Small subgraphs in the trace of a random walk

Michael Krivelevich[*] Peleg Michaeli

School of Mathematical Sciences
Raymond and Beverly Sackler Faculty of Exact Sciences
Tel Aviv University
Tel Aviv, 6997801, Israel

{krivelev,peleg.michaeli}@math.tau.ac.il

Submitted: May 29, 2016; Accepted: Jan 30, 2017; Published: Feb 17, 2017
Mathematics Subject Classifications: 05C81, 05C80, 60G50

Abstract

We consider the combinatorial properties of the trace of a random walk on the complete graph and on the random graph $G(n,p)$. In particular, we study the appearance of a fixed subgraph in the trace. We prove that for a subgraph containing a cycle, the threshold for its appearance in the trace of a random walk of length m is essentially equal to the threshold for its appearance in the random graph drawn from $G(n,m)$. In the case where the base graph is the complete graph, we show that a fixed forest appears in the trace typically much earlier than it appears in $G(n,m)$.

Keywords: random walk, random graph, small subgraph

1 Introduction

For a positive integer n and a real $p \in [0,1]$, we denote by $G(n,p)$ the probability space of all (simple) labelled graphs on the vertex set $[n] = \{1,\dots,n\}$, where every pair of vertices is connected independently with probability p. A closely related model, which we denote by $G(n,m)$, is the *uniform* probability space over all graphs on n vertices with m edges. Both models have been extensively studied since first introduced by Gilbert [7], and by Erdős and Rényi [4,5].

One of the problems studied in [5] was the problem of finding the threshold for the appearance of a fixed subgraph. Formally, given a fixed graph H, one is interested in the smallest value of p_0 such that when $p \gg p_0$ the random graph $G(n,p)$ contains a copy of H *with high probability* (**whp**), that is, with probability tending to 1 as n grows. It turns

[*]Research supported in part by a USA-Israel BSF Grant and by a grant from Israel Science Foundation.

out that the threshold for the appearance of H is determined by $m_0(H)$, the maximum edge density of all of its non-empty subgraphs. In symbols,

$$m_0(H) = \max\left\{\frac{|E(H')|}{|V(H')|} \mid H' \subseteq H,\ |V(H')| > 0\right\}.$$

The problem of finding the threshold for every fixed subgraph was settled by Bollobás [3] in 1981, and the result can be stated as follows (see also [1, Section 4.4] or [9, Theorem 3.4]).

Theorem 1.1. *Let H be a fixed non-empty graph and let $G \sim G(n,p)$. Then*

$$\lim_{n\to\infty} \mathbb{P}(H \subseteq G) = \begin{cases} 0 & p \ll n^{-1/m_0(H)} \\ 1 & p \gg n^{-1/m_0(H)}. \end{cases}$$

Theorem 1.2. *Let H be a fixed non-empty graph and let $G \sim G(n,m)$. Then*

$$\lim_{n\to\infty} \mathbb{P}(H \subseteq G) = \begin{cases} 0 & m \ll n^{2-1/m_0(H)} \\ 1 & m \gg n^{2-1/m_0(H)}. \end{cases}$$

Here and later, the notation $f \gg g$ means that $f/g \to \infty$. For a vertex v, denote by $N(v)$ the set of its neighbours, and let $N^+(v) = \{v\} \cup N(v)$. Given a (finite) base graph $G = (V,E)$, a (lazy) *simple random walk* on G is a stochastic process $(X_0, X_1, \ldots)$ where X_0 is sampled uniformly at random from V, and for $t \geqslant 0$, X_{t+1} is sampled uniformly at random from $N^+(X_t)$, independently of the past. The *trace* of the random walk at time t is the (random) subgraph $\Gamma_t \subseteq G$ on the same vertex set, whose edges consist of all edges traversed by the walk by time t, excluding loops and suppressing possible edge multiplicity. Formally,

$$E(\Gamma_t) = \{\{X_{s-1}, X_s\} \mid 0 < s \leqslant t,\ X_{s-1} \neq X_s\}.$$

Note. There are various definitions of laziness of random walks, perhaps the most common is staying put with probability $1/2$ (see, e.g., [12]); however, for the case of random walks on the complete graph on n vertices, a random walk which stays put with probability $1/n$ yields an independent sequence of uniformly distributed locations, which is far easier to handle. We decided therefore to adopt here a general definition of laziness which, in the case of the complete graph, behaves like that. However, as the thresholds discussed in this work are coarse, the results below can be applied for more traditional definitions of laziness, as well as for non-lazy random walks.

In [2] it was shown that the trace of a random walk whose length is proportional to n^2 on (dense) *quasirandom* graphs (including dense random graphs) on n vertices is typically quasirandom. In [6], several results were given concerning graph-theoretic properties of the trace, for sparser base graphs and shorter random walks. In this paper we continue this study of the structure of the trace, finding thresholds for the appearance of fixed subgraphs. Our first result, which is analogous to Theorem 1.2, considers the random walk on the random graph $G(n,p)$, and is restricted to fixed subgraphs containing a cycle. As we will see later, that restriction is necessary, as the statement is simply false for forests.

Note that the condition $m_0(H) \geqslant 1$ is equivalent to the condition of containing a cycle.

Theorem 1.3. *Let H be a fixed graph with $m_0(H) \geqslant 1$, let $\varepsilon > 0$, $p \geqslant n^{-1/m_0(H)+\varepsilon}$ and $G \sim G(n,p)$, and let Γ_t be the trace of a random walk of length t on G. Then*

$$\lim_{n\to\infty} \mathbb{P}\left(H \subseteq \Gamma_t\right) = \begin{cases} 0 & t \ll n^{2-1/m_0(H)} \\ 1 & t \gg n^{2-1/m_0(H)}. \end{cases}$$

Remark 1.4. When proving the above theorem, we do not really require that G is random, but rather that it possesses some pseudo-random properties, which occur with high probability in $G(n,p)$.

The complementary case $m_0(H) < 1$ is in fact quite different, and we were able to find the threshold in that case for random walks on the complete graph K_n only. We will discuss potential difficulties in this aspect in Section 4. Denote by $\mathrm{odd}(G)$ the number of odd degree vertices in G.

Theorem 1.5. *Let T be a fixed tree on at least 2 vertices with $\mathrm{odd}(T) = \theta$. Let Γ_t be the trace of a random walk of length t on K_n. Then*

$$\lim_{n\to\infty} \mathbb{P}\left(T \subseteq \Gamma_t\right) = \begin{cases} 0 & t \ll n^{1-2/\theta} \\ 1 & t \gg n^{1-2/\theta}. \end{cases}$$

In particular, the theorem implies that the probability that the trace contains a fixed path (the case $\theta = 2$) as a subgraph is $1 - o(1)$ if $t \gg 1$. The corollary below follows easily from Theorem 1.5.

Corollary 1.6. *Let F be a non-empty fixed forest, and let $T_1, \ldots, T_z$ be its connected components. Let $\theta = \max_{i\in[z]} \{\mathrm{odd}(T_i)\}$. Let Γ_t be the trace of a random walk of length t on K_n. Then*

$$\lim_{n\to\infty} \mathbb{P}\left(F \subseteq \Gamma_t\right) = \begin{cases} 0 & t \ll n^{1-2/\theta} \\ 1 & t \gg n^{1-2/\theta}. \end{cases}$$

The overall proof strategy of Theorems 1.3 and 1.5 is to apply the first and the second moment methods. Our key lemma (Lemma 2.1) estimates the probability that the random walk on a random graph will traverse the edges of a *fixed* copy of a constant-sized graph H. We find that if $t \gg n$, the probability for the appearance of a copy in the trace is asymptotically equivalent to the probability of its appearance in a uniform random choice of a subgraph of $G(n,p)$ with t edges, and if $t \ll n$, it is determined by a structural property of H, namely, by the smallest number ρ for which H admits a trail decomposition with ρ parts. For the proof of the key lemma we use standard tools from Markov chain theory, and, in particular, a result about the mixing time of random graphs.

The rest of the paper is organized as follows. In Section 2 we state the key lemma and present some preliminary results to be used in its proof. The lemma itself is proved in Section 2.1, and in Section 2.2 we use it to prove Theorem 1.3. Section 3 contains the proofs of Theorem 1.5 and Corollary 1.6. Finally, in Section 4, we conclude with some remarks and open problems.

2 Walking on $G(n,p)$

Recall that a *walk* on G is a sequence of vertices $v_1, \ldots, v_t$ such that for $1 \leqslant i < t$, $\{v_i, v_{i+1}\}$ is an edge of G, and that a *trail* on G is a walk in which all of these edges are distinct. Denote by $\rho(G)$ the *trail decomposition number* of G, that is, the minimum number of edge-disjoint trails in G whose union is the edge set of G.

