) of [28]). Although this bound does not directly yield the correct lower bound for binary trees, it forms the basis of a later analysis in Theorem C.6 of [28].

Finally, we also provide asymptotically tight worst-case bounds on t_{meet} and t_{coal} . We show that on any graph the coalescence time must be at least $\Omega(\log n)$ and is no more than $O(n^3)$. For regular (and in particular vertex-transitive) graphs these bounds become $\Omega(n)$ and $O(n^2)$ (See also Table 2 on page 51 of [28], which also contains an explanation why these bounds are asymptotically tight.) These two new upper bounds for general and regular graphs complete the picture of worst-case bounds:

THEOREM 2.5. *The following hold for graphs of the stated kind.*

- (i) *For any graph G we have $t_{\text{meet}} \in [\Omega(1), O(n^3)]$ and $t_{\text{coal}} \in [\Omega(\log n), O(n^3)]$.*

⁷In Proposition B.9 of [28], we prove this formally by following the proof for the continuous setting [1, Proposition 14.5].

⁸Alternatively, the same bound as the known bound can also be derived from the bound on the conductance in [9] together with Cheeger's inequality.

- (ii) For any regular graph G we have $t_{\text{meet}}, t_{\text{coal}} \in [\Omega(n), O(n^2)]$.

The proof of Theorem 2.5 appears in Section 5 of [28].

2.1 Summary of Technical Contributions: Our work also makes several technical contributions, which might be of interest for future research on coalescing walks and other stochastic processes; these are explained in greater detail in Section 3. Below we give a very brief summary.

- **Conditional Expectation Approach.** Most of our results make use of the conditional expectation approach given in (3.1), a very simple yet extremely powerful tool, which to the best of our knowledge has not been used in the context of meeting and coalescing times before.
- **Division of Particles into two Groups.** One basic ingredient in our proof is a domination result that allows us to divide random walks into a group of “destroyers” (\mathcal{G}_1), which are particles that cannot be eliminated, and a group of remaining particles (\mathcal{G}_2), which can be eliminated by any other random walk. This domination result might be helpful to analyze other stochastic processes involving different types of particles, e.g. [12].
- **New Concentration Inequalities.** We derive a new concentration inequality for random walks on graphs in Section 4.1 of [28]. Unlike previous approaches which are based on the mixing time (or the closely related spectral gap), our new inequality depends only on the hitting time and improves on the existing bounds when the mixing time is close to the hitting time. These tighter inequalities are required to derive worst-case upper bounds on the coalescence time.

3 Proof Ideas and Technical Contributions

When dealing with processes involving concurrent random walks, a significant challenge is to understand the behavior of “short” random walks. This challenge appears in several settings, *e.g.*, in the context of cover time of multiple random walks [4, 23], where Efremenko and Reingold [23, Section 6] highlight the difficulty in analyzing the hitting time distribution before its expectation. In the context of concentration inequalities for Markov chains, Lezaud [29, p. 863] points out the requirement to spend at least mixing time steps before taking any samples. Related to that, in property testing, dealing with graphs that are far from expanders has

been mentioned as one of the major challenges to test the expansion of the graph by Czumaj and Sohler [21].

In our setting, we also face these generic problems and devise different methods to get a handle on the meeting time distribution before its expectation. Despite our focus being on coalescing and meeting times, several of our approaches can be leveraged to derive new bounds on other random walk quantities such as hitting times or cover times (see Appendix B of [28]).

3.1 Bounds on t_{coal} in terms of t_{mix} and t_{meet} :

The key ingredient in the proof of Theorem 2.1, where we express t_{coal} as a tradeoff between t_{meet} and t_{mix} is a better understanding of meeting events prior to the meeting time. More precisely, we derive a tight bound on the probability p_ℓ that two random walks meet before ℓ time-steps, for ℓ in the range $[t_{\text{mix}}, t_{\text{meet}}]$. Arguing about meeting probabilities of walks that are much shorter than t_{meet} allows us to understand the rate at which the number of *alive* random walks is decreasing.

