Received: 10 March 2023 | Revised: 28 September 2023 | Accepted: 2 November 2023
DOI: 10.1002/jgt.23056

ARTICLE

WILEY

# Simple versus nonsimple loops on random regular graphs

**Benjamin Dozier**[1] | **Jenya Sapir**[2]

[1]Department of Mathematics, Cornell University, Ithaca, New York, USA
[2]Department of Mathematics, Binghamton University, Binghamton, New York, USA

**Correspondence**
Benjamin Dozier, Department of Mathematics, Cornell University, Ithaca, NY, USA.
Email: benjamin.dozier@cornell.edu

**Abstract**
In this note, we solve the "birthday problem" for loops on random regular graphs. Namely, for fixed $d \geq 3$, we prove that on a random $d$-regular graph with $n$ vertices, as $n$ approaches infinity, with high probability: (i) almost all primitive nonbacktracking loops of length $k \prec \sqrt{n}$ are simple, that is, do not self-intersect, and (ii) almost all primitive nonbacktracking loops of length $k \succ \sqrt{n}$ self-intersect.

**KEYWORDS**
nonbacktracking walk, random graphs, regular graphs, spectral graph theory

## 1 | INTRODUCTION

The issue of the number of simple versus nonsimple loops on random regular graphs plays an important role in controlling their spectral gap [5, 8, 9]. The related study of nonbacktracking random walk on regular graphs has been pursued in [1, 13, 14] among other papers.

Our interest in this topic arose from consideration of the analogous question for closed geodesics on Weil–Petersson random hyperbolic surfaces, raised by Lipnowski-Wright [12, conjecture 1.2] in their study of spectral gap for these surfaces. There is a fruitful analogy between large regular graphs and hyperbolic surfaces of high genus (e.g., [6, 7]). Closed geodesics, as well as simple closed geodesics, on hyperbolic surfaces, are objects of intense study; their analogs on graphs are nonbacktracking loops.

The shortest nonbacktracking loop on any regular graph is simple, that is, passes through each vertex at most once. As we consider nonbacktracking loops of length $k$ getting larger, eventually all of them are nonsimple. At what length $k$ (as a function of $n$) do the nonsimple loops start to become more common? The scale at which the transition happens will depend on the shape of the graph; in this paper, we answer the question for the case of *random* regular graphs of fixed degree $d \geq 3$. Our result is that the transition occurs around the threshold $k = \sqrt{n}$.

  

## 1.1 | Terminology

We define a *walk* of length $k$ on a (multi)graph to be a sequence of oriented edges* $e_1 e_2 \cdots e_k$ with the terminal vertex of $e_i$ equal to the initial vertex of $e_{i+1}$ for $i = 1, ..., k-1$. We say a walk is a *loop* if the terminal vertex of $e_k$ is equal to the initial vertex of $e_1$. If the edge sequences of two loops differ by a cyclic shift, we consider the two loops to be the same. We say a walk is *nonbacktracking* if $\bar{e}_i \neq e_{i+1}$ (here $\bar{e}_i$ denotes $e_i$ with orientation reversed) for $i = 1, ..., k-1$. We say a loop is nonbacktracking if *all* of the walks associated to that loop are nonbacktracking (in particular, when a distinguished start vertex is chosen, we must have "nonbacktracking at the start").† A walk is said to be *simple* if the terminal endpoints of the edges are distinct, and a loop is said to be simple if all (equivalently, any) of the associated walks are simple. We say a loop is *primitive* if it is not the repetition of a shorter loop.

Our notion of simple loop coincides with the usual meaning of the term *oriented cycle* in the graph theory literature.

## 1.2 | Model of random regular graphs and notation

Let $G(d, n)$ be a random $d$-regular graph on $n$ vertices, chosen using the *configuration model*. One sample from this distribution is as follows. Begin with $n$ vertices, each with $d$ unpaired half-edges. At each stage, choose some unpaired half-edge and pair it with a different unpaired half-edge chosen uniformly at random. Repeat until all half-edges are paired. The resulting "graph" can have self-loops and multiple edges between a pair of vertices; nevertheless, we will abuse terminology and call them graphs.

Given $\Gamma$ a $d$-regular graph let:

- $N_{simp}(\Gamma, k)$ be the number of *simple* nonbacktracking loops on $\Gamma$ of length $k$,
- $N_{prim}(\Gamma, k)$ be the number of *primitive* nonbacktracking loops on $\Gamma$ of length $k$.

We use $N_{simp}(k)$ to denote the functions on the set of all $d$-regular graphs sending $\Gamma$ to the integer $N_{simp}(\Gamma, k)$, and we use $N_{prim}(k)$ similarly.

Let $\mathcal{P}_n[\cdot]$ denote the probability of an event depending on a graph $\Gamma$ drawn from the distribution $G(d, n)$, and let $\mathbb{E}_n[\cdot]$ denote the expected value. (Note that these quantities depend also on the degree $d$, but we will always think of $d$ as fixed). We will use the shorthand $\mathbb{E}_n[X(k)] := \mathbb{E}_n[X(\Gamma, k)]$ for $X$ any of the quantities in the list above, and likewise for $\mathcal{P}_n[X]$.

## 1.3 | Main results

In this paper, $A \prec B$ means that $A = o(B)$, and $A \sim B$ means $\lim \frac{A}{B} = 1$.

*For oriented edges, there is a small subtlety with self-loops. To define oriented edge, we begin by decomposing each edge of the graph into two half-edges, each of which has an incident vertex. An orientation of an edge is then a choice of an ordering of its two half-edges. Notice that with this defintion, any edge (including a self-loop) has exactly two orientations.

†Note that with our definitions, if $e$ is an oriented edge that is a self-loop, then the walk $ee \cdots e$ is considered to be nonbacktracking.