We begin with a key lemma. In what follows, we use $\mathbb{P}$ to denote the probability given that the initial distribution of the walk is uniform, and $\mathbb{P}_\mu$ to denote the probability given that the initial distribution is μ.

Lemma 2.1. *Let $\varepsilon, \gamma > 0$, $p \geqslant n^{-1+\varepsilon}$, $G \sim G(n,p)$ and $p^{-1} \ll t = O\left(n^{2-\gamma} p\right)$. Let H be a fixed graph with $\ell \geqslant 1$ edges and $\rho\left(H\right) = \rho$. Then,* ***whp*** *(over the distribution of G), for each fixed copy H_0 of H in G,*

$$\mathbb{P}\left(H_0 \subseteq \Gamma_t \mid G\right) = \Theta\left((np)^{-\ell} \sum_{r=\rho}^{\ell} \left(\frac{t}{n}\right)^r\right).$$

Moreover, if $t \gg n$, then

$$\mathbb{P}\left(H_0 \subseteq \Gamma_t \mid G\right) = \left(\frac{2t}{n^2 p}\right)^{\ell} (1 + o(1)).$$

The assumption that $p^{-1} \ll t = O\left(n^{2-\gamma} p\right)$ in the statement of the lemma is artificial. The upper bound on t is essential for proving (in Lemma 2.6) that the random walk traverses all edges at most a constant number of times with very high probability – a fact which is clearly not true for every t. The lower bound on t is used to show that it is "too expensive" for the walk to traverse an edge of H_0 more than once (see (8)). As we will see later, these bounds on t do not affect the proofs of our main theorems.

Before proving the lemma, we state a simple corollary.

Corollary 2.2. *Let H be a fixed graph with k vertices, $\ell \geqslant 1$ edges, $m_0(H) = m_0$ and $\rho(H) = \rho$. Let $\varepsilon, \gamma > 0$, $\nu = \max\{m_0, 1\}$, $p \geqslant n^{-1/\nu+\varepsilon}$, $G \sim G(n,p)$ and $p^{-1} \ll t = O\left(n^{2-\gamma} p\right)$. Finally, let Z be a random variable counting the number of copies of H in Γ_t (where multiple edges are ignored). Then,* ***whp*** *(over the distribution of G),*

$$\mathbb{E}\left(Z \mid G\right) = \Theta\left(n^{k-\ell} \sum_{r=\rho}^{\ell} \left(\frac{t}{n}\right)^r\right).$$

Proof (of the corollary). Since $p \geqslant n^{-1/\nu+\varepsilon} \gg n^{-1/m_0}$, the number of copies of H in G is **whp** asymptotically equal to its expectation (see for example [9, Remark 3.7]) which is $\Theta\left(n^k p^\ell\right)$. The result then follows from Lemma 2.1 and the linearity of expectation. □

Our goal now is to prove Lemma 2.1. In what follows, $\varepsilon, \gamma > 0$ are fixed constants, $p \geqslant n^{-1+\varepsilon}$, $G \sim G(n,p)$, $X_0, X_1, \ldots, X_t$ is a (lazy, simple) random walk on G starting at a uniformly chosen vertex, Γ_t is its trace and $p^{-1} \ll t = O\left(n^{2-\gamma}p\right)$. The *transition probability* of X from u to v is the probability

$$p_{uv} = \mathbb{P}\left(X_{t+1} = v \mid X_t = u\right) = \mathbb{P}\left(X_1 = v \mid X_0 = u\right),$$

and for an integer $s \geqslant 0$ we denote

$$p_{uv}^s = \mathbb{P}\left(X_{t+s} = v \mid X_t = u\right) = \mathbb{P}\left(X_s = v \mid X_0 = u\right).$$

Since, as is well known, G is **whp** connected, the sequence X forms an irreducible Markov chain, hence it has a unique stationary distribution given by (see, e.g., [12, Section 1.5])

$$\pi_v = \frac{d(v)}{\sum_{u\in[n]} d(u)} = \frac{d(v)}{2\left|E\right|}.$$

The following lemma about the degree distribution in $G(n,p)$ can easily be proved using standard estimates for the tail of the binomial distribution.

Lemma 2.3. *With high probability, $d(v) \sim np$, and thus $\pi_v \sim n^{-1}$, for every $v \in [n]$.*

We will use the fact that the random walk on $G(n,p)$ "mixes well". Roughly speaking, this means that the walk quickly forgets its starting point, and the distribution of its location quickly approaches stationarity. Recall that the *total variation distance* between the distribution of X_t and the stationary distribution is

$$d_{\mathrm{TV}}(X_t, \pi) = \frac{1}{2}\sum_{v\in[n]} \left|\mathbb{P}\left(X_t = v\right) - \pi_v\right|.$$

In [8], Hildebrand showed[1] that there exists a constant $s = s(\varepsilon)$ for which, **whp** (and regardless of the starting distribution),

$$d_{\mathrm{TV}}(X_s, \pi) < 1/e.$$

It follows (see, e.g., [12, Section 4.5]) that for an integer $\ell > 0$,

$$d_{\mathrm{TV}}(X_{\ell s}, \pi) < (2/e)^\ell.$$

We therefore obtain the following.

[1] Hildebrand shows this for a non-lazy random walk. However, as the probability that the lazy walk stays put at least once in a walk of fixed length is $o(1)$, we may ignore this difference here.

Claim 2.4. *For every $x > 0$ there exists $B = B(\varepsilon, x) = O(\ln n)$ such that **whp***

$$d_{\mathrm{TV}}(X_B, \pi) = o(n^{-x}).$$

Let x be a large positive constant to be determined later. Say that a vertex distribution π' is *almost stationary* if $d_{\mathrm{TV}}(\pi', \pi) = o\left(n^{-x}\right)$. The last corollary practically means that regardless of the starting distribution, after B steps, say, the distribution of the walk is almost stationary.

For a vertex v, let $\mathbf{n}_v$ be the uniform distribution over $N(v)$, and for $s > 0$ denote by $\eta(v, s)$ the number of exits the walk has made from vertex v by time s. Formally,

$$\eta(v, s) = \left|\{i \in [s] \mid X_{i-1} = v,\ X_i \neq v\}\right|.$$

A key observation is that typically no vertex is visited too many times, hence no edge is traversed too many times. This is stated in the following two lemmas.

Lemma 2.5. *For every $\alpha > 0$ there exists $\gamma' > 0$ such that **whp** (over the distribution of G), the probability that the random walk (of length t) visits at least one of the vertices more than $n^{1-\gamma'}p$ times is $o\left(n^{-\alpha}\right)$.*

Proof. First note that we may assume that $\gamma \leqslant \varepsilon$; otherwise, let $t_\varepsilon = n^{2-\varepsilon}p \gg n^{2-\gamma}p = \Omega(t)$. We can now prove the lemma for a walk of length t_ε, and conclude that the result holds for the walk of length t.

Fix $v \in [n]$ and let $s = n^{1-\gamma/2}$. Observe that in order to exit v, starting at a vertex which is not v, the walk must first enter it, and in view of Lemma 2.3 the probability for that to happen at any given step is $O\left(1/\left(np\right)\right)$. It follows that **whp** (over the distribution of G),

$$Q := \mathbb{P}_{\mathbf{n}_v}\left(\eta(v, B) \geqslant 1 \mid G\right) = O\left(\frac{B}{np}\right) = O\left(\frac{\ln n}{n^\varepsilon}\right) = o\left(n^{-\gamma/2}\right).$$

For an integer $a > 0$, let

$$P_\mu(a) := \mathbb{P}_\mu\left(\eta(v, s) \geqslant a \mid G\right).$$

Note that for an almost stationary distribution π', and for large enough x, by the union bound we have that **whp**

$$P_{\pi'}(1) \leqslant P_\pi(1) + o\left(n^{-x}\right) = O(s/n) = O\left(n^{-\gamma/2}\right),$$

and for $a > 1$, there exists an almost stationary distribution π'' for which

$$\begin{aligned} P_{\pi'}(a) \leqslant P_\pi(a) + o\left(n^{-x}\right) &\leqslant P_\pi(a-1)\left(Q + P_{\pi''}(1)\right) + o\left(n^{-x}\right) \\ &= P_\pi(a-1) \cdot O\left(n^{-\gamma/2}\right) + o\left(n^{-x}\right), \end{aligned}$$

as the probability of visiting v at least a times is at most the probability of visiting it $a-1$ times, and conditioning on that, the probability of visiting it once more, which is at most the probability of visiting it during the first B steps after exiting from it, plus the probability of visiting it at least once during s steps, starting from (another) almost stationary distribution π''. By induction, for $a > 2(\alpha+2)/\gamma$ and $x > a\gamma/2$,

$$P_{\pi'}(a) \leqslant P_\pi(1) \cdot O\left(n^{-(a-1)\gamma/2}\right) + o\left(n^{-x}\right) = O\left(n^{-a\gamma/2}\right) = o\left(n^{-\alpha-2}\right). \tag{1}$$