Optimistically, one may hope that starting with k random walks, as there are $\binom{k}{2}$ possible meeting events, roughly $\binom{k}{2} \cdot p_\ell$ meetings may have occurred after ℓ time-steps. However, the non-independence of these events turns out to be a serious issue and we require a significantly more sophisticated approach to account for the dependencies. We divide the k random walks into disjoint groups \mathcal{G}_1 and \mathcal{G}_2 (with $|\mathcal{G}_1|$ usually being much smaller than $|\mathcal{G}_2|$) and walks of \mathcal{G}_1 can’t be eliminated. The domination of the real process by the group-restricted one is established by introducing a formal concept called immortal process at the beginning of Section 3.1 of [28]. In this stochastic process, we can expose the random walks of \mathcal{G}_1 first and consider meetings with random walks in \mathcal{G}_2 (for an illustration, see Figure 2 on page 11 of [28]). Conditioning on a specific exposed walk in \mathcal{G}_1 , the events of the different walks in \mathcal{G}_2 meeting this exposed walk are indeed independent. In fact, we will also use the symmetric case where the roles of \mathcal{G}_1 and \mathcal{G}_2 are switched. Thus, the problem then reduces to calculating the probability of a random walk in \mathcal{G}_2 having a ‘good trajectory’, *i.e.*, one which many random walks in \mathcal{G}_1 would meet with large enough probability.

Surprisingly, it suffices to divide trajectories into only two categories. Although, one may expect that a more fine-grained classification of trajectories would result in better bounds, this turns out not to be the case. In fact, the bound that we derive on the coalescing time in Theorem 2.1 is tight, and this is precisely due to the tightness of Lemma 3.3 of [28]. The tightness is established by the following construction (cf. Figure 1). The graph is designed such that the vast majority of

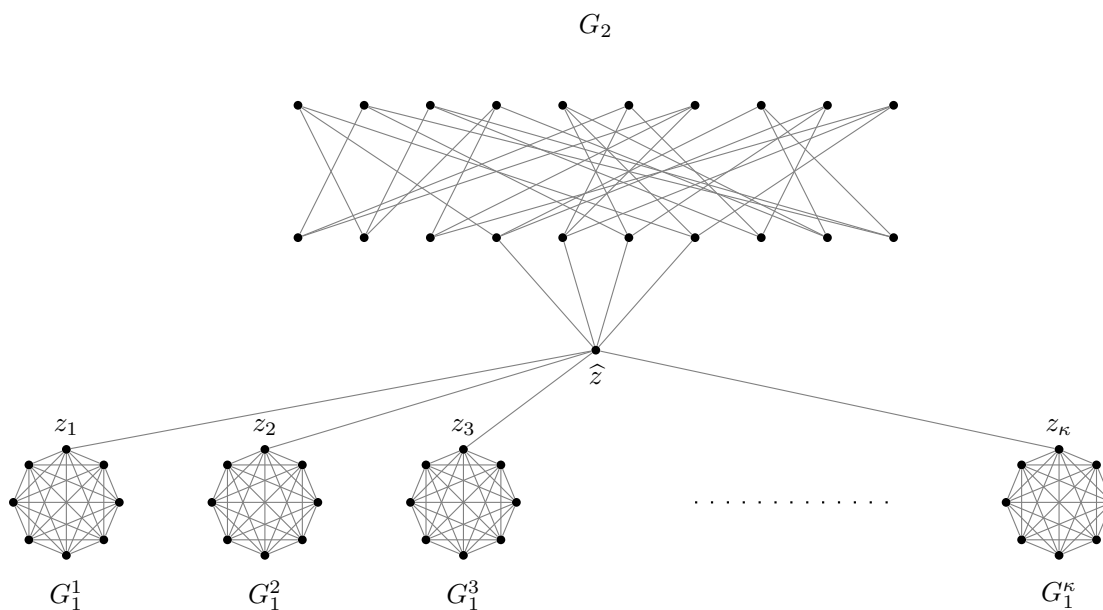


Figure 1: The graph that exhibits $t_{\text{coal}} = \Omega(t_{\text{meet}} + \sqrt{t_{\text{meet}}/t_{\text{mix}}} \cdot \log n \cdot t_{\text{mix}})$.

meetings (between any two random walks) occur in a relatively small part of the graph (G_2 in Figure 1). On average, it takes a considerable number of time-steps before random walks actually get to this part of the graph. What this implies is that for relatively short trajectories (of length significantly smaller than t_{meet}), it is quite likely that other random walks will not meet them. There is a bit of a dichotomy here, once a walk reaches G_2 it is likely that many random walks will meet it; however, a random walk not reaching G_2 is unlikely to be met by any other random walk.