**Theorem 1.1** (Low length regime). *Take $d \geq 3$ fixed. Suppose $k$ is some function of $n$ satisfying $k \prec \sqrt{n}$. Fix $\epsilon > 0$. Then*

$$\mathcal{P}_n[N_{simp}(k) \geq (1-\epsilon)N_{prim}(k)] \to 1,$$

*as $n \to \infty$.*

**Theorem 1.2** (High length regime). *Take $d \geq 3$ fixed. Suppose $k$ is some function of $n$ satisfying $k \succ \sqrt{n}$. Fix $\epsilon > 0$. Then*

$$\mathcal{P}_n[N_{simp}(k) \leq \epsilon \cdot N_{prim}(k)] \to 1,$$

*as $n \to \infty$.*

In both theorems above, when $d$ is odd, for there to be any $d$-regular graphs on $n$ vertices, $n$ must be even. Thus in those cases, we take $n \to \infty$ along the even integers.

*Remark* 1.3. A considerable subtlety is that Theorem 1.1 would *not* be true if we replaced $N_{prim}(k)$ by $N_{all}(k)$, the number of nonbacktracking loops, without the primitive condition. In particular, it would fail for $k$ a fixed composite integer $pq$. In fact, in that case, it is known (by [3]) that $N_{simp}(k)$ and $N_{simp}(p)$ asymptotically have finite positive mean, and since $\mathbb{E}_n[N_{all}(k)] \geq \mathbb{E}_n[N_{simp}(k)] + \mathbb{E}_n[N_{simp}(p)]$, we get that $N_{all}(k)$ must be greater than $N_{simp}(k)$ by a definite factor a positive proportion of the time.

*Remark* 1.4. The above two theorems also hold if we replace the configuration model $G(d, n)$ with the *uniform model* over all $d$-regular graphs, without self-loops or multiple edges, on $n$ vertices. These versions can be deduced from the theorems above together with the result that the probability that a (multi)graph from $G(d, n)$ has neither self-loops nor multiple edges tends to a positive constant as $n \to \infty$ (see [3] or [4, theorem 2.16]).

### 1.3.1 | Nonrandom graphs

Note that the analog of Theorem 1.1 for fixed sequences of regular graphs (rather than random ones) fails. Families of graphs with diameter linear in $n$ provide counter-examples, since in this case walks behave like random walk on a line and thus even short walks (and loops) are likely to self-intersect.

On the other hand, we do not know if the nonrandom version of Theorem 1.2 holds:

**Question 1.5.** Let $n_j$ be an increasing sequence of positive integers, $\Gamma_{n_j}$ a $d$-regular graph on $n_j$ vertices, and $k$ some function of $n$ with $k(n) \succ \sqrt{n}$. Is it true that for any fixed $\epsilon > 0$,

$$N_{simp}(\Gamma_{n_j}, k(n_j)) < \epsilon \cdot N_{prim}(\Gamma_{n_j}, k(n_j))$$

for all $j$ sufficiently large (depending on $\epsilon$)?

## 1.4 | Discussion of the proofs

### 1.4.1 | Heuristic

A random loop of length $k$ on a random graph should behave in some sense like choosing a list of $k$ vertices uniformly at random from all $n$ vertices and making the loop travel through them in order. If this were the case, then the solution to the standard "birthday problem" for picking $k$ birthdays randomly from among $n$ suggests that the transition between all the vertices being distinct versus having at least one repetition should occur around $k = \sqrt{n}$.

Of course, on a $d$-regular graph, a nonbacktracking walk can only be extended by one step in $d - 1$ ways, so certainly, not all $n$ vertices are equally likely to come after some given vertex. However, since we are choosing the graph randomly as well, one can in fact think of the next vertex along a walk as being randomly selected from the $n$ vertices. This is because we can build the graph and walk simultaneously, only making choices about the graph when the walk forces us to. There are, however, two issues with doing this:

(i) it only gives control over expected values of counts; to get asymptotic almost sure (a.a.s.) control requires additional work,
(ii) it fails if the walk already self-intersects, since in that case we would not get the necessary freedom in the choice of next edge to follow.

If our ultimate goal was to study *walks* rather than loops, neither of these issues would be problematic. In fact, the technique suggested above easily gives that the transition for a random nonbacktracking walk on a random graph to be self-intersecting occurs around $k = \sqrt{n}$. Conditioning on the walk being a loop makes things considerably harder. Unlike the number of nonbacktracking walks, the number of nonbacktracking loops depends on the particular graph. This means that knowing the expected number of simple loops is not enough; we must know additional information about the distribution of the number of simple loops, and separate information about the count of primitive loops.

## 1.5 | Outline of paper

- In Section 2, we compute the expected number of simple loops. In the low length regime $k \prec \sqrt{n}$, we also bound the second moment of the number of simple loops and then use this to control the a.a.s. behavior. Our bound on the second moment is the primary novel ingredient in this paper.
- In Section 3, we study the count of primitive loops. In the *very* low length regime $k \prec n^{1/4}$, we control the expected value, while for $k \succ \log n$ we control the a.a.s. behavior.
- In Section 4, we combine the results from the previous sections to prove both the main theorems.

# 2 | COUNTING SIMPLE LOOPS

Estimating *expected* counts of simple loops, which we do below in Proposition 2.1 and Proposition 2.2, is well-understood and quick. The idea is to build the graph and walk simultaneously. We will also need control of the a.a.s. behavior, so we need to rule out high variance. In the low length regime $1 \prec k \prec \sqrt{n}$, we estimate the second moment (Proposition 2.3), and then use the second moment method to deduce the a.a.s. behavior (Proposition 2.4). In the high-length regime $k \succ \sqrt{n}$, since we are trying to prove an inequality of the form $N_{simp}(\Gamma, k) \leq \epsilon N_{prim}(\Gamma, k)$ for asymptotically almost all $\Gamma$, the first-moment method will suffice.