Now, let

$$L = \lceil t/(s+B) \rceil = O\left(n^{1-\gamma/2}p\right) = o(n).$$

Consider dividing $[t]$ into L segments of length at most s, with "buffers" of length B between them (and before the first). It follows from (1) and the union bound that (**whp** over the distribution of G) with probability $o\left(n^{-\alpha-1}\right)$ there exists a segment in which the walk exits v at least a times. Considering the possible visits in the buffers between the segments as well (at most BL such visits), we conclude that with probability $o\left(n^{-\alpha-1}\right)$ the walk exits v more than $n^{1-\gamma'}p$ times by time t, for $\gamma' = \gamma/3$, say. The union bound over all vertices yields the desired result. □

Lemma 2.6. *For every $\alpha > 0$ there exists $M > 0$ such that **whp** (over the distribution of G), the probability that the random walk (of length t) traverses at least one of the edges more than M times is $o\left(n^{-\alpha}\right)$.*

Proof. For a vertex v and integer $i \geqslant 0$, let $x_v^i \sim \mathbf{n}_v$, independently of each other. Think of the random walk X_t as follows. X_0 is sampled uniformly at random from V, and at each time $t \geqslant 0$, X_{t+1} is determined as follows: with probability $1/\left(d\left(X_t\right)+1\right)$ it equals X_t, and with the remaining probability it equals $x_{X_t}^{\eta(X_t,t)}$. We think of x_v^i as being sampled before the walk is performed, and the walk, when it exits v for the i'th time, simply reveals x_v^i[2].

Let (u,v) be a directed edge. Let x_{uv}^i be the indicator of the event $x_u^i = v$. The number of traversals of (u,v) during the first η exits from u is therefore (**whp**) the sum of η independent Bernoulli-distributed random variables with success probability (roughly) $1/(np)$. Thus, the probability that (u,v) was traversed at least M times during the first η exits from u equals the probability that a binomial random variable with η trials and success probability (roughly) $1/(np)$ is at least M. The probability that (u,v) was traversed at least M times is at most the probability that it was traversed at least M times during the first η exits from u in addition to the probability that the walk has exited u more than η times.

Thus, by the union bound, the probability that there exists (u,v) which was traversed at least M times by time t is at most

$$n^2 \cdot \mathbb{P}\left(\operatorname{Bin}\left(\eta, \frac{(1+o(1))}{np}\right) \geqslant M\right) + \mathbb{P}\left(\exists u :\ \eta(u,t) > \eta\right).$$

Choosing $\eta = 2n^{1-\gamma'}p$, with the right γ', Lemma 2.5 tells us that the second term is $o\left(n^{-\alpha}\right)$, and standard concentration results for the binomial distribution tell us that for large enough M the first term is $o\left(n^{-\alpha}\right)$, concluding the proof. □

[2] This is somewhat similar to the *list model* described in [2].

For a set $W \subseteq [t]$ denote by $r(W)$ the minimum number of integer intervals whose union is W. In symbols,

$$r(W) = |\{1 \leqslant i \leqslant t \mid i \in W \wedge i+1 \notin W\}| \,.$$

For W with $r(W) = r$ write

$$W = \{t_1, t_1+1, \ldots, t_1+a_1-1, t_2, t_2+1, \ldots, t_2+a_2-1, \ldots, t_r, t_r+1, \ldots, t_r+a_r-1\} \,,$$

where $t_i - 1 \notin W$ for $i \in [r]$ and $t_i + a_i < t_j$ for $1 \leqslant i < j \leqslant r$. If $t_{i+1} - (t_i + a_i) < 3B$, we say that the $(i+1)$'th run is *defective*, and we denote by

$$q(W) = |\{i \in [r-1] \mid t_{i+1} - (t_i + a_i) < 3B\}|$$

the number of defective runs in W. Let

$$\mathcal{W}_{w,r} = \{W \subseteq [t] \mid |W| = w, \ r(W) = r\} \,,$$

and

$$\mathcal{W}_{w,r,q} = \{W \subseteq [t] \mid |W| = w, \ r(W) = r, \ q(W) = q\} \,.$$

Claim 2.7. *For every $1 \leqslant r \leqslant w$,*

$$|\mathcal{W}_{w,r}| = \binom{w-1}{r-1}\binom{t-w+1}{r}.$$

Proof. For every $\mathbf{a} = (a_i)_{i=1}^r$ with $a_i > 0$ and $\sum_{i=1}^r a_i = w$, let $\mathcal{W}_{\mathbf{a}}$ be the set of W's in $\mathcal{W}_{w,r}$ with run lengths $a_1, \ldots, a_r$. The cardinality of $\mathcal{W}_{\mathbf{a}}$ is the number of ways to locate r runs with lengths $a_1, \ldots, a_r$ in $[t]$ so that any two distinct runs will be separated by at least 1. For every $\mathbf{a}$, this number is the number of integer solutions to the equation

$$\sum_{i=0}^{r} b_i = t - w, \qquad \begin{cases} b_0, b_r \geqslant 0 & \\ b_i \geqslant 1 & 1 \leqslant i \leqslant r-1, \end{cases}$$

where we think of b_0 as the space before the first run, b_r the space after the last run, and for $1 \leqslant i \leqslant r-1$, b_i is the space between the i'th run and the one following it. Thus

$$|\mathcal{W}_{\mathbf{a}}| = \binom{t-w+1}{r}.$$

Since the number of $\mathbf{a}$'s with $a_i > 0$ and $\sum_{i=1}^r a_i = w$ is the number of integer solutions to the equation

$$\sum_{i=1}^{r} a_i = w, \qquad \forall 1 \leqslant i \leqslant r, \ a_i > 0,$$

it follows that

$$|\mathcal{W}_{w,r}| = \binom{w-1}{r-1}\binom{t-w+1}{r}. \qquad \square$$

Lemma 2.8. *Let $K > 0$ be fixed, let $0 \leqslant q < r \leqslant w < K$ and suppose $t \gg 1$. Then,*

$$|\mathcal{W}_{w,r,q}| = O\left(B^q t^{r-q}\right).$$

Proof. Given a set $J \subseteq [r-1]$ with $|J| = q$, $I = [r-1] \smallsetminus J$ and $\mathbf{b} = (b_j)_{j \in J}$ with $1 \leqslant b_j < 3B$ for $j \in J$, let $A_{J,\mathbf{b}}$ be the set of $W \in \mathcal{W}_{w,r}$ for which for every $j \in J$, $t_{j+1} - (t_j + a_j) = b_j$. The cardinality of $A_{J,\mathbf{b}}$ is the number of solutions to the integer equation

$$b_0 + b_r + \sum_{i \in I} b_i = t - w - \sum_{j \in J} b_j, \qquad \begin{cases} b_0, b_r \geqslant 0 \\ b_i \geqslant 1 & i \in I, \end{cases}$$

which is clearly at most the number of integer solutions to the equation

$$b_0 + b_r + \sum_{i \in I} b_i = t, \qquad \begin{cases} b_0, b_r \geqslant 0 \\ b_i \geqslant 1 & i \in I. \end{cases}$$

It was shown in Claim 2.7 that $|\mathcal{W}_{w,r}| = \Theta(t^r)$. By a similar argument, $|A_{J,\mathbf{b}}| = O(t^{r-q})$. The union bound over all choices of J and $\mathbf{b}$ yields

$$|\mathcal{W}_{w,r,q}| \leqslant \binom{r-1}{q} (3B)^q \cdot O\left(t^{r-q}\right) = O\left(B^q t^{r-q}\right). \qquad \square$$

For $i \in [t]$ let $e_i = \{X_{i-1}, X_i\}$ and let $\vec{e}_i = (X_{i-1}, X_i)$. For a fixed subgraph H of G let $W(H) \subseteq [t]$ be the (random) set of times in which an edge from H had been traversed. That is,

$$W(H) = \{i \in [t] \mid e_i \in E(H)\}.$$

We are now ready to prove our key lemma.