Equipped with Theorem 2.1, we can bound $t_{\text{coal}} = \Theta(t_{\text{meet}})$ for all graphs satisfying $t_{\text{meet}}/t_{\text{mix}} \geq \log^2 n$. Therefore, the problem of bounding t_{coal} reduces to bounding t_{meet} .

For some of the other results including Theorem 2.2 and Theorem 2.3, we will need a more fine-grained approach to derive lower (or upper bounds) on the probability that two walks meet during a certain number of steps, which may or may not be smaller than the mixing time or meeting time. The starting point is the following simple observation. If we have two random walks $(X_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$, and count the number of collisions $Z := \sum_{t=0}^{\tau-1} \mathbf{1}_{X_t=Y_t}$ before time-step $\tau \in \mathbb{N}$, then

$$(3.1) \quad \mathbb{P}[Z \geq 1] = \frac{\mathbb{E}[Z]}{\mathbb{E}[Z \mid Z \geq 1]}.$$

If we further assume that both walks start from the

stationary distribution, then we have

$$\mathbb{P}[Z \geq 1] = \frac{\tau \cdot \|\pi\|_2^2}{\mathbb{E}[Z \mid Z \geq 1]}.$$

To the best of our knowledge, this is the first application of this formula to meeting (and coalescence) times. However, we should mention that variants of this formula have been used by Cooper and Frieze in several works (e.g., [16]) to derive accurate bounds on the hitting (and cover time) on various classes of random graphs, and in Barlow et al. [5] to bound the collisions of random walks on infinite graphs. Using (3.1), we are able to obtain several improvements to existing bounds on the meeting time, and as a consequence for coalescing time. We believe that our work further highlights the power of this basic identity.

The crux of (3.1) is that in order to lower (or upper) bound the probability that the two walks meet, we need to derive a corresponding bound on $\mathbb{E}[Z \mid Z \geq 1]$, i.e., the number of collisions conditioning on the occurrence of at least one collision. Our results employ various tools to get a handle on this quantity, but here we mention one that is quite intuitive:

$$(3.2) \quad \mathbb{E}[Z \mid Z \geq 1] \leq \max_{u \in V} \sum_{t=0}^{\tau-1} \sum_{v \in V} (p_{u,v}^t)^2.$$

The inner summand $\sum_{v \in V} (p_{u,v}^t)^2$ is the probability that two walks starting from the same vertex u will meet after a further t steps. Thus, summing over t and

conditioning on the first meeting happening (*i.e.*, the condition $Z \geq 1$) at some vertex u before time-step τ yields the bound in (3.2). Despite the seemingly crude nature of this bound, it can be used to derive new results for t_{hit} , t_{meet} and t_{coal} that significantly improve over the state-of-the-art for regular graphs (see Appendix B of [28], or the last paragraph in this section for a summary).

3.2 Bounds on t_{coal} in terms of t_{hit} : The derivation of our bounds on t_{coal} in terms of t_{hit} (Theorem 2.3) are based on two general reduction results, that might be useful in other applications:

THEOREM 3.1. (REDUCTION RESULTS) *The following results hold for any graph G :*

1. *The coalescence process reduces the number of walks from n to $O(\log^3 n)$ in $O(t_{\text{hit}})$ steps with probability at least $1 - n^{-1}$. (see Theorem 4.3 of [28])*
2. *The coalescence process reduces the number walks from $\log^4 n$ to $(\Delta/d)^{O(1)}$ in $O(t_{\text{hit}})$ steps in expectation, where Δ is the maximum degree and d is the average degree (see Theorem 4.4 of [28])*

A basic ingredient are new concentration inequalities, which are derived in Section 4.1 of [28]. Our concentration inequalities yield sufficiently strong bounds for upper tails of returns (or other, possibly more complex random variables) by a random walk of length t_{hit} , while most of the existing bounds (*e.g.*, [11, 29]) require that the expectation of the random variable is at least as large as t_{mix} . While $t_{\text{mix}} \leq t_{\text{hit}}$ in general, the challenging case in our analysis is when $t_{\text{mix}} \approx t_{\text{hit}}$ and in this cases our concentration inequalities provide stronger upper tails than the existing ones.