## 2.1 | Expected count of simple loops

**Proposition 2.1.** *Let $k \prec \sqrt{n}$. Then*

$$\mathbb{E}_n[N_{simp}(k)] \sim \frac{(d-1)^k}{k},$$

*as $n \to \infty$.*

*Proof.* The main idea is to build the graph and walk simultaneously. A similar result, with a similar proof, appears in [5, lemma 4] though the random model they use is somewhat different. Let $p$ denote the probability that a randomly chosen nonbacktracking walk of length $k$ on a random regular graph is closed and simple. To compute $p$, we will consider choosing the random graph at the same time as the random walk, only making choices about which half-edges are paired in the graph when we are forced to. The first factor below is the probability that the first half-edge the path follows is paired with a half-edge that leads to a different vertex (since there are $dn - 1$ half-edges it could be paired with, of which $d - 1$ are incident to the start vertex). Continuing in this way, we see that

$$p = \left(1 - \frac{d-1}{dn-1}\right)\left(1 - \frac{(d-1)+(d-2)}{dn-3}\right)\cdots\left(1 - \frac{(d-1)+(k-2)(d-2)}{dn-(2k-3)}\right) \cdot \left(\frac{d-1}{dn-(2k-1)}\right). \tag{1}$$

Since, a fortiori, $k \prec n$, we get

$$p \sim \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{k-1}{n}\right)\left(\frac{d-1}{d} \cdot \frac{1}{n}\right).$$

Then using the estimate $1 - x = \exp(-x + O(x^2))$ for small $x$ repeatedly gives

$$p \sim \exp\left(\frac{-k(k-1)}{2n}\right)\left(\frac{d-1}{d} \cdot \frac{1}{n}\right) \tag{2}$$

$$\sim \frac{d-1}{d} \cdot \frac{1}{n}, \tag{3}$$

where in the last step we have used the assumption that $k \prec \sqrt{n}$.

Now to compute the desired expected value, we note that every regular graph has $nd(d-1)^{k-1}$ nonbacktracking walks of length $k$. Each simple loop will be counted $k$ times in this way (recall that our walks and loops are all counted with *orientation*). Hence

$$\begin{aligned}\mathbb{E}_n[k \cdot N_{simp}(k)] &= nd(d-1)^{k-1}p \\ &\sim nd(d-1)^{k-1}\left(\frac{d-1}{d} \cdot \frac{1}{n}\right) \\ &\sim (d-1)^k,\end{aligned}$$

which then gives the desired result. □

**Proposition 2.2.** *Suppose $k$ is some function of $n$ satisfying $k \succ \sqrt{n}$. Then*

$$\mathbb{E}_n[N_{simp}(k)] \prec \frac{(d-1)^k}{k},$$

*as $n \to \infty$.*

*Proof.* Let $p$ denote the probability that a randomly chosen nonbacktracking walk of length $k$ on a random regular graph is closed and simple. We recall the exact formula (1) for $p$:

$$\begin{aligned}p = &\left(1 - \frac{d-1}{dn-1}\right)\left(1 - \frac{(d-1)+(d-2)}{dn-3}\right)\cdots\left(1 - \frac{(d-1)+(k-2)(d-2)}{dn-(2k-3)}\right) \\ &\cdot\left(\frac{d-1}{dn-(2k-1)}\right).\end{aligned}$$

Note that the factors in the first line above are decreasing. Hence we can upper bound each of the factors in the latter half by

$$1 - \frac{(d-1)+(k/2-1)(d-2)}{dn-(k-1)} \le 1 - \Omega(k/n).$$

(Here $f = \Omega(g)$ means $g = O(f)$, and the implicit constant can depend on $d$. We are also using that $k \le n$; otherwise, there are no simple walks of length $k$). Combining these bounds with the obvious upper bound of 1 for the first terms, and the bound $\frac{d-1}{dn-(2k-1)} = O(1/n)$ for the last term (since $d \ge 3$ and $k \le n$) gives

$$p \le (1 - \Omega(k/n))^{k/2-2} \cdot O(1/n).$$

Next, using the approximation $1 - x = \exp(-x + O(x^2))$ and the assumption that $k \succ \sqrt{n}$, we get from the above that

$$p = o(1) \cdot \frac{1}{n}.$$

Then, as in the proof of Proposition 2.1, we get

$$\begin{aligned}\mathbb{E}_n[k \cdot N_{simp}(k)] &= nd(d-1)^{k-1}p \\ &= nd(d-1)^{k-1} \cdot o(1) \cdot \frac{1}{n}. \\ &= o((d-1)^k),\end{aligned}$$

which implies the desired result. □

## 2.2 | Variance of count of simple loops

**Proposition 2.3.** *Let* $1 \prec k \prec \sqrt{n}$. *Then*

$$\mathbb{E}_n\left[N^2_{simp}(k)\right] \leq (1 + o(1)) \cdot \frac{(d-1)^{2k}}{k^2},$$

*as* $n \to \infty$.

*Proof.* We consider $\Gamma$ drawn from the configuration model $G(d, n)$. Let $L_{dist}$ denote the number of ordered pairs $(\gamma_1, \gamma_2)$, where each $\gamma_i$ is a simple loop of length $k$ on the graph $\Gamma$ with a choice of distinguished point on the loop, and such that $\gamma_1 \neq \gamma_2$ and $\gamma_1 \neq \overline{\gamma}_2$ (here $\overline{\gamma}$ denotes the loop $\gamma$ but with orientation reversed).

Since there are exactly $nd(d-1)^{k-1}$ nonbacktracking paths of length $k$, we have

$$\mathbb{E}_n[L_{dist}] = (nd(d-1)^{k-1})^2 p,$$

where $p$ is the probability for $\Gamma$ a randomly chosen regular graph and $\gamma_1, \gamma_2$ randomly chosen nonbacktracking walks on $\Gamma$, that $\gamma_1, \gamma_2$ are both simple loops with $\gamma_1 \neq \gamma_2, \gamma_1 \neq \overline{\gamma}_2$. We can write $p = p_1 + p_2$, where

- $p_1$ is the probability (under the same choices as above) that $\gamma_1, \gamma_2$ are both simple loops and that the initial vertex of $\gamma_2$ is *not* one of the vertices of $\gamma_1$,
- $p_2$ is the probability that $\gamma_1, \gamma_2$ are both simple loops, the initial vertex of $\gamma_2$ *is* one of the vertices of $\gamma_1$, and $\gamma_1 \neq \gamma_2, \gamma_1 \neq \overline{\gamma}_2$.

To bound these probabilities from above, we will again consider choosing the random graph at the same time as the random walks, only making choices about the graph when we are forced to.