2.1 Proof of Lemma 2.1

Let $\varepsilon, \gamma > 0$, $p \geqslant n^{-1+\varepsilon}$, $G \sim G(n,p)$ and $p^{-1} \ll t = O(n^{2-\gamma} p)$. As promised in Remark 1.4, we assume that G possesses the properties guaranteed **whp** by Lemmas 2.3 and 2.6 and Claim 2.4. Let H be a fixed graph with $\ell \geqslant 1$ edges, k vertices and $\rho(H) = \rho$, and let H_0 be a copy of H in G. Let A be the event $H_0 \subseteq \Gamma_t$, and for any $W \subseteq [t]$ let A_W be the event $A \wedge (W(H_0) = W)$. Our goal now is to estimate $\mathbb{P}(A)$.

Claim 2.9. *If $\mathbb{P}(A_W)$ is positive then*

- $\ell \leqslant |W| \leqslant t$,
- $1 \leqslant r(W) \leqslant |W|$,
- $0 \leqslant q(W) < r(W)$, *and*
- $r(W) \geqslant \ell + \rho - |W|$.

Proof. The only non-obvious claim is that $r(W) \geqslant \ell + \rho - |W|$. We will prove it by decomposing H_0 into at most $|W|+r(W)-\ell$ trails. Suppose $W_1, \ldots, W_r$ are the $r = r(W)$ runs of W, and let $w_1, \ldots, w_r$ be their lengths. Let ℓ_i be the number of edges of H_0 that were traversed by W_i but not by W_j for $j < i$. By removing from W_i every edge that was previously traversed by either W_i or by an earlier run, we create at most $1 + (w_i - \ell_i)$ edge-disjoint trails, which are disjoint to every trail created so far. At the end of this process we have created at most

$$\sum_{i=1}^{r}(1 + w_i - \ell_i) = r + |W| - \ell$$

edge-disjoint trails covering H. □

As a result of Claim 2.9, letting $r_w = \max\{1, \ell + \rho - w\}$, we have:

$$\mathbb{P}(A) = \sum_{w=\ell}^{t} \sum_{r=r_w}^{w} \sum_{q=0}^{r-1} \sum_{W \in \mathcal{W}_{w,r,q}} \mathbb{P}(A_W). \tag{2}$$

Upper bound

Let $M > 0$ be such that the probability that any edge was traversed at least M times is $o\left(n^{-3\ell}\right)$, as guaranteed by Lemma 2.6, and let $K = \ell M$. Write

$$\Lambda_{w,r,q} = \sum_{W \in \mathcal{W}_{w,r,q}} \mathbb{P}(A_W), \qquad \Lambda_{w,r} = \sum_{q=0}^{r-1} \Lambda_{w,r,q}, \qquad \Lambda^{+}_{w,r} = \Lambda_{w,r} - \Lambda_{w,r,0},$$

and

$$\Lambda_1 = \sum_{w=K}^{t} \sum_{r=r_w}^{w} \Lambda_{w,r}, \qquad \Lambda_2 = \sum_{w=\ell+1}^{K-1} \sum_{r=r_w}^{w} \Lambda_{w,r}, \qquad \Lambda_3 = \sum_{r=\rho}^{\ell} \Lambda_{\ell,r},$$

so, noting that $r_\ell = \rho$ it follows from (2) that

$$\mathbb{P}(A) = \Lambda_1 + \Lambda_2 + \Lambda_3. \tag{3}$$

Now, according to the choice of K, we have that

$$\Lambda_1 \ll n^{-3\ell} \ll (np)^{-\ell} \sum_{r=\rho}^{\ell} \left(\frac{t}{n}\right)^{r}. \tag{4}$$

Let $W \in \mathcal{W}_{w,r,q}$ with $w < K$. In these settings,

$$\mathbb{P}(A_W) \leqslant \mathbb{P}(W \subseteq W(H_0)) = O\left(n^{-r+q}(np)^{-w-q}\right), \tag{5}$$

as at the beginning of any non-defective run the probability that the walk will be at a vertex of H_0 is $\Theta(1/n)$ (and there are $r - q$ non-defective runs), at the beginning of any defective run the probability that the walk will be at a vertex of H_0 is $O\left(1/\left(np\right)\right)$, and at any time of W, the probability that the walk will traverse an edge of H_0 is $O\left(1/\left(np\right)\right)$.

If $w < K$, Lemma 2.8 states that $|\mathcal{W}_{w,r,q}| = O\left(B^q t^{r-q}\right)$, and therefore it follows from (5) and since $B \ll tp$, that

$$\Lambda^+_{w,r} = \sum_{q=1}^{r-1} O\left(B^q t^{r-q} n^{-r+q} \left(np\right)^{-w-q}\right)$$
$$= O\left(\left(np\right)^{-w} \left(\frac{t}{n}\right)^r \sum_{q=1}^{r-1} \left(\frac{B}{tp}\right)^q \right) \ll \left(np\right)^{-w} \left(\frac{t}{n}\right)^r, \tag{6}$$

and

$$\Lambda_{w,r,0} = O\left(\left(np\right)^{-w} \left(\frac{t}{n}\right)^r \right),$$

and therefore

$$\Lambda_{w,r} = O\left(\left(np\right)^{-w} \left(\frac{t}{n}\right)^r \right). \tag{7}$$

Suppose that $\ell < w < K$. If $t \geqslant n$ then, since $t \ll n^2 p$ and using (7),

$$\sum_{r=r_w}^{w} \Lambda_{w,r} = O\left(\left(np\right)^{-w} \left(\frac{t}{n}\right)^w \right) \ll \left(np\right)^{-\ell} \left(\frac{t}{n}\right)^\ell = \Theta\left(\left(np\right)^{-\ell} \sum_{r=\rho}^{\ell} \left(\frac{t}{n}\right)^r \right),$$

and if $t < n$ then, since $t \gg p^{-1}$ and using (7),

$$\sum_{r=r_w}^{w} \Lambda_{w,r} = O\left(\left(np\right)^{-w} \left(\frac{t}{n}\right)^{r_w} \right)$$
$$= O\left(\left(np\right)^{-\ell} \left(\frac{t}{n}\right)^\rho \cdot \left(\frac{t}{n}\right)^{\ell-w} \left(np\right)^{\ell-w} \right)$$
$$\ll \left(np\right)^{-\ell} \left(\frac{t}{n}\right)^\rho = \Theta\left(\left(np\right)^{-\ell} \sum_{r=\rho}^{\ell} \left(\frac{t}{n}\right)^r \right),$$

and therefore

$$\Lambda_2 \ll \left(np\right)^{-\ell} \sum_{r=\rho}^{\ell} \left(\frac{t}{n}\right)^r. \tag{8}$$

Finally, using (7),

$$\Lambda_3 = O\left(\left(np\right)^{-\ell} \sum_{r=\rho}^{\ell} \left(\frac{t}{n}\right)^r \right), \tag{9}$$

and therefore, using (3), (4), (8) and (9),

$$\mathbb{P}\left(A\right) = O\left(\left(np\right)^{-\ell} \sum_{r=\rho}^{\ell} \left(\frac{t}{n}\right)^r \right).$$

This concludes the proof of the upper bound of the first part of the lemma.

Lower bound

Let $\Gamma_W = \{\{X_{i-1}, X_i\} \mid i \in W\}$.

Claim 2.10. *For* $W \in \mathcal{W}_{\ell,r,0}$,

$$\mathbb{P}(A_W) \sim \mathbb{P}(H_0 \subseteq \Gamma_W).$$

Proof. First note that

$$\begin{aligned}\mathbb{P}(A_W) &= \mathbb{P}((W(H_0) = W) \wedge (H_0 \subseteq \Gamma_W)) \\ &= \mathbb{P}(W(H_0) \subseteq W \mid H_0 \subseteq \Gamma_W) \cdot \mathbb{P}(H_0 \subseteq \Gamma_W).\end{aligned}$$