Equipped with these concentration results, the proof of Theorem 3.1 (Part 1) is surprisingly simple and rests again on (3.1). First, by a straightforward bucketing argument on the degree distribution, we show that with high probability, we can find for each random walk $(X_t)_{t \geq 0}$ with label i a set S (depending on the trajectory of X_t), so that with high probability, (i) each vertex in S is visited frequently during $O(t_{\text{hit}})$ steps, and (ii) each vertex in S has the same degree up to constant factors. Conditioning on this, it follows that a second random walk $(Y_t)_{t \geq 0}$ will have sufficient collisions with $(X_t)_{t \geq 0}$ in expectation, *i.e.*, $\mathbb{E}[Z]$ is large enough. To bound $\mathbb{E}[Z \mid Z \geq 1]$, we use the concentration inequalities to establish that with high probability, the trajectory $(X_t)_{t \geq 0}$ will be good in the sense that $\mathbb{E}[Z \mid (x_0, x_1, \dots), Z \geq 1]$ is not too large. Combining these bounds yields $\mathbb{P}[Z \geq 1] = \Omega(1/\log^3 n)$, and a straightforward division into groups \mathcal{G}_1 and \mathcal{G}_2 of sizes

$\Theta(\log^3 n)$ and $n - |\mathcal{G}_1|$ shows that all random walks in \mathcal{G}_2 can be eliminated in $O(t_{\text{hit}})$ steps.

The proof of the second reduction result (Theorem 3.1 (Part 2)) is more involved, although it again revolves around (3.1). The issue is that we can no longer repeat the simple bucketing argument described above about the degree distribution, since the number of buckets may vastly exceed the number of walks. Furthermore, we may no longer obtain “w.h.p.”-bounds on the probability for certain good events. For all these reasons, a refined approach is needed.

Our analysis allocates small phases of length $O(t_{\text{hit}}/\kappa)$ in order to halve the number of random walks, where $k = \kappa^c$ is the number of walks at the beginning of the phase, for some suitably large constant c . The first step is to show that starting from any vertex, there exists a large set of vertices, so that each vertex is visited the “right” amount of time, but also that it was not too unexpected to visit that vertex. The latter condition is quite subtle, but it allows us to arrange a proper scheduling of the walks to show that, regardless of which vertices the random walk i decides to visit in that set, there are enough walks that are able to reach these vertices by then. In other words, it rules out the possibility that, despite two random walks visiting the same set of vertices, they never collide (for an illustration, see Figure 5 on page 45 of [28]). Using our concentration bounds with a careful choice of the slackness parameters in terms of κ , the above approach can eventually be shown to reduce the number of random walks k by a constant fraction within $O(t_{\text{hit}}/\kappa)$ steps. Repeating this iteratively yields the bound $O(t_{\text{hit}})$.

3.3 Bounds on t_{hit} and Worst-Case Bounds :

With the two reduction results, Theorem 2.3 follows immediately. Furthermore, the aforementioned results can be also used to derive worst-case upper and lower bounds on meeting and coalescing time on general and regular graphs that are tight up to constant factors. Some of these were known, or follow directly from existing results, the others are novel to the best of our knowledge.

We proceed by establishing that $t_{\text{coal}} = O(n^3)$ on all graphs. The proof of $t_{\text{coal}} = O(n^3)$ (Theorem 2.5) follows by first applying both reductions (Theorem 3.1 (Parts 1 & 2)) to reduce the number of walks from n to $(\Delta/d)^{O(1)} \leq (n^2/|E|)^{O(1)}$ in $O(t_{\text{hit}})$. We have, by Proposition B.9 of [28], $t_{\text{meet}} \leq 4t_{\text{hit}} = O(n \cdot |E|)$, where this last bound follows from [3].

Finally, combining the bound $t_{\text{meet}} = O(n \cdot |E|)$ together with $t_{\text{coal}}(S_0) = O(t_{\text{meet}} \cdot \log(|S_0|))$ (Proposition 3.4 of [28]) for any set of start vertices S_0 , yields

that after additional

$$O(t_{\text{meet}} \cdot \log(|S_0|)) = O(n \cdot |E| \cdot \log(n^2/|E|)) = O(n^3)$$

steps the coalescing terminates. The fact that this is tight can be easily verified by considering the Barbell graph.⁹

For regular graphs, the same argument as before shows that $t_{\text{coal}} = O(n^2)$, and this is matched by the cycle, for instance. The proofs of the other results are straightforward, and we refer the reader to Section 5.3 of [28].