To bound $p_1$, note that the probability that the last edge of $\gamma_1$ goes back to its initial vertex $v_1$ is at most $(1 + o(1))\frac{d-1}{d}\frac{1}{n}$, where the $o(1)$ bound on the error uses that $k \prec n$. The analogous statement is true for $\gamma_2$, even conditioning on $\gamma_1$, since the initial vertex of $\gamma_2$ is disjoint from $\gamma_1$. Hence we get

$$p_1 \leq (1 + o(1))\left(\frac{d-1}{d} \cdot \frac{1}{n}\right)^2.$$

For $p_2$, we get a factor of $O(1/n)$ from the condition that $\gamma_1$ is a simple loop (a better bound is used in the preceding analysis for $p_1$; this is all i.e., needed here). We also get a factor of $O(k/n)$ from the condition that the initial vertex of $\gamma_2$ coincides with a vertex of $\gamma_1$. Now if $(\gamma_1, \gamma_2)$ is a pair of simple loops counted in $L_{dist}$, there exists a unique edge $e$ of $\gamma_2$ such that $e$ is part of neither $\gamma_1$ nor $\bar{\gamma}_1$, and such that after $\gamma_2$ traverses $e$, it exactly follows either $\gamma_1$ or $\bar{\gamma}_1$ until its final vertex (Figure 1). There are at most $k$ choices of where along $\gamma_2$ this edge $e$ is, and the probability that the forward endpoint of $e$ coincides with the appropriate vertex of $\gamma_1$ is $O(1/n)$. Thus we get a further factor of $O(k/n)$. Putting this all together, we find

$$\begin{aligned} p_2 &= O\left(\frac{1}{n} \cdot \frac{k}{n} \cdot \frac{k}{n}\right) \\ &= o\left(\frac{1}{n^2}\right), \end{aligned}$$

where in the last step, we have used the assumption that $k^2 \prec n$.

Putting the two estimates together gives

$$\begin{aligned} p = p_1 + p_2 &\leq (1 + o(1))\left(\frac{d-1}{d} \cdot \frac{1}{n}\right)^2 + o\left(\frac{1}{n^2}\right) \\ &\leq (1 + o(1))\left(\frac{d-1}{d} \cdot \frac{1}{n}\right)^2. \end{aligned}$$

Now returning to $L_{dist}$, we get

$$\begin{aligned} \mathbb{E}_n[L_{dist}] &= (nd(d-1)^{k-1})^2 p \\ &\leq (nd(d-1)^{k-1})^2 (1 + o(1))\left(\frac{d-1}{d} \cdot \frac{1}{n}\right)^2 \\ &\leq (1 + o(1)) \cdot (d-1)^{2k}. \end{aligned}$$

To finish the proof, we combine the above with the term coming from pairs of loops that are either identical or differ only in orientation. There are $k^2 N_{simp}^2(\Gamma, k)$ pairs of loops with distinguished start points. So using Proposition 2.1 we compute:

$$\begin{aligned} \mathbb{E}_n\left[k^2 \cdot N_{simp}^2(k)\right] &= \mathbb{E}_n[L_{dist} + 2k^2 \cdot N_{simp}(k)] = \mathbb{E}_n[L_{dist}] + \mathbb{E}_n[2k^2 \cdot N_{simp}(k)] \\ &\leq (1 + o(1)) \cdot (d-1)^{2k} + 2(1 + o(1)) \cdot k(d-1)^k \\ &\leq (1 + o(1)) \cdot (d-1)^{2k}, \end{aligned}$$

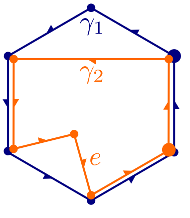


FIGURE 1 When the initial vertex of $\gamma_2$ lies on $\gamma_1$. [Color figure can be viewed at wileyonlinelibrary.com]

where in the last line we have used the assumption that $1 \prec k$. Dividing by $k^2$ gives the desired result. ☐

### 2.3 | Asymptotic almost sure behavior of simple loops

**Proposition 2.4.** *Let* $1 \prec k \prec \sqrt{n}$. *Fix* $\epsilon > 0$. *Then*

$$\lim_{n\to\infty} \mathcal{P}_n\left[\left|\frac{N_{simp}(k)}{(d-1)^k/k} - 1\right| < \epsilon\right] = 1.$$

*Remark* 2.5. Note that above is false for $k$ a constant, since in that case $N_{simp}(k)$ has a Poisson distribution with positive variance (3, theorem 2).

*Proof.* This follows from the second-moment method (Chebyshev inequality), using the expected value estimate of Proposition 2.1 and the second-moment estimate of Proposition 2.3. ☐

## 3 | COUNTING PRIMITIVE LOOPS

When the length is very low ($k \prec n^{1/4}$), we will show that the probability that a nonbacktracking walk forms *two* distinct, primitive loops in its induced subgraph is so low that this probability is still negligible conditioned on the event that the walk is a loop. It follows that in this regime the expected number of *primitive* loops (Proposition 3.1) has the same asymptotics as the expected number of simple loops, computed in the previous section.

When $k \succ \log n$, we take the standard approach of using the spectral gap for the adjacency matrix $A$ to control the a.a.s. count of loops (Proposition 3.7). The trace of $A^k$ counts all loops of length $k$, without the nonbacktracking condition. Since we are interested in nonbacktracking loops, we study the related "nonbacktracking matrix" $\tilde{A}$ whose eigenvalues can be computed in terms of those of $A$.

### 3.1 | Length $k \prec n^{1/4}$

**Proposition 3.1.** *Let* $d \geq 3$, *and let* $k \prec n^{1/4}$. *Then*

$$\mathbb{E}_n[N_{prim}(k)] \sim \frac{(d-1)^k}{k}$$

*as* $n \to \infty$.

*Proof.* We will compute the expected value as follows. On any graph from $G(d, n)$, the total number of nonbacktracking walks $\gamma$ of length $k$ equals

$$nd(d-1)^{k-1}, \tag{4}$$

since there are $n$ choices of starting vertex, then $d$ choices for the first outgoing edge, and $(d-1)$ choices for the succeeding outgoing edges (note that if the walk happens to be a loop, then the resulting loop could potentially backtrack at the starting vertex).