Now, conditioning on $H_0 \subseteq \Gamma_W$, the probability that an edge of H_0 is ever traversed during times not in W, can be bounded from above as follows. Let

$$W_B = \{s \in [t] \mid \exists s' \in W,\ |s - s'| \leqslant B\}.$$

Let $\vec{e} = (u, v)$ be an arbitrary edge of H_0 with a direction assigned to it. Let $i \in [t] \setminus W_B$, and assume first that i is between two consecutive runs of W. Let i_0 be the maximal element in W with $i_0 < i$, and let i_1 be the minimal element in W with $i < i_1$. Write $s_0 = i - i_0$, $s_1 = i_1 - i$. Observing that for every two vertices v_1, v_2 and $s \geqslant B$ we have $p^s_{v_1 v_2} \sim n^{-1}$, we have that for every u_0, u_1,

$$\begin{aligned}\mathbb{P}(X_{i-1} = u \mid X_i = v,\ X_{i_0} = u_0) &= \frac{p^{s_0-1}_{u_0 u} p_{uv}}{\sum_{w \in N^+(v)} p^{s_0-1}_{u_0 w} p_{wv}} \\ &\sim \frac{p_{uv}}{\sum_{w \in N^+(v)} p_{wv}} \sim \frac{p_{vu}}{\sum_{w \in N^+(v)} p_{vw}} = p_{vu} \sim \frac{1}{np},\end{aligned}$$

and

$$\mathbb{P}(X_i = v \mid X_{i_0} = u_0,\ X_{i_1} = u_1) = \frac{p^{s_0}_{u_0 v} p^{s_1}_{v u_1}}{\sum_{w \in [n]} p^{s_0}_{u_0 w} p^{s_1}_{w u_1}} \sim \frac{1}{n},$$

thus

$$\begin{aligned}&\mathbb{P}(\vec{e}_i = \vec{e} \mid H_0 \subseteq \Gamma_W, X_{i_0} = u_0, X_{i_1} = u_1) \\ &= \mathbb{P}(X_{i-1} = u, X_i = v \mid X_{i_0} = u_0, X_{i_1} = u_1) \\ &= \mathbb{P}(X_{i-1} = u \mid X_i = v,\ X_{i_0} = u_0) \cdot \mathbb{P}(X_i = v \mid X_{i_0} = u_0,\ X_{i_1} = u_1) \sim \frac{1}{n^2 p}.\end{aligned}$$

Since this holds for every u_0, u_1, the probability that $i \in W(H_0)$ is $O(1/(n^2 p))$. Now let $i \in W_B \setminus W$, and let i_0, i_1 and s_0, s_1 be as before. Since $W \in \mathcal{W}_{\ell,r,0}$, $s_0 + s_1 \geqslant 3B$. Suppose first that $s_0 > B$. In that case,

$$\begin{aligned}\mathbb{P}(\vec{e}_i = \vec{e} \mid H_0 \subseteq \Gamma_W, X_{i_0} = u_0, X_{i_1} = u_1) &= \mathbb{P}(X_{i-1} = u, X_i = v \mid X_{i_0} = u_0, X_{i_1} = u_1) \\ &\leqslant \mathbb{P}(X_{i-1} = u \mid X_i = v,\ X_{i_0} = u_0) \sim \frac{1}{np}.\end{aligned}$$

If on the other hand $s_0 \leqslant B$ then $s_1 > B$ and we may use the reversibility of the walk to obtain a similar bound for $\mathbb{P}(\vec{e}_i = \vec{e} \mid H_0 \subseteq \Gamma_W)$, and therefore, since this holds for every u_0, u_1, the probability that $i \in W(H_0)$ is $O(1/np)$.

If $i < \min W$ (or $i > \max W$), letting i_1 (i_0, respectively) be as before, a similar argument, now conditioning only on the location of X at time i_1 (at time i_0, respectively), gives the same bounds.

Since $|W_B| = O(B)$, $B \ll np$ and $t \ll n^2p$ we have that

$$\begin{aligned}\mathbb{P}\left(W\left(H_0\right) \nsubseteq W \mid H_0 \subseteq \Gamma_W\right) &= \mathbb{P}\left(\exists i \notin W,\ i \in W(H_0) \mid H_0 \subseteq \Gamma_W\right) \\ &= O\left(B(np)^{-1} + t\left(n^2p\right)^{-1}\right) = o(1),\end{aligned}$$

and thus

$$\mathbb{P}\left(A_W\right) \sim \mathbb{P}\left(H_0 \subseteq \Gamma_W\right). \qquad \square$$

Now, let $W \in \mathcal{W}_{\ell,r,0}$ with $\rho \leqslant r \leqslant \ell$. In this case,

$$\mathbb{P}\left(H_0 \subseteq \Gamma_W\right) = \Omega\left((np)^{-\ell} n^{-r}\right).$$

This can be seen as follows. Let

$$f_1^1, \ldots, f_{\ell_1}^1, \ldots, f_1^r, \ldots, f_{\ell_r}^r$$

be a decomposition of the edges of H_0 into r trails (think of the edges f_i^j as directed edges, with the direction induced by the j'th trail), and write $f_i^j = (u_i^j, v_i^j)$. At the beginning of the j'th run of W (which is non-defective), the probability that the walk will be at u_1^j is $\Omega(1/n)$, and the i'th time in the j'th run of W, the probability that the traversed edge is f_i^j, given that the location of the walk before that move is u_i^j, is $\Omega(1/(np))$. Using Claim 2.10 we have that

$$\mathbb{P}\left(A_W\right) = \Omega\left((np)^{-\ell} n^{-r}\right).$$

Therefore,

$$\Lambda_{\ell,r} \geqslant \Lambda_{\ell,r,0} = \sum_{W \in \mathcal{W}_{\ell,r,0}} \mathbb{P}\left(A_W\right) = \Omega\left((np)^{-\ell}\left(\frac{t}{n}\right)^r\right),$$

and thus

$$\Lambda_3 = \Omega\left((np)^{-\ell}\sum_{r=\rho}^{\ell}\left(\frac{t}{n}\right)^r\right).$$

Using (3) we have that

$$\mathbb{P}(A) = \Omega\left((np)^{-\ell}\sum_{r=\rho}^{\ell}\left(\frac{t}{n}\right)^r\right).$$

This concludes the proof of the lower bound of the first part of the lemma.

The case $t \gg n$ In this case, according to (4),

$$\Lambda_1 \ll \left(\frac{t}{n^2p}\right)^{\ell}, \tag{10}$$

and according to (8),

$$\Lambda_2 \ll \left(\frac{t}{n^2p}\right)^{\ell}. \tag{11}$$

Let $W \in \mathcal{W}_{\ell,\ell,0}$. In this case we can give a more accurate estimate on $\mathbb{P}(A_W)$. There are $\ell!$ ways to order the edges of H_0 by their traversal times, and for each such ordering, as all the runs are non-defective and of length 1, the probability that the walk will traverse an edge at a prescribed time is approximately the inverse of the number of edges in G. Therefore, using Claim 2.10, we have that

$$\mathbb{P}(A_W) \sim \ell! \cdot \left(\frac{2}{n^2p}\right)^{\ell}.$$

According to Claim 2.7 and Lemma 2.8,

$$|\mathcal{W}_{\ell,\ell,0}| \sim \binom{\ell-1}{\ell-1}\binom{t-\ell+1}{\ell} = \binom{t-\ell+1}{\ell},$$

and thus

$$\Lambda_{\ell,\ell,0} \sim \binom{t-\ell+1}{\ell} \cdot \ell! \cdot \left(\frac{2}{n^2p}\right)^{\ell} \sim \left(\frac{2t}{n^2p}\right)^{\ell}.$$

It follows from (6) that

$$\Lambda^{+}_{\ell,\ell} \ll \left(\frac{t}{n^2p}\right)^{\ell},$$

hence

$$\Lambda_{\ell,\ell} = \Lambda_{\ell,\ell,0} + \Lambda^{+}_{\ell,\ell} \sim \left(\frac{2t}{n^2p}\right)^{\ell}.$$

Now suppose that $\rho \leqslant r < \ell$. It follows from (7) that

$$\Lambda_{\ell,r} = O\left((np)^{-\ell}\left(\frac{t}{n}\right)^{r}\right) \ll \left(\frac{t}{n^2p}\right)^{\ell},$$

thus

$$\Lambda_3 \sim \Lambda_{\ell,\ell} \sim \left(\frac{2t}{n^2p}\right)^{\ell}. \tag{12}$$

It follows from (3), together with (10), (11) and (12), that if $t \gg n$,

$$\mathbb{P}(A_W) \sim \left(\frac{2t}{n^2p}\right)^{\ell},$$

concluding the proof of the second part of the lemma. □

2.2 Proof of Theorem 1.3

Throughout this subsection H is a fixed graph with k vertices, ℓ edges and $m_0(H) = m_0 \geqslant 1$, $\varepsilon > 0$, $p \geqslant n^{-1/m_0+\varepsilon}$ and G is sampled according to $G(n,p)$.