3.4 Bounds on t_{meet} and Other Results: In Appendix B of [28], we derive several bounds on t_{meet} . These bounds are derived more directly by (3.1) and/or (3.2), and involve other quantities such as $\|\pi\|_2^2$ or the eigenvalue gap $1 - \lambda_2$. One important technical contribution is to combine routine spectral methods involving the spectral representation and fundamental matrices that have been used in previous works, *e.g.*, Cooper et al. [13] with some short-time bounds on the t -th step probabilities. This allows us to improve several bounds, not only on t_{meet} and t_{coal} but also t_{hit} and t_{cov} , by significantly reducing the dependency on the spectral gap or mixing time—by almost a square root factor. As a corollary, we also derive a new bound on the cover time for regular graphs that considerably improves over the best known bound by Broder and Karlin [10] from 1989.

3.5 Concrete Topologies: Finally, in Appendix C of [28], we apply the derived upper and lower bounds on t_{meet} and t_{coal} on various fundamental topologies including grids, expanders and hypercubes. In most cases, these results follow immediately from the general bounds by plugging in corresponding values for $\|\pi\|_2^2$, t_{hit} or t_{mix} . One exception is the binary tree, for which it seems surprisingly non-trivial to derive a lower bound of $t_{\text{meet}} = \Omega(n \log n)$. Here again we use a refinement of (3.1) that restricts the vertices to leaf-nodes u , for which $\sum_{t=1}^{t_{\text{mix}}} (\sum_{v \in V} p_{u,v}^t)^2 = \Omega(\log n)$. The matching upper bound $t_{\text{meet}} = O(n \log n)$ follows from $t_{\text{coal}} = O(t_{\text{hit}})$ for almost-regular graphs (Theorem 2.3).

Of particular interest might be the analysis of “real-world” graph models. There we show how to utilize our bounds from earlier sections to establish $t_{\text{coal}} = \Theta(t_{\text{meet}})$ on two random graph models, leading to bounds on t_{coal} that are sublinear in the number of vertices.

3.6 Discussion and Future Work In this work we derived several novel bounds on t_{coal} . Our first main

result implies that a gap of just $\Omega(\log^2 n)$ between t_{mix} and t_{meet} is sufficient to have $t_{\text{coal}} = \Theta(t_{\text{meet}})$. We also proved that this result is essentially tight. Further, we derived several new bounds on t_{coal} based on t_{hit} . For almost-regular-graphs, our new result implies the following hierarchy for the discrete-time setting,

$$t_{\text{meet}} \leq t_{\text{coal}} = O(t_{\text{hit}}),$$

which refines the already known result $t_{\text{meet}} = O(t_{\text{hit}})$. Finally, we also determined tight worst-case lower and upper bound for t_{coal} .

For future work, an obvious problem is to extend the $t_{\text{coal}} = O(t_{\text{hit}})$ result to all graphs (so far, we only know $t_{\text{coal}} = O(t_{\text{hit}} \cdot \log \log n)$). Even more ambitious would be to try to prove that the continuous-time variant and the discrete-time process are (asymptotically) equivalent, as this would immediately resolve the $t_{\text{coal}} = O(t_{\text{hit}})$ problem. A different direction may be to further explore lower bounds on t_{meet} .

Acknowledgments. The authors would like to thank Petra Berenbrink, Robert Elsässer, Nikolaos Fountoulakis, Peter Kling, Roberto Oliveira and Perla Sousi for helpful discussions and in particular Yuval Peres for pointing out how to further improve the bound on the hitting time (Theorem 2.4). Moreover, the authors would like to thank the anonymous reviewer for pointing out a mistake in the domination described in Section 3.1 of [28] in an earlier version of that work.

