

The Birthday Paradox for non-backtracking walks on regular graphs

Benjamin Dozier *

August 26, 2026

Abstract

We show a birthday paradox for random non-backtracking walk on regular graphs of degree at least 3: such a walk of length k has high probability of self-intersecting when k is significantly greater than $\sqrt{n}$, where n is the number of vertices of the graph. This resolves a conjecture of Noga Alon and Yuval Peres for the fixed degree case.

1 Introduction

The classical “birthday paradox” in elementary probability states that if we have n objects and draw k of them at random with replacement, uniformly and independently, we are likely to draw at least one object more than once provided that k is significantly greater than $\sqrt{n}$. In this paper we prove a version of this in which the objects are not drawn uniformly at random, but rather are the vertices of a random non-backtracking walk on a (regular) graph.

We consider (undirected) graphs without multiple edges or self-loops. When we speak of an edge, it will always come with an orientation. A *non-backtracking walk* (NBW) on such a graph is a sequence of edges such that (i) the endpoint of each edge (other than the last) is equal to the startpoint of the next, and (ii) no edge is followed by its opposite. Such a walk has a *self-intersection* if it hits some vertex more than once; a walk without self-intersection is called *simple*. We prove:

Theorem 1.1. *Fix $d \geq 3$. For any $\epsilon > 0$, there exists a $C = C(d, \epsilon)$ such that if Γ is a d -regular graph on n vertices, and $k > C\sqrt{n}$, then for any edge δ , the probability that a length k random non-backtracking walk starting with δ is simple is less than ϵ .*

Remark 1.2. For $d = 2$ the statement is false. For instance, if Γ is a cycle of length n , then all NBWs of length $k < n$ are simple.

*Department of Mathematics, Cornell University, benjamin.dozier@cornell.edu. Research supported in part by NSF Grant DMS-2247244 and the Simons Foundation.

Remark 1.3. If we remove the non-backtracking assumption, a much stronger statement is true. In this case, at each step the walk has a definite chance of repeating the previous edge and thus producing a self-intersection. Hence the relevant probability is related to a geometric random variable, and thus C can be chosen (depending on d, ϵ) so that the result holds for any $k > C$.

Remark 1.4. The result is sharp in the sense that if we replace $\sqrt{n}$ in the statement by some function $f(n)$ with $f(n) = o(\sqrt{n})$, then it becomes false. This can be seen by considering (uniformly) *random* d -regular graphs. For closed loops, this is shown in [DS24, Theorem 1.1]; as mentioned there, for walks the same result holds and the argument is considerably easier (using the strategy of “build both the walk and the graph step-by-step”).

Noga Alon and Yuval Peres conjectured the following more general result; our theorem above is the fixed degree case.

Conjecture 1.5 (Alon-Peres). *For any $\epsilon > 0$, there exists a $C = C(\epsilon)$ such that if Γ is a graph on n vertices with all degrees at least 3, and $k > C\sqrt{n}$, then for any edge δ , the probability that a length k random non-backtracking walk starting with δ is simple is less than ϵ .*

One could also explore analogs of these statements for (non-backtracking) *closed loops*, rather than walks.

Sketch of proof. It turns out to be useful to do pre-sampling, i.e. work with walks with additional information, namely data that gives for each edge, a new edge to follow after that one; we call this a plan. The ratio of simple plans to all plans is equal to the analogous ratio for walks. Given a plan whose walk is simple, we produce many slight deviations, many of which will have to self-intersect. In fact, many of the deviating paths, if simple, must return rather quickly to the original path, since otherwise, they would be forced to be disjoint (here is where it is important to consider plans, rather than plain walks). And then of these deviations that eventually return to the original path, an excursion reversal argument shows that at least (approximately) half must self-intersect.

The above produces for each simple plan, many self-intersecting ones. Since the excursions are relatively short, each self-intersecting plan can only arise like this in a few ways. Putting these facts together gives many more self-intersecting plans than simple ones.