2.2.1 Proof of the negative part

Assume $t \ll n^{2-1/m_0}$. Since $p^{-1} \leqslant n^{1/m_0-\varepsilon} \ll n \leqslant n^{2-1/m_0}$ we may assume without loss of generality that $t \gg p^{-1}$. In addition, letting $\gamma \leqslant \varepsilon$ we have that $t = O(n^{2-\gamma}p)$. Let $H' \subseteq H$ with k_0 vertices and ℓ_0 edges be such that $\ell_0/k_0 = m_0$, and write $\rho = \rho(H')$. Let Z, Z' count the number of appearances of a copy of H, H' in Γ_t, respectively. From Corollary 2.2 it follows that **whp**

$$\mathbb{E}\left(Z' \mid G\right) = O\left(n^{k_0-\ell_0}\sum_{r=\rho}^{\ell_0}\left(\frac{t}{n}\right)^r\right).$$

Now, if $m_0 = 1$ then $k_0 = \ell_0$ and $t \ll n$ and thus **whp** $\mathbb{E}\left(Z' \mid G\right) = o(1)$. If $m_0 > 1$ then $k_0 - \ell_0 \leqslant -1$; in that case, if $t < n$ then **whp** $\mathbb{E}\left(Z' \mid G\right) = O\left(n^{-1}\right) = o(1)$, and if $t \geqslant n$ we have that **whp**

$$\mathbb{E}\left(Z' \mid G\right) = O\left(n^{k_0-2\ell_0}t^{\ell_0}\right) = o\left(n^{k_0-2\ell_0}n^{2\ell_0-k_0}\right) = o(1).$$

Since the non-appearance of a copy of H' in Γ_t implies that of H, Markov's inequality yields the desired result. □

2.2.2 Proof of the positive part

Assume $t \gg n^{2-1/m_0} \geqslant n$. We also assume, without loss of generality, that $t = O\left(n^{2-\gamma}p\right)$ for sufficiently small $\gamma > 0$. For two graphs H_1, H_2 denote by $H_1 \cup H_2$ the graph whose vertex set is $V(H_1) \cup V(H_2)$ and whose edge set is $E(H_1) \cup E(H_2)$ (where multiple edges are ignored). If H_1, H_2 are not vertex-disjoint we say they *intersect* and denote it by $H_1 \sim H_2$.

Lemma 2.11. *Let H_1, H_2 be two intersecting labelled copies of H in G, and let $H^* = H_1 \cup H_2$. Let Z, Z^* count the number of appearances of a copy of H, H^* in Γ_t, respectively. Then,* ***whp****,*

$$\mathbb{E}\left(Z^* \mid G\right) \ll \mathbb{E}^2\left(Z \mid G\right).$$

Proof. According to Corollary 2.2, since $t \gg n$, **whp**

$$\mathbb{E}\left(Z \mid G\right) = \Theta\left(n^{k-2\ell}t^{\ell}\right),$$

and thus

$$\mathbb{E}^2\left(Z \mid G\right) = \Theta\left(n^{2k-4\ell}t^{2\ell}\right).$$

Let k', ℓ' be the number of vertices and edges in the intersection $H_1 \cap H_2$, respectively, and note that H^* has $2k - k'$ vertices and $2\ell - \ell'$ edges. We therefore have that, **whp**,

$$\mathbb{E}\left(Z^* \mid G\right) = \Theta\left(n^{(2k-k')-2(2\ell-\ell')}t^{2\ell-\ell'}\right) = \Theta\left(n^{2k-k'-4\ell+2\ell'}t^{2\ell-\ell'}\right),$$

and thus

$$\frac{\mathbb{E}^2\left(Z \mid G\right)}{\mathbb{E}\left(Z^* \mid G\right)} = \Theta\left(n^{k'-2\ell'}t^{\ell'}\right),$$

so, as H_1, H_2 are intersecting, either $\ell' = 0$ and $k' > 0$, in which case the above expression is $\omega(1)$, or $\ell' > 0$, in which case $t^{\ell'} \gg n^{2\ell'-\ell'/m_0}$ and the above expression is (since $m_0 \geqslant \ell'/k'$),

$$\omega\left(n^{k'-\ell'/m_0}\right) = \omega(1). \qquad \square$$

The following lemma shows that if two copies of H are not vertex-intersecting, then the events of their appearances in the trace are almost independent, in the sense that their covariance is very small.

Lemma 2.12. *Let H_1, H_2 be two vertex-disjoint labelled copies of H in G. Let A_i be the event "$H_i \subseteq \Gamma_t$", and let Z_i be its indicator, $i = 1, 2$. Then* ***whp***

$$\operatorname{Cov}(Z_i, Z_j \mid G) = o\left(t^{2\ell} n^{-4\ell} p^{-2\ell}\right).$$

Proof. According to Lemma 2.1 and since $t \gg n$, **whp**,

$$\mathbb{P}(A_i \mid G) \sim (2t)^{\ell} \left(n^2 p\right)^{-\ell},$$

and, since H_1, H_2 are vertex disjoint,

$$\mathbb{P}(A_1 \wedge A_2 \mid G) \sim (2t)^{2\ell} \left(n^2 p\right)^{-2\ell},$$

and finally

$$\mathbb{P}(A_1 \mid G) \cdot \mathbb{P}(A_2 \mid G) = \mathbb{P}^2(A_i \mid G) \sim (2t)^{2\ell} \left(n^2 p\right)^{-2\ell},$$

thus

$$\operatorname{Cov}(Z_i, Z_j \mid G) = o\left(t^{2\ell} n^{-4\ell} p^{-2\ell}\right). \qquad \square$$

We now employ the second moment method to prove the positive part of the theorem.

Proof of the positive part of Theorem 1.3. Let Z count the number of copies of H in Γ_t. Recall (e.g. from the proof of Lemma 2.11) that **whp**

$$\mathbb{E}(Z \mid G) = \Theta\left(n^{k-2\ell} t^{\ell}\right),$$

which is $\omega(1)$, since $m_0 \geqslant \ell/k$.

Let Y denote the number of copies of H in G, and recall that **whp** $Y \sim \mathbb{E}(Y)$. Let $\mathcal{H} = \{H_1, H_2, \ldots H_Y\}$ be the set of all copies of H in G, let Z_i be the indicator of the event "$H_i \subseteq \Gamma_t$", let $\mathcal{U}$ be the set of all possible unions of two intersecting (distinct) copies of H, and for $H^* \in \mathcal{U}$, let Z_{H^*} be the random variable counting the number of copies of H^* in Γ_t.

Write $i \sim j$ if $H_i \sim H_j$, and $i \nsim j$ otherwise. Since $|\mathcal{U}| = O(1)$, and using Lemma 2.12, it follows that, **whp**,

$$\operatorname{Var}(Z \mid G) = \sum_{i=1}^{Y} \sum_{j=1}^{Y} \operatorname{Cov}(Z_i, Z_j \mid G)$$

$$= \sum_{i=1}^{Y} \mathrm{Var}\left(Z_i \mid G\right) + \sum_{i \sim j} \mathrm{Cov}\left(Z_i, Z_j \mid G\right) + \sum_{i \nsim j} \mathrm{Cov}\left(Z_i, Z_j \mid G\right)$$
$$\leqslant \sum_{i=1}^{Y} \mathbb{E}\left(Z_i \mid G\right) + \sum_{H_i \sim H_j} \mathbb{P}\left(H_i \cup H_j \subseteq \Gamma_t \mid G\right) + o\left(n^{2k} p^{2\ell} \cdot t^{2\ell} n^{-4\ell} p^{-2\ell}\right)$$
$$= \mathbb{E}\left(Z \mid G\right) + 2 \sum_{H^* \in \mathcal{U}} \mathbb{E}\left(Z_{H^*}\right) + o\left(\mathbb{E}^2\left(Z \mid G\right)\right) = o\left(\mathbb{E}^2\left(Z \mid G\right)\right).$$

Chebyshev's inequality then yields the desired result. □

3 Walking on K_n, traversing trees

Recall that $\rho(G)$ denotes the minimum number of edge-disjoint trails in G whose union is the edge set of G. In order to prove Theorem 1.5, we will prove the following theorem instead.

Theorem 3.1. *Let T be a fixed tree on at least 2 vertices with $\rho(T) = \rho$. Let Γ_t be the trace of a random walk of length t on K_n. Then,*

$$\lim_{n \to \infty} \mathbb{P}\left(T \subseteq \Gamma_t\right) = \begin{cases} 0 & t \ll n^{1-1/\rho} \\ 1 & t \gg n^{1-1/\rho}. \end{cases}$$

The following lemma shows that Theorems 1.5 and 3.1 are in fact equivalent.