References

- [1] D. Aldous and J. Fill. *Reversible Markov Chains and Random Walks on Graphs*. Unpublished. <http://www.stat.berkeley.edu/~aldous/RWG/book.html>. 2002 (cit. on pp. 1, 2, 4).
- [2] D. Aldous. “Meeting times for independent Markov chains”. In: *Stochastic Processes and their Applications* 38.2 (1991), pp. 185–193 (cit. on p. 2).
- [3] R. Aleliunas, R. Karp, R. Lipton, L. Lovász, and C. Rackoff. “Random Walks, Universal Traversal Sequences, and the Complexity of Maze Problems”. In: *FOCS*. 1979, pp. 218–223 (cit. on p. 7).
- [4] N. Alon, C. Avin, M. Koucký, G. Kozma, Z. Lotker, and M. R. Tuttle. “Many Random Walks Are Faster Than One”. In: *Combinatorics, Probability & Computing* (2011) (cit. on p. 5).
- [5] M. Barlow, Y. Peres, and P. Sousi. “Collisions of Random Walks”. In: *Annales de l’Institut Henri Poincaré* 48.4 (2012), pp. 922–946 (cit. on p. 6).

⁹This n -vertex graph is constructed by taking two cliques of size $n/4$ each, and connecting them through a path of length $n/2$.

- [6] L. Becchetti, A. E. F. Clementi, E. Natale, F. Pasquale, and L. Trevisan. “Stabilizing Consensus with Many Opinions”. In: *SODA*. 2016, pp. 620–635 (cit. on p. 2).
- [7] L. Becchetti, A. Clementi, E. Natale, F. Pasquale, and R. Silvestri. “Plurality Consensus in the Gossip Model”. In: *SODA*. 2015, pp. 371–390 (cit. on p. 2).
- [8] P. Berenbrink, A. E. F. Clementi, R. Elsässer, P. Kling, F. Mallmann-Trenn, and E. Natale. “Ignore or Comply?: On Breaking Symmetry in Consensus”. In: *PODC*. 2017, pp. 335–344. URL: <http://doi.acm.org/10.1145/3087801.3087817> (cit. on p. 2).
- [9] P. Berenbrink, G. Giakkoupis, A.-M. Kermarrec, and F. Mallmann-Trenn. “Bounds on the Voter Model in Dynamic Networks”. In: *ICALP*. 2016 (cit. on pp. 1, 2, 4).
- [10] A. Z. Broder and A. R. Karlin. “Bounds on the cover time”. In: *Journal of Theoretical Probability* 2.1 (1989), pp. 101–120 (cit. on pp. 1, 4, 8).
- [11] K.-M. Chung, H. Lam, Z. Liu, and M. Mitzenmacher. “Chernoff-Hoeffding Bounds for Markov Chains: Generalized and Simplified”. In: *STACS*. 2012, pp. 124–135 (cit. on p. 7).
- [12] C. Cooper, A. Frieze, and T. Radzik. “Multiple Random Walks in Random Regular Graphs”. In: *SIAM J. Discret. Math.* 23.4 (2009), pp. 1738–1761 (cit. on pp. 1, 4, 5).
- [13] C. Cooper, R. Elsässer, H. Ono, and T. Radzik. “Coalescing Random Walks and Voting on Connected Graphs”. In: *SIAM Journal on Discrete Mathematics* 27.4 (2013), pp. 1748–1758 (cit. on pp. 1–4, 8).
- [14] C. Cooper, R. Elsässer, and T. Radzik. “The Power of Two Choices in Distributed Voting”. In: *Proc. ICALP*. 2014, pp. 435–446 (cit. on p. 2).
- [15] C. Cooper, R. Elsässer, T. Radzik, N. Rivera, and T. Shiraga. “Fast Consensus for Voting on General Expander Graphs”. In: *DISC*. 2015, pp. 248–262 (cit. on p. 2).
- [16] C. Cooper and A. M. Frieze. “The Cover Time of Random Regular Graphs”. In: *SIAM J. Discrete Math.* 18.4 (2005), pp. 728–740 (cit. on p. 6).
- [17] C. Cooper, A. Frieze, and T. Radzik. “Multiple Random Walks and Interacting Particle Systems”. In: *Automata, Languages and Programming*. Ed. by S. Albers, A. Marchetti-Spaccamela, Y. Matias, S. Nikolettseas, and W. Thomas. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 399–410 (cit. on p. 1).
- [18] C. Cooper, T. Radzik, N. Rivera, and T. Shiraga. “Fast plurality consensus in regular expanders”. In: *CoRR* abs/1605.08403 (2016). URL: <http://arxiv.org/abs/1605.08403> (cit. on p. 2).
- [19] C. Cooper and N. Rivera. “The Linear Voting Model”. In: *ICALP*. Vol. 55. 2016, 144:1–144:12 (cit. on p. 2).
- [20] J. Cox. “Coalescing Random Walks and Voter Model Consensus Times on the Torus in \mathbb{Z}^d ”. In: *The Annals of Probability* 17.4 (1989), pp. 1333–1366 (cit. on p. 2).
- [21] A. Czumaj and C. Sohler. “Testing Expansion in Bounded-Degree Graphs”. In: *Combinatorics, Probability & Computing* 19.5-6 (2010), pp. 693–709 (cit. on p. 5).
- [22] B. Doerr, L. A. Goldberg, L. Minder, T. Sauerwald, and C. Scheideler. “Stabilizing Consensus With the Power of Two Choices”. In: *Proc. SPAA*. 2011, pp. 149–158 (cit. on p. 1).
- [23] K. Efremenko and O. Reingold. “How Well Do Random Walks Parallelize?” In: *RANDOM*. 2009, pp. 476–489 (cit. on p. 5).
- [24] R. Elsässer, T. Friedetzky, D. Kaaser, F. Mallmann-Trenn, and H. Trinker. “Efficient k-Party Voting with Two Choices”. In: *CoRR* abs/1602.04667 (2016). URL: <http://arxiv.org/abs/1602.04667> (cit. on p. 2).
- [25] M. Gauci, M. E. Ortiz, M. Rubenstein, and R. Nagpal. “Error Cascades in Collective Behavior: A Case Study of the Gradient Algorithm on 1000 Physical Agents”. In: *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2017, São Paulo, Brazil, May 8-12, 2017*. 2017, pp. 1404–1412. URL: <http://dl.acm.org/citation.cfm?id=3091319> (cit. on p. 1).
- [26] M. Ghaffari and J. Lengler. “Tight Analysis for the 3-Majority Consensus Dynamics”. In: *CoRR* abs/1705.05583 (2017). URL: <http://arxiv.org/abs/1705.05583> (cit. on p. 2).
- [27] Y. Hassin and D. Peleg. “Distributed Probabilistic Polling and Applications to Proportionate Agreement”. In: *Information and Computation* 171.2 (2001), pp. 248–268 (cit. on pp. 2, 4).