Prior work on NBW and self-intersection. Non-backtracking random walk on regular graphs has been studied in [ABLS07, LS10, LP16] among other papers. Yadin has studied simple walks on graphs of large girth [Yad16]. The relative number of simple and self-intersecting closed loops on random regular graphs plays an important role in the study of their spectral gap; see [BS87, Fri91, Fri08]. In [DS24],

the author and Sapir show a birthday paradox for closed loops on random regular graphs.

The birthday paradox in the related context of Markov processes is relevant to the effectiveness of the Pollard rho algorithm for discrete logarithms. Results in this context have been proved in [MV06, KMPT10].

After completing a draft of this paper, the author learned of independent work [BGJ⁺26] on related questions, posted to the arxiv a few weeks before. They give an upper bound on self-intersection time in a more general setting, but need an additional $\log n$ factor in their bound. They also prove a bound without the log factor, but under the assumption that the graphs are fixed degree expanders.

AI use and autoformalization. The case of degree 3 was proved by GPT 5.6 Sol Ultra, in about 2 hours of reasoning time. That proof was digested by the author, who generalized it to all degrees, requiring some additional work. The paper was written entirely by the author, who developed the organization, terminology, notation, exposition, and way that corner cases are handled. But the key ideas of using “plans” (the author’s terminology) to get disjointness of deviations, as well as the excursion reversal argument, were found by the AI. The paper was proofread by the AI, and the author made several minor corrections based on the resulting feedback. The figure was produced by the AI based on a hand-drawn sketch of the author.

Using Codex, the main theorem and proof have been (auto)formalized in Lean 4; see ancillary files. These will also be made available on the **Palomar** registry.

Acknowledgements. I am grateful to Noga Alon for telling me about his conjecture with Peres, and for later pointing out to me the related work [BGJ⁺26]. I thank Ariel Yadin and Jenya Sapir for interesting discussions about the problem.

2 Proof of Theorem 1.1

Let $N(k, \delta)$ denote the number of length k NBWs on Γ starting with δ , and let $N_{\text{simp}}(k, \delta)$ denote the number of those that are simple.

Plans and their associated walks. A key idea in the proof is to consider walks with some additional information about where to go, even at unvisited vertices. Given an edge δ , denote by $\bar{\delta}$ the edge with the reverse orientation. We define a *startless plan* on Γ to be a function ϕ from the set of edges to itself such that (i) for each edge (u, v) , $\phi(u, v)$ equals (v, w) for some vertex $w \neq u$, and (ii) ϕ is injective. The map ϕ should be thought of as determining where a walker goes next given the current edge. Condition (i) corresponds to non-backtracking, while (ii) means that a single edge also determines a full past trajectory. One can alternatively think of a startless plan as specifying for each vertex v a derangement π (permutation without

fixed points) on the d incoming edges; the restriction of ϕ to the edges ending at v is then given by composing π with the orientation reversing involution on edges.

A *plan* ρ is a pair (ϕ, δ) where ϕ is a startless plan and δ is an edge, the *start*. This determines an *associated NBW* $\gamma_k(\rho)$ (of any length k we choose) given by starting with δ and then repeatedly applying ϕ . We also define $\gamma(\rho)$ to be the infinite sequence produced like this. We define a *k-simple plan* to be one for which the associated walk of length k is simple. Let Ω^δ denote the set of plans with start δ , and let $\Omega_{simp}^\delta(k)$ be the set of k -simple plans with start δ .

A crucial fact is that the ratio of the number of k -simple plans to all plans is exactly the probability that we are interested in. The reason is that both can be thought of as encoding a choice of length k walk, together with the same total number of other choices associated with edges that are not involved in the walk (for walks that are self-intersecting, this encoding is less direct). We state this as a lemma, and give a more computational proof:

Lemma 2.1. *For $k < n$:*

$$\frac{|\Omega_{simp}^\delta(k)|}{|\Omega^\delta|} = \frac{N_{simp}(k, \delta)}{N(k, \delta)}. \quad (1)$$