We now consider choosing such a $\gamma$ randomly, that is, we choose a graph randomly according to $G(d, n)$, pick a random start vertex, and then pick a random nonbacktracking walk starting at that vertex. Combined with (4), to prove the Proposition, it will suffice to compute the probability $p_{prim}$ that $\gamma$ is a loop that is primitive and nonbacktracking.

We begin by showing that $\gamma$ forming multiple loops is unlikely.

*Claim* 3.2. The probability that the induced graph formed by $\gamma$ has at least two distinct primitive loops is $O(k^4/n^2)$.

*Proof.* A similar result appears as [5, lemma 3] though the random model there is somewhat different. We proceed by building the random graph and the random walk at the same time. If there are two loops it means that we had at least two "free choices" of an edge that came back to vertices already on the walk. By free, we mean that we picked a half-edge to follow that was unpaired. The probability that the half-edge that this gets paired to is incident to one of the vertices already in the walk is $O(k/(n-k))$ which, since we are assuming $k \prec n^{1/4}$, is $O(k/n)$. There are $\binom{k}{2}$ choices for the two steps at which the collision occurs. Thus the probability of interest is $O(k^2(k/n)^2)$, as desired. □

*Claim* 3.3. Given a nonbacktracking, primitive, nonsimple loop $\gamma$ on a graph, its induced subgraph must contain at least two distinct primitive loops.

*Proof.* Let $k$ be the length of the loop. Choose a starting point $v_1$, and let $\gamma$ traverse the vertices $v_1, \ldots, v_k, v_{k+1} = v_1$ in that order. We can choose the starting point such that a simple loop $\alpha$ is formed by $v_1, v_2, \ldots, v_i = v_1$ for some $1 < i < k$ (we use both nonsimple and nonbacktracking properties here). After $v_i$, the walk may follow this loop $\alpha$ several times, but since the loop $\gamma$ is primitive, eventually it must depart from $\alpha$, say at step $j$. A new loop $\beta$ is then formed somewhere between steps $j$ and $k$. □

Let $p_s$ be the probability that $\gamma$ is a simple loop. In (3), we showed that when $k \prec \sqrt{n}$,

$$p_s \sim \frac{1}{n}\frac{d-1}{d}.$$

By the two claims above, the probability that $\gamma$ is a primitive loop that is nonbacktracking (including at start vertex) and *nonsimple* is

$$p_{ns} = O(k^4/n^2) = o(1/n).$$

Combinging these two, we get that the probability that $\gamma$ is a primitive, nonbacktracking loop is

$$p_{prim} = p_s + p_{ns} \sim \frac{1}{n}\frac{d-1}{d} + o(1/n) \sim \frac{1}{n}\frac{d-1}{d}.$$

We combine this with (4) to compute the expected number of primitive, nonbacktracking loops of length $k$, with a distinguished start-vertex to be

$$nd(d-1)^{k-1} \cdot p_{prim} \sim nd(d-1)^{k-1} \cdot \frac{1}{n}\frac{d-1}{d} = (d-1)^k.$$

There are $k$ choices of distinguished start point along the loop, so if we forget this, then the quantity goes down by a factor of $k$, giving the desired result. □

## 3.2 | Length $k \succ \log n$

### 3.2.1 | Nonbacktracking matrix

Much of the material in this section is well-known, but does not appear in the literature in the exact form that we need, so we give an exposition for the reader's convenience.

Let $A$ be the $n \times n$ adjacency matrix of $d$-regular graph $\Gamma$. Since $\Gamma$ is undirected, this graph is symmetric, so $A$ has $n$ real eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

Since $\Gamma$ is $d$-regular, $\lambda_1 = d$. Let $\lambda = \max(|\lambda_2|, |\lambda_n|)$.

Powers of $A$ and associated spectral information can be used to understand counts of walks with backtracking allowed. Since we are interested in nonbacktracking walks, we consider the related Markov process that keeps track of edges in a path with orientation. This Markov process is well-established in the literature (see, e.g., [1, 2, 13]). We redefine it here for completeness.

The nonbacktracking Markov process $\tilde{\Gamma}$ is defined as follows. For each edge of $\Gamma$, we let each state of $\tilde{\Gamma}$ correspond to an edge of $\Gamma$ with a choice of orientation. In other words, the number of states of $\tilde{\Gamma}$ is exactly $nd$, or twice the number of edges of $\Gamma$. Consider two states $e_1, e_2$ of $\tilde{\Gamma}$ (i.e., two edges of $\Gamma$ with choice of orientation). We make a connection from $e_1$ to $e_2$ if $e_1, e_2$ are not opposite orientations of the same edge, and the forward endpoint of $e_1$ equals the back endpoint of $e_2$ (these conditions mean that going from $e_1$ to $e_2$ is a valid move in a nonbacktracking walk). In particular, if $e_1$ is represents an oriented self-loop in $\Gamma$, then it is connected to itself in $\tilde{\Gamma}$. And if two vertices in $\Gamma$ are joined by multiple edges, then these edges induce a complete bipartite subgraph in $\tilde{\Gamma}$. In this way, each vertex of $\tilde{\Gamma}$ has $d-1$ incoming edges and $d-1$ outgoing edges. Observe that there is a bijection between nonbacktracking loops on $\Gamma$ and all (directed) loops on $\tilde{\Gamma}$. We consider the Markov process corresponding to $\tilde{\Gamma}$. In particular, given a state $e_1$ we think of this process as choosing the next allowable state uniformly at random.

To count nonbacktracking loops, we will study the adjacency matrix $\tilde{A}$ of $\tilde{\Gamma}$. We call $\tilde{A}$ the *nonbacktracking matrix* associated to the graph $\Gamma$. We denote by $\mu$ the magnitude of the second largest number in magnitude among the eigenvalues of $\tilde{A}$ (listed with multiplicity).

The following lemma says that if $A$ has spectral gap, then so does $\tilde{A}$.