Lemma 3.2. *For every connected G, $\rho(G) = \max\{\mathrm{odd}(G)/2, 1\}$.*

Proof. If $\mathrm{odd}(G) = 0$ then G is Eulerian, thus $\rho(G) = 1$. Otherwise, let $\mathrm{odd}(G) = 2k$, and let $v_1, v_2, \ldots, v_{2k}$ be the odd degree vertices. Create G' by adding the edges $\{v_{2i-1}, v_{2i}\}$. G' is Eulerian; consider a tour (closed trail) T in G', and remove the added edges from that tour. That creates exactly k trails which make a partition of $E(G)$, thus $\rho(G) \leqslant k$. On the other hand, every trail removed from $E(G)$ decreases $\mathrm{odd}(G)$ by at most 2, hence $\rho(G) \geqslant k$. □

3.1 Proof of Theorem 3.1

Throughout this section T is a fixed non-empty tree with k vertices, $\ell = k - 1$ edges and $\rho(T) = \rho$.

3.1.1 Proof of the negative part

Assume $1 \ll t \ll n^{1-1/\rho}$. Let Z count the number of copies of T in Γ_t. According to Corollary 2.2,

$$\mathbb{E}\left(Z\right) = \Theta\left(n \sum_{r=\rho}^{k-1} \left(\frac{t}{n}\right)^r\right).$$

Since $t \ll n$, we have that

$$\mathbb{E}(Z) = \Theta\left(n\left(\frac{t}{n}\right)^{\rho}\right) = \Theta\left(n^{1-\rho}t^{\rho}\right) = o\left(n^{1-\rho}n^{\rho-1}\right) = o(1).$$

Markov's inequality then yields the result. □

3.1.2 Proof of the positive part

We will need a couple of lemmas in order to prove the positive part of the theorem.

Lemma 3.3. *Let $T_1 \subseteq T_2$ be two trees. Then $\rho(T_1) \leqslant \rho(T_2)$.*

Note. The above lemma does not hold for T_1, T_2 which are not trees. For example, the star S_3 with three leaves has $\rho(S_3) = 2$, but if $G = S_3 + e$ for any edge e in the complement of S_3, then $\rho(G) = 1$. Similarly, the path P_3 of length 3 has $\rho(P_3) = 1$, but $G = P_3 - e$ where e is the middle edge, is a forest with $\rho(G) = 2$.

Proof. It suffices to show that every trail in T_2, restricted to the edges of T_1, is a trail in T_1. Let P be a trail in T_2. Since T_2 is a tree, P is a path. Suppose to the contrary that the restriction of P to the edges of T_1, P', is not a path. Thus, it must have at least two connected components. Let u_1 and v_1 be two vertices of P' which belong to two distinct connected components. Thus in T_2 there are two distinct paths from u_1 to v_1, one which passes through P and one which passes through T_1, in contradiction to the fact that T_2 is a tree. □

Alternative proof. In view of Lemma 3.2 it suffices to show that $\text{odd}(T_1) \leqslant \text{odd}(T_2)$, and this can be verified by starting with T_1 and incrementally adding edges until reaching T_2, showing that each addition of an edge may not decrease the number of odd degree vertices. □

Lemma 3.4. *Let T_1, T_2 be two intersecting labelled copies of T in K_n. Let k' and ℓ' denote the number of vertices and edges, respectively, of the intersection $T_1 \cap T_2$, and let $\hat{\rho} = \rho(T_1 \cup T_2)$. Then*

$$k' - \ell' - 2 + \hat{\rho}/\rho \geqslant 0.$$

Proof. Observe that $T_1 \cap T_2$ is a forest. If it is not a tree, then $k' - \ell' \geqslant 2$ and the claim follows. Consider now the case where $T_1 \cap T_2$ is a tree. In that case, $k' - \ell' = 1$, thus it suffices to show that $\hat{\rho} \geqslant \rho$. Note that in that case it also follows that $T_1 \cup T_2$ is a tree, since it is connected with $2k - k'$ vertices and $2\ell - \ell'$ edges, and

$$(2k - k') - (2\ell - \ell') = 2(k - \ell) - (k' - \ell') = 1.$$

It follows that T is a subtree of $T_1 \cup T_2$, thus by Lemma 3.3, $\rho \leqslant \hat{\rho}$. □

In what follows, assume $n^{1-1/\rho} \ll t$. We also assume without loss of generality that $t \ll n$. The following lemma is the equivalent of Lemma 2.11 for the case of traversing trees.

Lemma 3.5. *Let T_1, T_2 be two intersecting labelled copies of T in K_n, and let $T^* = T_1 \cup T_2$. Let Z, Z^* count the number of appearances of a copy of T, T^* in Γ_t, respectively. Then*

$$\mathbb{E}(Z^*) \ll \mathbb{E}^2(Z).$$

Proof. According to Corollary 2.2 and since $n^{-1/\rho} \ll t/n \ll 1$, we have that

$$\mathbb{E}(Z) = \Theta\left(n\left(\frac{t}{n}\right)^{\rho}\right) = \omega(1),$$

and thus

$$\mathbb{E}^2(Z) = \Theta\left(n^{2-2\rho}t^{2\rho}\right).$$

Write $\hat{\rho} = \rho(T^*)$. Let k', ℓ' be the number of vertices and edges of the intersection $T_1 \cap T_2$, respectively. Since $T_1 \cap T_2$ is a non-empty forest, $k' > \ell'$. From Corollary 2.2, and since $t/n \ll 1$, we have that

$$\mathbb{E}(Z^*) = \Theta\left(n^{2k-k'-(2\ell-\ell')}\left(\frac{t}{n}\right)^{\hat{\rho}}\right) = \Theta\left(n^{2+\ell'-k'-\hat{\rho}}t^{\hat{\rho}}\right).$$

Now, if $\hat{\rho} \geqslant 2\rho$, then $(t/n)^{2\rho-\hat{\rho}} = \Omega(1)$ and

$$\frac{\mathbb{E}^2(Z)}{\mathbb{E}(Z^*)} = \Theta\left(n^{k'-\ell'+\hat{\rho}-2\rho}t^{2\rho-\hat{\rho}}\right) = \Omega\left(n^{k'-\ell'}\right) = \omega(1).$$

On the other hand, if $\hat{\rho} < 2\rho$, then $(t/n)^{2\rho-\hat{\rho}} \gg n^{\hat{\rho}/\rho-2}$ and

$$\frac{\mathbb{E}^2(Z)}{\mathbb{E}(Z^*)} = \Theta\left(n^{k'-\ell'+\hat{\rho}-2\rho}t^{2\rho-\hat{\rho}}\right) = \omega\left(n^{k'-\ell'-2+\hat{\rho}/\rho}\right),$$

and it follows from Lemma 3.4 that the last term is $\omega(1)$. □

Our next goal is to show that the events of the appearances of two vertex-disjoint graphs in the trace are not positively correlated. To that aim, we use a correlation inequality proved in [13]. For finite non-empty sets T and V, say that a collection $\mathcal{F}$ of families $(W_v)_{v\in V}$ of subsets of T is *decreasing* if for every family $(W_v)_{v\in V} \in \mathcal{F}$, if $(W'_v)_{v\in V}$ satisfies $W'_v \subseteq W_v$ for every $v \in V$, then $(W'_v)_{v\in V} \in \mathcal{F}$.

Lemma 3.6 ([13, Section 2])**.** *Let T and I be finite non-empty sets. Let I be partitioned into two non-empty sets J and K. Let $\mathcal{F}$ be a decreasing collection of families $(W_v)_{v\in J}$ and let $\mathcal{G}$ be a decreasing collection of families $(W_v)_{v\in K}$. Let $(x_j)_{j\in T}$ be a family of independent random variables, each taking values in some set containing I, and, for each $v \in I$, let $S_v = \{j \in T \mid x_j = v\}$. Let F be the event "$(S_v)_{v\in J} \in \mathcal{F}$" and let G be the event "$(S_v)_{v\in K} \in \mathcal{G}$". In these settings,*

$$\mathbb{P}(F \wedge G) \leqslant \mathbb{P}(F)\,\mathbb{P}(G).$$

Corollary 3.7. *Let H_1, H_2 be two vertex-disjoint subgraphs of K_n. For $i \in [2]$, let A_i be the event "$H_i \subseteq \Gamma_t$". Then A_1, A_2 are not positively correlated.*

Proof. It is easy to verify that if two events are not positively correlated then neither are their complements. It therefore suffices to prove that the complements B_1, B_2 of A_1, A_2 are not positively correlated. For $i \in [2]$ let $H_i = (V_i, E_i)$. We say that a family $(W_v)_{v \in V_i}$ of sets of times in $\{0, 1, \ldots, t\}$ *misses an edge* $\{u, v\} \in E_i$ if there is no $j \in [t]$ such that either $j - 1 \in W_u$ and $j \in W_v$ or $j - 1 \in W_v$ and $j \in W_u$. Let $\mathcal{F}, \mathcal{G}$ be the collections of all families of sets of times which miss at least one edge from E_1, E_2, respectively, and observe that $\mathcal{F}, \mathcal{G}$ are decreasing.