Proof. Let R_d denote the number of derangements (permutations without fixed points) on the set $\{1, \dots, d\}$. Fixing some $1 < j \leq d$, let R'_d be the number of these derangements that map j to 1 (note by symmetry, the number doesn't depend on the choice of j). Since there are $d - 1$ choices of this j , we have $R_d = (d - 1)R'_d$. We compute:

$$\frac{|\Omega_{simp}^\delta(k)|}{|\Omega^\delta|} = \frac{N_{simp}(k, \delta) (R'_d)^{k-1} (R_d)^{n-(k-1)}}{(R_d)^{k-1} (R_d)^{n-(k-1)}} = \frac{N_{simp}(k, \delta)}{(d - 1)^{k-1}} \quad (2)$$

$$= \frac{N_{simp}(k, \delta)}{N(k, \delta)}. \quad (3)$$

■

Deviating from the plan. From here, to establish Theorem 1.1, for each k -simple plan, we will consider many deviations away from it, producing many new walks, many of which will turn out to self-intersect (actually, we will only prove this for the *average* over all k -simple plans with a fixed start). The reason that the extra information of a plan is helpful is that it will force certain deviations to be disjoint.

Our notion of deviation concerns changing a plan at a single vertex. When $d = 3$, there are only two choices for the data of a plan at a particular vertex, and thus there is a unique way to change the plan there. However, when $d > 3$, there are more choices, but what ϕ does to edges not along the associated walk will not be of much concern. Nevertheless, it is useful to pick specific values for these (largely

irrelevant) choices, which motivates the following definition; this will come up again in (Biii) of proof of Lemma 2.2. Suppose we have chosen a vertex v , a derangement π of the incoming edges at v , and a pair (δ, β) , where δ and β are incoming and outgoing edges at v , respectively, with $\delta \neq \bar{\beta}$. We pick arbitrarily a *distinguished derangement contingency*, which is a derangement π' of the incoming edges of v such that $\pi'(\delta) = \beta$. (This choice will actually only affect parts of the plan that are not relevant to the associated walk, which is why the word “contingency” is used.)

Given a k -simple plan $\rho = (\phi, \delta)$, consider the edges $\delta = \delta_1, \delta_2, \dots$ of the associated walk $\gamma(\rho)$. We define a *deviation* of ρ to be a pair (ρ', ρ) , where the plan $\rho' = (\phi', \delta)$ is obtained from ϕ by changing the derangement at the endpoint of δ_i for some $i = 1, \dots, k-1$ to the distinguished derangement contingency corresponding to (δ_i, β) , for some $\beta \neq \delta_{i+1}$. Note that there are $(k-1)(d-2)$ deviations of a k -simple plan.

We say a deviation (ρ', ρ) obtained from changing $\phi(\delta_i)$ is *L-simple* if its associated length $i + L$ walk is simple. We want to choose L large enough that many deviations are not *L-simple*.

Lemma 2.2. *There exists C' (depending on d) such that if $L > C'n/k$ and δ is an edge of Γ , then at least $1/4$ proportion of the deviations of k -simple plans starting at δ are not *L-simple*.¹*

Note that deviations of two different plans could result in the same plan, but these are regarded as distinct deviations, since a deviation was defined as a pair (ρ', ρ) , i.e. it remembers the plan it deviated from.

Proof. There are two types of deviations of a k -simple plan ρ , and we will handle them with two different methods, a packing argument and an “excursion reversal” argument:

- (A) *Non-returning* deviations ρ' of ρ , i.e. those for which the associated walk does not return to a vertex of $\gamma_k(\rho)$ within L steps of deviating. We will bound these with a packing argument using disjointness.

Consider non-returning deviations that are *L-simple*, which in particular implies that the L steps after deviating are on distinct edges. Note that a plan determines the walk once a single edge along it is known (here we use that the function ϕ in the definition of plan is assumed to be injective). It follows that any two non-returning deviations ρ', ρ'' of ρ must have the pieces consisting of the last L edges of their associated walks disjoint. Note that there are nd edges of Γ , so we can have at most nd/L such disjoint pieces. So for $L > C'n/k$, we get that the number of these deviations is at most $\frac{nd}{C'n/k} = (d/C')k$, and choosing $C' > 8d$, we get that there are fewer than $k/8$ such deviations of ρ . In

¹In this lemma, it is important that $d \geq 3$; when $d = 2$, there are no deviations, and so the proportion is not meaningful.

total there are $(k-1)(d-2)$ deviations of ρ , so by taking C' a bit larger, we find that the proportion of deviations of ρ that are non-returning and L -simple is less than $1/8$. And so, a fortiori, considering *all* possible k -simple ρ with start δ , we see that the proportion of deviations that are non-returning and L -simple is again at most $1/8$.