**Lemma 3.4.** *Let $A$ be the adjacency matrix of any regular graph, and $\tilde{A}$ the corresponding nonbacktracking matrix. For every $\epsilon > 0$, there is some $\delta > 0$ so that if*

$$\lambda < d - \epsilon$$

*then*

$$\mu < (d-1) - \delta.$$

*Proof.* If $\Gamma$ is not connected, then there is no spectral gap: $\lambda = d$. If $\Gamma$ is connected, it is well known that the eigenvalues of $\tilde{A}$ can be expressed in terms of those of $A$ as

$$\left(\mu_1^+, \mu_1^-, \ldots, \mu_n^+, \mu_n^-, \overbrace{1, \ldots, 1}^{nd/2-n}, \overbrace{-1, \ldots, -1}^{nd/2-n}\right),$$

where

$$\mu_i^\pm := \frac{\lambda_i \pm \sqrt{\lambda_i^2 - 4(d-1)}}{2}.$$

This was originally shown using Ihara's theorem; see [10, 11]. The statement above appears as part of [13, proposition 3.1]. It can be proved by explicitly finding the $\tilde{A}$ eigenspaces for $\pm 1$ as well as eigenvectors of $\tilde{A}$ corresponding to those of $A$ (see [13, first part of proof of proposition 3.1 and remark 3.4]).

Let $f(x) = \frac{x + \sqrt{x^2 - 4(d-1)}}{2}$.

*Claim* 3.5.

$$\mu \le \max\{\sqrt{d-1}, f(d-\epsilon)\}.$$

(Note that if necessary, we can decrease $\epsilon$ so that $f(d-\epsilon)$ is real).

*Proof.* Note that since $\lambda_1 = d$, we have $\mu_1^+ = d-1$ and $\mu_1^- = 1$. So it suffices to show that the remaining eigenvalues of $\tilde{A}$ are all at most $\max\{\sqrt{d-1}, f(d-\epsilon)\}$ in magnitude.

If $|\lambda_i| < 2\sqrt{d-1}$, then both $\mu_i^\pm$ are nonreal, and have magnitude equal to $\sqrt{d-1}$, which can be seen by multiplying by the conjugate.

If $|\lambda_i| \ge 2\sqrt{d-1}$, and $i \ne 1$, then both $\mu_i^\pm$ are real. In this case, if $\lambda_i \ge 0$, then

$$0 \le \mu_i^- \le \mu_i^+ = f(\lambda_i) \le f(\lambda) \le f(d-\epsilon),$$

where in the last two inequalities, we have used that $f$ is an increasing function on $[2\sqrt{d-1}, \infty)$. When $\lambda_i \le 0$, arguing similarly gives $|\mu_i^\pm| \le f(d-\epsilon)$, completing the proof. $\square$

Applying this claim we get

$$\begin{aligned}\mu \le \max\{\sqrt{d-1}, f(d-\epsilon)\} &\le f(d) - \min\{f(d) - \sqrt{d-1}, f(d) - f(d-\epsilon)\} \\ &= (d-1) - \min\{d-1-\sqrt{d-1}, f(d) - f(d-\epsilon)\}.\end{aligned}$$

Combining this with the fact that $f$ is a *strictly* increasing function on $[2\sqrt{d-1}, \infty)$ gives the existence of $\delta$ with the desired property. □

**Lemma 3.6.** *Let $\tilde{A}$ be the nonbacktracking matrix for a graph from $G(d, n)$, and let $\mu$ be the largest magnitude among the eigenvalues of $\tilde{A}$ other than $\mu_1$. Then there exists $\delta > 0$ such that*

$$\lim_{n\to\infty} \mathcal{P}_n[\mu < (d-1) - \delta] = 1.$$

*Proof.* Let $\lambda$ be the quantity for $A$ defined above. It is well known (see e.g., [5, theorem 13]) that there is some $\epsilon > 0$ such that

$$\lim_{n\to\infty} \mathcal{P}_n[\lambda < d - \epsilon] = 1.$$

Then applying Lemma 3.4 to such an $A$ gives a $\delta$ such that $\mu < (d-1) - \delta$. □

**Proposition 3.7.** *Suppose $k$ is some function of $n$ satisfying $k \succ \log n$. Fix $\epsilon > 0$. Then*

$$\mathcal{P}_n\left[\left|\frac{N_{prim}(k)}{(d-1)^k/k} - 1\right| < \epsilon\right] \to 1,$$

*as $n \to \infty$.*

*Proof.* The nonbacktracking matrix $\tilde{A}$ associated to a $d$-regular graph $\Gamma$ has dimensions $nd \times nd$. Let $m = nd$.

Let

$$\mu_1 = d - 1, \mu_2, \ldots, \mu_m,$$

be the eigenvalues of $\tilde{A}$, listed with (algebraic) multiplicity, arranged in order of decreasing magnitude.

By Lemma 3.6 there exists $\delta > 0$ such that

$$|\mu_i| < (d-1) - \delta, \text{ for } i = 2, \ldots, m, \tag{5}$$

with probability tending to 1 as $n \to \infty$.

Now note that the $j$th diagonal entry of $\tilde{A}^k$ counts the number of directed walks on $\tilde{\Gamma}$ that start and end at the $j$th vertex. We denote by $N_{tr}(k)$ the number of nonbacktracking walks on $\Gamma$ that start and end at the same vertex. Then

$$\begin{aligned} N_{tr}(k) = \mathrm{Trace}(\tilde{A}^k) &= \mu_1^k + \mu_2^k + \cdots + \mu_m^k \\ &= (d-1)^k + \mu_2^k + \cdots + \mu_m^k \\ &= (d-1)^k + O(|\mu_2|^k + \cdots + |\mu_m|^k) \\ &= (d-1)^k + O(m \cdot \max_{i\geq 2} |\mu_i|^k). \end{aligned}$$

By the above spectral gap bound (5) on the $\mu_i$, we get that, with probability tending to 1 as $n \to \infty$,

$$N_{tr}(k) = (d-1)^k + O(m((d-1)-\delta)^k).$$

Recall that $m = nd$, so when $k \succ \log n$, the above gives

$$N_{tr}(k) = (1 + o(1)) \cdot (d-1)^k \tag{6}$$

with probability tending to 1 as $n \to \infty$.