For $v \in V$, let S_v be the (random) set of times at which the walk was located at v. We can now write B_1, B_2 as the events "$(S_v)_{v \in V_i} \in \mathcal{F}$", "$(S_v)_{v \in V_2} \in \mathcal{G}$", respectively. Since $X_0, \ldots, X_t$ are independent, it follows from Lemma 3.6 (with $J = V_1$, $K = V_2$, $T = \{0, \ldots, t\}$ and $x_j = X_j$) that $\mathbb{P}(B_1 \wedge B_2) \leqslant \mathbb{P}(B_1)\mathbb{P}(B_2)$. □

Proof of the positive part of Theorem 3.1. Recall that $n^{1-1/\rho} \ll t \ll n$. Let Z count the number of copies of T in Γ_t. Recall (e.g. from the proof of Lemma 3.5) that

$$\mathbb{E}(Z) = \Theta\left(n\left(\frac{t}{n}\right)^{\rho}\right) = \omega(1).$$

Let $\mathcal{T} = \{T_1, T_2, \ldots, T_y\}$ be the set of all copies of T in K_n, let Z_i be the indicator of the event "$T_i \subseteq \Gamma_t$", let $\mathcal{U}$ be the set of all possible unions of two intersecting (distinct) copies of T, and for $H \in \mathcal{U}$, let Z_H be the random variable counting the number of copies of H in Γ_t. Write $i \sim j$ if Z_i and Z_j are positively correlated, and recall (from Corollary 3.7) that if $i \sim j$ then $T_i \sim T_j$ (that is, T_i, T_j intersect). It follows that

$$\begin{aligned}
\mathrm{Var}(Z) &= \sum_{i=1}^{y}\sum_{j=1}^{y} \mathrm{Cov}(Z_i, Z_j) \\
&\leqslant \sum_{i=1}^{y} \mathbb{E}(Z_i) + \sum_{i \sim j} \mathbb{P}(T_i \cup T_j \subseteq \Gamma_t) \\
&\leqslant \mathbb{E}(Z) + \sum_{T_i \sim T_j} \mathbb{P}(T_i \cup T_j \subseteq \Gamma_t) = \mathbb{E}(Z) + 2\sum_{H \in \mathcal{U}} \mathbb{E}(Z_H).
\end{aligned}$$

Since $|\mathcal{U}| = O(1)$, it follows from Lemma 3.5 that $\mathrm{Var}(Z) = o(\mathbb{E}^2(Z))$, and thus from Chebyshev's inequality it follows that $Z > 0$ **whp**. □

3.2 Proof of Corollary 1.6

Suppose first that $t \ll n^{1-2/\theta}$. Let $i \in [z]$ such that $\mathrm{odd}(T_i) = \theta$. By Theorem 1.5, **whp** T_i is not a subgraph of Γ_t, and hence F is not a subgraph of Γ_t.

Now suppose that $t \gg n^{1-2/\theta}$. We assume without loss of generality that $t \ll n$. Let $s = \lfloor t/z \rfloor$, and for $i \in [z]$ let Γ_i be the trace restricted to the times $[(i-1)s, is-1)$. For $i \in [z]$, let A_i be the event "$T_i \subseteq \Gamma_i$", and let T_i' be the first copy of T_i in Γ_i (if there exists one; let it be an arbitrary tree otherwise). Note that the events A_i are mutually independent. Let

$$U_i = \bigcup_{1 \leqslant j < i} V\left(T'_j\right),$$

let B_i be the event that an edge from Γ_i intersects U_i, and let $C_i = A_i \wedge \overline{B_i}$. Observe that for $U \subseteq [n]$ with $|U| = O(1)$, the probability that an edge from Γ_i intersects U is $O(s|U|/n) = o(1)$. It follows, using Theorem 1.5, that conditioning on $C_1, \ldots, C_{i-1}$, the probability of C_i is $1 - o(1)$, and therefore, **whp**, the trace contains vertex-disjoint copies $T'_1, \ldots, T'_z$ of $T_1, \ldots, T_z$, hence it contains a copy of F. □

4 Concluding remarks and open problems

Our results give another confirmation to the assertion that random walks which are long enough to typically cover a random graph, which is itself dense enough to be typically connected, leave a trace which "behaves" much like a random graph with a similar density. On the other hand, at least on the complete graph, the results suggest that if the random walk is of sublinear length then it leaves a trace which is very different from a random graph with similar edge density. In what other aspects do the two models differ?

In Theorem 1.5 we have found, in particular, that a fixed path P appears in the trace of a random walk on the complete graph **whp** as long as $t \gg 1$. In fact, it is not difficult to show that if P is a path of length $\ell \ll \sqrt{n}$ and $t \geqslant \ell$, then Γ_t contains a copy of P **whp**. This is true since a random walk of length $t \ll \sqrt{n}$ typically does not intersect itself. It may be interesting to find thresholds for the appearance of other "large" trees. It may also be interesting to find the threshold for the appearance of forests in the trace of a random walk on a random graph. Is it true, for example, that if $p \geqslant n^{-1+\varepsilon}$ for some $\varepsilon > 0$ then the thresholds are the same as in the case of $p = 1$? A slight variation in the proof of Lemma 3.5 works for random graphs as well, as long as $\varepsilon \geqslant 1/\rho$, but our use of Lemma 3.6 already assumes that the locations of the random walk are independent of each other.

Another possible direction would be to study the trace of the walk on other expander graphs, such as (n, d, λ)-graphs (see [11] for a survey), or on other random graphs, such as random regular graphs. The small subgraph problem for random regular graphs of growing degree was settled by Kim, Sudakov and Vu [10]. They have shown that the degree threshold for the appearance of a copy of H in a random regular graph is $n^{1-1/m_0(H)}$, as long as H contains a cycle. Is it true that for $d \geqslant n^{1-1/m_0(H)+\varepsilon}$, the time threshold for the appearance of H in the trace of a random walk on a random d-regular graph is also typically $n^{2-1/m_0(H)}$, as in Theorem 1.3?

Acknowledgements

The authors wish to thank Alan Frieze, Asaf Nachmias and Yinon Spinka for useful discussions, and two anonymous referees for their careful reading of the paper and valuable comments and suggestions.

References

[1] Noga Alon and Joel H. Spencer, **The probabilistic method**, 4th ed., Wiley Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., Hoboken, NJ, 2016. MR3524748 ↑2

[2] Ben Barber and Eoin Long, *Random walks on quasirandom graphs*, Electronic Journal of Combinatorics **20** (2013), no. 4, #P25. MR3158264 ↑2, 7

[3] Béla Bollobás, *Random graphs*, Combinatorics (Swansea, 1981), 1981, pp. 80–102. MR633650 ↑2

[4] Paul Erdős and Alfréd Rényi, *On random graphs. I*, Publicationes Mathematicae Debrecen **6** (1959), 290–297. MR0120167 ↑1

[5] Paul Erdős and Alfréd Rényi, *On the evolution of random graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **5** (1960), 17–61. MR0125031 ↑1

[6] Alan Frieze, Michael Krivelevich, Peleg Michaeli and Ron Peled, *On the trace of random walks on random graphs* (August 28, 2015). `arXiv:1508.07355`. ↑2

[7] Edgar N. Gilbert, *Random graphs*, Annals of Mathematical Statistics **30** (1959), 1141–1144. MR0108839 ↑1

[8] Martin Hildebrand, *Random walks on random simple graphs*, Random Structures & Algorithms **8** (1996), no. 4, 301–318. MR1603253 ↑5

[9] Svante Janson, Tomasz Łuczak and Andrzej Ruciński, **Random graphs**, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000. MR1782847 ↑2, 5

[10] Jeong Han Kim, Benny Sudakov and Van Vu, *Small subgraphs of random regular graphs*, Discrete Mathematics **307** (2007), no. 15, 1961–1967. MR2320201 ↑21

[11] Michael Krivelevich and Benny Sudakov, *Pseudo-random graphs*, More sets, graphs and numbers, 2006, pp. 199–262. MR2223394 ↑21

[12] David A. Levin, Yuval Peres and Elizabeth L. Wilmer, **Markov chains and mixing times**, American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson. MR2466937 ↑2, 5

[13] Colin McDiarmid, *On a correlation inequality of Farr*, Combinatorics, Probability and Computing **1** (1992), no. 2, 157–160. MR1179245 ↑19