(B) *Returning* deviations ρ' , i.e. those for which $\gamma(\rho')$ returns to a vertex w of $\gamma_k(\rho)$ within L steps of deviating (and we take w to be the first such vertex along $\gamma(\rho')$). We will separate into three cases based on w ; the first two are unusual corner cases that can be dealt with quickly.

- (i) Suppose that w is the startpoint of the start edge δ . Such a deviation is clearly not L -simple, and so no argument is needed here.
- (ii) Suppose that w is the endpoint of δ_k . Note that for fixed plan ρ , such a ρ' is determined by the edge through which it comes back to w . Hence there are at most $d-1$ such deviations. The total number of deviations of ρ is $(k-1)(d-2)$, so this accounts for at most $2/(k-1)$ proportion of deviations. For $k > 17$, this proportion is less than $1/8$. For $k \leq 17$, note that we can take $C' > 17$ and so $L > C'n/k \geq n$, and thus there are no L -simple deviations, since there are only n vertices in our graph; in this case the proportion of such deviations that are also L -simple is zero.
- (iii) Suppose ρ' is *middle-returning*, i.e. returning but not of type (Bi) or (Bii). To handle these we define an “excursion reversing” involution R on the set of all middle-returning deviations of k -simple plans starting at δ . This map will have the property that for each (ρ', ρ) , it cannot be that both (ρ', ρ) , $R(\rho', \rho)$ are L -simple. The idea is that the plan is reversed along the edges of the “excursion”. See Figure 1.

Suppose that $\rho = (\phi, \delta)$, and that $\rho' = (\phi', \delta)$ deviates at index i . The *excursion* is the sequence of edges $e_1 = \phi'(\delta_i), \dots, e_j := (\phi')^{\circ j}(\delta_i)$, where e_j has endpoint w on $\gamma_k(\rho)$, but no other e_η with $\eta < j$ has endpoint on $\gamma_k(\rho)$. By the returning assumption, $j \leq L$. We take $R(\rho', \rho)$ to be $(\tilde{\rho}', \tilde{\rho})$, defined as follows. For each vertex v along the excursion (but not on $\gamma_k(\rho)$), we modify ρ by taking the derangement π at v and replacing it with the inverse derangement. The plan is not changed away from the excursion. This gives us $\tilde{\rho}$. To get $\tilde{\rho}'$, we modify $\tilde{\rho}$ only by changing the plan for w . Since we are excluding deviations of type (Bii), w is an allowed location to deviate. And since we are excluding type (Bi), there is a (unique) edge δ_ℓ of $\gamma_k(\rho)$ that has w as its endpoint. We define $\tilde{\phi}'$ at w so that $\tilde{\phi}'(\delta_\ell) = \bar{e}_j$, using the distinguished derangement contingency that agrees with this.

These definitions mean that $\gamma(\rho')$ and $\gamma(\tilde{\rho}')$ both traverse the same excursion edges, but in opposite directions.