Now we can express $N_{tr}(k)$ in terms of $N_{prim}(\Gamma, k)$. Each loop counted by $N_{tr}$ has a period $r|k$, and there are $r$ distinct choices of starting edge. So we have:

$$N_{tr}(k) = \sum_{r|k} r \cdot N_{prim}(\Gamma, r).$$

We isolate $N_{prim}(\Gamma, k)$ and use the immediate universal inequality $N_{prim}(\Gamma, r) \leq nd^r$ holding for any $\Gamma, r, n$, together with (6) to get, with probability tending to 1 as $n \to \infty$,

$$\begin{aligned} k \cdot N_{prim}(\Gamma, k) &= N_{tr}(k) - \sum_{r|k, r\neq k} r \cdot N_{prim}(\Gamma, r) \\ &= N_{tr}(k) - O\left(\sum_{r|k, r\neq k} r \cdot n \cdot d^r\right) \\ &= (1 + o(1)) \cdot (d-1)^k - O(k^2 \cdot n \cdot d^{k/2}) \\ &= (1 + o(1)) \cdot (d-1)^k - o((d-1)^k), \end{aligned}$$

where we have used that $k \succ \log n$ to get the last line. The desired result follows. □

*Remark* 3.8. When $d \geq 5$, one can accurately bound the expectation of $N_{prim}(\Gamma, k)$ for $k \prec \log n$ using more quantitative information about the eigenvalues of random regular graphs (namely [8, theorem 3.1]). But these methods do not seem to work for $d = 3, 4$.

## 4 | SIMPLE VERSUS PRIMITIVE LOOPS

In this section, we prove the main theorems by combining the results that we've proved about counting simple and primitive loops in the previous two sections.

### 4.1 | Low length regime

*Proof of Theorem* 1.1. We will prove the theorem when the order of growth $k$ lies in one of three specific ranges: (1) $k$ constant, (2) $1 \prec k \prec n^{1/4}$, (3) $\log n \prec k(n) \prec \sqrt{n}$. In general, $k(n)$ need not stay in any of these three ranges; however the general case reduces to these. In fact, to show that the desired limit of probabilities is 1, it suffices to show that any subsequence $\{n_i\}$ has a further subsequence $\{n_{i_j}\}$ along which the limit is 1. Given $\{n_i\}$, we can always find $\{n_{i_j}\}$ along which the growth of $k$ falls into one of the three cases. Hence, by the below, the limit of the probabilities along $\{n_{i_j}\}$ will be 1.

*Case* 1: $k$ is constant.

Our proof uses (i) $N_{simp}(k) \leq N_{prim}(k)$, (ii) the integrality of $N_{simp}(k)$, $N_{prim}(k)$, and (iii) our results on the expectation of $N_{simp}(k)$, $N_{prim}(k)$, in particular that these converge to the same *constant* value in this regime.

Since $N_{simp}(k) \leq N_{prim}(k)$, and both are valued in nonnegative integers, we have

$$\mathbb{E}_n[N_{prim}(k)] - \mathbb{E}_n[N_{simp}(k)] \geq \mathcal{P}_n[N_{prim}(k) \neq N_{simp}(k)] \cdot 1.$$

By Proposition 2.1 and Proposition 3.1, the quantities $\mathbb{E}[N_{prim}(k)]$ and $\mathbb{E}[N_{simp}(k)]$ both converge to the constant $(d-1)^k/k$, and hence the left-hand side of the above tends to 0 as $n \to \infty$. It follows that

$$\lim_{n\to\infty} \mathcal{P}_n[N_{simp}(k) \neq N_{prim}(k)] = 0.$$

So for any $\epsilon > 0$, we get that

$$\mathcal{P}_n[N_{simp}(k) \geq (1-\epsilon)N_{prim}(k)] \geq \mathcal{P}_n[N_{simp}(k) = N_{prim}(k)] \to 1,$$

as $n \to \infty$, as desired.

*Case* 2: $1 \prec k(n) \prec n^{1/4}$

In previous sections, we have gained control over the expectation of $N_{simp}(k)$ and $N_{prim}(k)$ in this regime, showing that they have the same asymptotic behavior. And of course, we also have $N_{simp}(k) \leq N_{prim}(k)$. However, these facts alone are not enough to deduce the desired result. For instance, we need to rule of the situation in which with probability 1/2, $N_{simp}(k) = 1$ and $N_{prim}(k) = 2$, and with probability 1/2, both $N_{simp}(k)$ and $N_{prim}(k)$ are around $2(d-1)^k/k$. Note that in this case, $\mathbb{E}[N_{simp}(k)], \mathbb{E}[N_{prim}(k)] \sim (d-1)^k/k$ as $k \to \infty$, but there is a 1/2 chance that the ratio is equal to 2. This type of situation will be ruled out by our control of the a.a.s. behavior of $N_{simp}(k)$ (which was proved by bounding the second moment of $N_{simp}(k)$).

We begin by defining two random variables

$$X := \frac{N_{simp}(\Gamma, k)}{(d-1)^k/k}, \; Y := \frac{N_{prim}(\Gamma, k)}{(d-1)^k/k}.$$

Note that $X \leq Y$ everywhere. Moreover, by our assumptions on $k(n)$, we have that $1 \prec k \prec n^{1/4}$. So $\mathbb{E}[X] \sim 1$ and $\mathbb{E}[Y] \sim 1$ by Propositions 2.1 and 3.1. Thus, $\mathbb{E}[Y] - \mathbb{E}[X] = o(1)$. In particular, for any $\epsilon > 0$, for $n$ large enough

$$\mathbb{E}[Y] - \epsilon \leq \mathbb{E}[X] \leq \mathbb{E}[Y].$$

So by Lemma 4.1, we get that

$$\mathcal{P}_n(X \geq Y - \sqrt{\epsilon}) \geq 1 - \sqrt{\epsilon}, \tag{7}$$

for $n$ large enough.

The above means that $X, Y$ are additively close with high probability, but the desired statement is about multiplicative closeness (which need not follow from additive closeness if both $X, Y$ are small). We will achieve this by bounding $X$ (and hence $Y$) from below almost surely.

By Proposition 2.4, we have that $Y \geq X > 1/2$ with probability at least $1 - \epsilon$ for all $n$ large enough. Hence

$$Y - \sqrt{\epsilon} > (1 - 2\sqrt{\epsilon})Y$$

with probability at least $1 - \epsilon$ for large $n$.