- [28] V. Kanade, F. Mallmann-Trenn, and T. Sauerwald. “On coalescence time in graphs-When is coalescing as fast as meeting?” In: *CoRR* abs/1611.02460 (2016). The references in the text correspond to v4 on arxiv. arXiv: [1611.02460](https://arxiv.org/abs/1611.02460). URL: <http://arxiv.org/abs/1611.02460> (cit. on pp. [1](#), [3–5](#), [7](#), [8](#)).
- [29] P. Lezaud. “Chernoff-type bound for finite Markov chains”. In: *Ann. Appl. Probab.* 8.3 (1989), pp. 849–867 (cit. on pp. [5](#), [7](#)).
- [30] R. I. Oliveira and Y. Peres. “Random walks on graphs: new bounds on hitting, meeting, coalescing and returning”. In: *arXiv preprint arXiv:1807.06858* (2018) (cit. on p. [4](#)).
- [31] R. I. Oliveira. “On the coalescence time of reversible random walks.” In: *Trans. Am. Math. Soc.* 364.4 (2012), pp. 2109–2128 (cit. on p. [2](#)).
- [32] R. I. Oliveira. “Mean field conditions for coalescing random walks”. In: *Ann. Probab.* 41.5 (2013), pp. 3420–3461 (cit. on pp. [2](#), [3](#)).
- [33] D. Peleg. “Local majorities, coalitions and monopolies in graphs: a review”. In: *Theoretical Computer Science* 282.2 (2002), pp. 231–257 (cit. on p. [1](#)).
- [34] E. Perron, D. Vasudevan, and M. Vojnović. “Using Three States for Binary Consensus on Complete Graphs”. In: *Proc. INFOCOM*. 2009, pp. 2527–2535 (cit. on p. [1](#)).