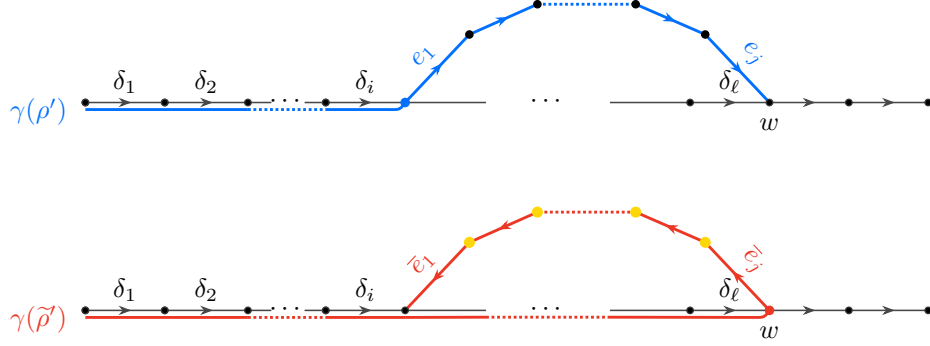


Figure 1: The mechanism by which at least half of middle-returning walks are self-intersecting. The top figure shows the walk associated to simple plan ρ (black edges), and a deviation ρ' at the i th index (in blue). The bottom figure represents the excursion reversal $R(\rho', \rho) = (\tilde{\rho}', \tilde{\rho})$. The plan $\tilde{\rho}$ is obtained by “reversing” along the excursion, i.e. the plan is changed at each yellow vertex by inverting the derangement there. Then deviation at w produces $\tilde{\rho}'$, whose associated walk is drawn in red. In this diagram $\ell > i$, and $\gamma_k(\rho')$ is simple, but then $\gamma_k(\tilde{\rho}')$ cannot be – it self-intersects at the endpoint of δ_i .

If $\ell \leq i$, i.e. the walk $\gamma(\rho')$ returns to $\gamma(\rho)$ at a vertex traversed no later than the excursion, then ρ' is clearly not L -simple. But otherwise $\tilde{\rho}'$ is not L -simple. Thus R has the properties claimed above.

Combining the three cases above, we get that among all deviations of k -simple plans starting with δ , at most $1/8 + 1/2 = 5/8$ are returning and L -simple.

Now any deviation is either non-returning or returning. Among all deviations of k -simple plans starting with δ , non-returning L -simple constitute at most $1/8$ by (A), while returning L -simple constitute at most $5/8$ by (B). Thus the total proportion of deviations that are L -simple is at most $5/8 + 1/8 = 3/4$, as desired. ■

Lemma 2.3. *For any $\epsilon > 0$, there exists such C such that for $k^2 > Cn$ we have*

$$\frac{|\Omega_{simp}^\delta(k)|}{|\Omega^\delta|} < \epsilon. \quad (4)$$

Proof. Choose $\eta > 0$ small (see later in proof) and take L an integer such that

$$\eta k > L > C'n/k, \quad (5)$$

where C' is the constant in Lemma 2.2 ; there is enough gap between the left and right hand sides above to do this if we take C a little larger than C'/η .

We will define a map D that associates to each plan $\rho \in \Omega_{simp}^\delta(k)$ a subset of Ω^δ , which will consist of some of the plans formed from deviations of ρ . Specifically, we consider deviations (ρ', ρ) , where ρ' deviates at some step $1, \dots, k-L$, and define $D(\rho)$ to be the set of such ρ' that are not L -simple; the condition on the deviation step implies that $\gamma_k(\rho')$ is not simple. We will use D to bound the ratio in (4).

First, we want to bound from below the average size of $D(\rho)$, over all $\rho \in \Omega_{simp}^\delta(k)$. By Lemma 2.2 at least $1/4$ proportion of deviations are not L -simple, but that includes those deviating at any step, while for D we imposed a restriction. However, the proportion of deviations excluded by this restriction is at most $L/(k-1)$, and since $\eta k > L$, even with the imposed restriction we still get at least proportion $1/4 - 2\eta$ of deviations are allowed, and provided $\eta < 1/16$, this is at least $1/8$. Thus the average size of $D(\rho)$ is at least $(k-1)/8 \geq k/16$.