Combining this with (7) gives that

$$\mathcal{P}_n[X \geq (1 - 2\sqrt{\epsilon})Y] \geq \mathcal{P}_n[X \geq Y - \sqrt{\epsilon}] - \epsilon \geq 1 - \sqrt{\epsilon} - \epsilon,$$

when $n$ is sufficiently large. This implies the desired result.

*Case* 3: $\log n \prec k(n) \prec \sqrt{n}$

The proof in this case, comes directly from our control of the a.a.s. behavior of both $N_{simp}(k), N_{prim}(k)$ in this regime.

Applying Proposition 2.4 and Proposition 3.7 gives that, for any $\epsilon > 0$, with probability approaching 1 as $n \to \infty$, we have both

$$N_{simp}(k) \geq (1 - \epsilon)(d - 1)^k/k,$$
$$N_{prim}(k) \leq (1 + \epsilon)(d - 1)^k/k.$$

It follows that

$$N_{simp}(k) \geq \frac{1 - \epsilon}{1 + \epsilon} \cdot N_{prim}(k),$$

with probability approaching 1 as $n \to \infty$. This implies the desired result. □

Below is a version with additive error of the basic probability fact that if one random variable dominates another and they have the same expectation, then they are equal almost everywhere.

**Lemma 4.1.** *Let $X, Y$ be random variables (on the same probability space) with $X \leq Y$ everywhere. Suppose that for some $\epsilon > 0$,*

$$\mathbb{E}[X] \geq \mathbb{E}[Y] - \epsilon.$$

*Then*

$$\mathcal{P}[X \geq Y - \sqrt{\epsilon}] \geq 1 - \sqrt{\epsilon}.$$

*Proof.* Applying Markov's inequality to the nonnegative random variable $Y - X$, and using that $\mathbb{E}[Y - X] \leq \epsilon$ gives

$$\mathcal{P}[Y - X > \sqrt{\epsilon}] \leq \frac{1}{\sqrt{\epsilon}} \cdot \mathbb{E}[Y - X] \leq \frac{1}{\sqrt{\epsilon}} \cdot \epsilon = \sqrt{\epsilon}.$$

The desired result follows by considering the complementary event. □

## 4.2 | High length regime

*Proof of Theorem* 1.2. In this regime the only input we need is the expected count of simple loops and the a.a.s. count of primitive loops.

Let

$$X = \frac{N_{simp}(\Gamma, k)}{(d-1)^k/k}, \; Y = \frac{N_{prim}(\Gamma, k)}{(d-1)^k/k}.$$

For the expected simple count, by Proposition 2.2,

$$\lim_{n \to \infty} \mathbb{E}_n[X] = 0.$$

We now can apply the first-moment method; by Markov's inequality, we have

$$\mathcal{P}_n[X \geq \epsilon/2] \leq \frac{\mathbb{E}_n[X]}{\epsilon/2} \to 0, \tag{8}$$

as $n \to \infty$.

For the a.a.s. primitive count, by Proposition 3.7, we have

$$\lim_{n \to \infty} \mathcal{P}_n[1 - \epsilon < Y < 1 + \epsilon] = 1. \tag{9}$$

Then

$$\mathcal{P}_n[N_{simp}(k) \geq \epsilon \cdot N_{prim}(k)] = \mathcal{P}_n[X \geq \epsilon \cdot Y] \leq \mathcal{P}_n[X \geq \epsilon/2] + \mathcal{P}_n[Y \leq 1/2],$$

and the above goes to 0 as $n \to \infty$, by (8) and (9). □

**ACKNOWLEDGMENTS**

We would like to thank Noga Alon, Lionel Levine, and Alex Wright for useful conversations.

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## REFERENCES


1. N. Alon, I. Benjamini, E. Lubetzky, and S. Sodin, *Non-backtracking random walks mix faster*, Commun. Contemp. Math. **9** (2007), no. 4, 585–603.
2. O. Angel, J. Friedman, and S. Hoory, *The non-backtracking spectrum of the universal cover of a graph*, Trans. Am. Math. Society **367** (2015), no. 6, 4287–4387. https://doi.org/10.1090/S0002-9947-2014-06255-7
3. B. Bollobás, *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs*, Eur. J. Combin. **1** (1980), no. 4, 311–316.
4. B Bollobás, *Random graphs*, 2nd ed. Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001.
5. A. Broder and E. Shamir, *On the second eigenvalue of random regular graphs*, In 28th Annual Symposium on Foundations of Computer Science (sfcs 1987). 1987, pp. 286–294.
6. R. Brooks and E. Makover, *Random construction of Riemann surfaces*, J. Differential Geom. **68** (2004), no. 1, 121–157.
7. T. Budzinski, N. Curien, and B. Petri, *On the minimal diameter of closed hyperbolic surfaces*, Duke Math. J. **170** (2021), no. 2, 365–377.
8. J. Friedman, *On the second eigenvalue and random walks in random $d$-regular graphs*, Combinatorica. **11** (1991), no. 4, 331–362.
9. J. Friedman, *A proof of Alon's second eigenvalue conjecture and related problems*, Mem. Amer. Math. Soc. **195** (2008), no. 910, viii+100.
10. Y. Ihara, *On discrete subgroups of the two by two projective linear group over $p$-adic fields*, J. Math. Soc. Japan. **18** (1966), 219–235.
11. M. Kotani and T. Sunada, *Zeta functions of finite graphs*, J. Math. Sci. Univ. Tokyo. **7** (2000), no. 1, 7–25.
12. M. Lipnowski and A. Wright, *Towards optimal spectral gaps in large genus*, Annal. of Prob. (2021). https://imstat.org/journals-and-publications/annals-of-probability/annals-of-probability-future-papers/
13. E. Lubetzky and Y. Peres, *Cutoff on all Ramanujan graphs*, Geom. Funct. Anal. **26** (2016), no. 4, 1190–1216.
14. E. Lubetzky and A. Sly, *Cutoff phenomena for random walks on random regular graphs*, Duke Math. J. **153** (2010), no. 3, 475–510.


**How to cite this article:** B. Dozier, and J. Sapir, *Simple versus nonsimple loops on random regular graphs*, J. Graph. Theory. (2023), 1–18. https://doi.org/10.1002/jgt.23056