Second, we want to control the number of distinct plans that admit a common deviation, i.e. we want for each plan $\rho' \in \Omega^\delta$, a bound on the number of distinct plans ρ for which $\rho' \in D(\rho)$. To this end, note that if this number is at least one, then ρ' must arise from a deviation of a k -simple plan, while $\gamma_k(\rho')$ is not simple. Consider the smallest j such that $\gamma_j(\rho')$ is not simple. We claim that for any ρ with $\rho' \in D(\rho)$ we must have that ρ' deviates from ρ at $i \in [j-L, j]$. In fact, if $i > j$, then $\gamma_j(\rho), \gamma_j(\rho')$ would coincide, but the former is simple, while the latter is not. On the other hand, if $i < j-L$, then, since ρ' is not L -simple, its associated walk $\gamma(\rho')$ would have a self-intersection before step $j-L+L = j$, again a contradiction. From this claim that we've established, it follows that the number of possible ρ is at most $R_d \cdot (L+1)$, since ρ, ρ' must agree except for the derangement at the single vertex (of which there are R_d choices), and there are at most $j - (j-L) + 1 = L+1$ choices of this vertex.

Now combining the results about the map D from the previous two paragraphs, we get that

$$\begin{aligned} |\Omega^\delta| &\geq \frac{(k/16) \cdot |\Omega_{simp}^\delta(k)|}{R_d \cdot (L+1)} \\ &\geq \frac{(L/\eta)/16 \cdot |\Omega_{simp}^\delta(k)|}{R_d \cdot (L+1)} \\ &\geq \frac{|\Omega_{simp}^\delta(k)|}{32\eta R_d}, \end{aligned}$$

and so taking η such that $32\eta R_d < \epsilon$ gives the desired result. ■

Proof of Theorem 1.1. For $k \geq n$ the statement is trivial, since there are no simple walks of this length. Now assuming $k < n$, we can apply Lemma 2.1 and then Lemma 2.3 (the C in the theorem can be taken to be the square root of the C in Lemma 2.3). ■

References

- [ABLS07] Noga Alon, Itai Benjamini, Eyal Lubetzky, and Sasha Sodin, *Non-backtracking random walks mix faster*, Commun. Contemp. Math. **9** (2007), no. 4, 585–603. MR 2348845 2
- [BGJ⁺26] Ferenc Bencs, Leslie Ann Goldberg, Matthew Jenssen, Mark Jerrum, Gabor Pete, Guus Regts, and Yitong Yin, *On the self-intersection time of non-backtracking random walks*, arXiv e-prints (2026), arXiv:2608.09729. 3
- [BS87] Andrei Broder and Eli Shamir, *On the second eigenvalue of random regular graphs*, 28th Annual Symposium on Foundations of Computer Science (sfcs 1987), 1987, pp. 286–294. 2
- [DS24] Benjamin Dozier and Jenya Sapir, *Simple versus nonsimple loops on random regular graphs*, J. Graph Theory **105** (2024), no. 4, 562–579. MR 4709725 2
- [Fri91] Joel Friedman, *On the second eigenvalue and random walks in random d -regular graphs*, Combinatorica **11** (1991), no. 4, 331–362. MR 1137767 2
- [Fri08] ———, *A proof of Alon’s second eigenvalue conjecture and related problems*, Mem. Amer. Math. Soc. **195** (2008), no. 910, viii+100. MR 2437174 2
- [KMPT10] Jeong Han Kim, Ravi Montenegro, Yuval Peres, and Prasad Tetali, *A birthday paradox for Markov chains with an optimal bound for collision in the Pollard rho algorithm for discrete logarithm*, Ann. Appl. Probab. **20** (2010), no. 2, 495–521. MR 2650040 3
- [LP16] Eyal Lubetzky and Yuval Peres, *Cutoff on all Ramanujan graphs*, Geom. Funct. Anal. **26** (2016), no. 4, 1190–1216. MR 3558308 2
- [LS10] Eyal Lubetzky and Allan Sly, *Cutoff phenomena for random walks on random regular graphs*, Duke Math. J. **153** (2010), no. 3, 475–510. MR 2667423 2
- [MV06] Stephen D. Miller and Ramarathnam Venkatesan, *Spectral analysis of Pollard rho collisions*, Algorithmic number theory, Lecture Notes in Comput. Sci., vol. 4076, Springer, Berlin, 2006, pp. 573–581. MR 2282950 3
- [Yad16] Ariel Yadin, *Self-avoiding walks on finite graphs of large girth*, ALEA Lat. Am. J. Probab. Math. Stat. **13** (2016), no. 2, 521–544. MR 3519758 2