Small subgraphs in the trace of a random walk

Michael Krivelevich * Peleg Michaeli †

May 15, 2016

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 $G(n,m)$.

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 $[n] = \{1,\ldots,n\}$, 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 [3, 4].

One of the problems studied in [4] 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 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 [2] in 1981, and the result can be stated as follows (see also [1, Section 4.4] or [8, Theorem 3.4]).

*School of Mathematical Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 6997801, Israel. E-mail: krivelev@post.tau.ac.il. Research supported in part by a USA-Israel BSF Grant and by a grant from Israel Science Foundation.

†School of Mathematical Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 6997801, Israel. E-mail: peleg.michaeli@math.tau.ac.il.
**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 $t \geq 0$, $X_{t+1}$ 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,

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

In [6], several results were given concerning graph-theoretic properties of the trace. 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 $m_0(H) \geq 1$ is equivalent to the condition of containing a cycle.

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

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

**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. Denote by $\text{odd}(G)$ the number of odd degree vertices in $G$.

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

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

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 $t \gg 1$. 

2
Corollary 1.6. Let $F$ be a non-empty fixed forest, and let $T_1, \ldots, T_z$ be its connected components. Let $\theta = \max_{i \in [z]} \{\text{odd}(T_i)\}$. Let $\Gamma_t$ be the trace of a random walk of length $t$ on $K_n$. Then,

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

The rest of the paper is organized as follows. In Section 2, we prove a key lemma which will be used in the proofs of the main theorems, and 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 $1 \leq i < t$, $\{v_i, v_{i+1}\}$ 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 minimum number of edge-disjoint trails in $G$ whose union is the edge set of $G$.

In what follows, we use $P$ to denote the probability given that the initial distribution of the walk is uniform, and $P_\mu$ to denote the probability given that the initial distribution is $\mu$.

We begin with a key lemma.

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

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

Moreover,

- if $p = 1$,

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

- if $t \gg n$,

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

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

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

Moreover,
\[ \mathbb{E}(Z \mid G) = \Theta\left(n^{k-\ell} \sum_{r=0}^{\ell} \left(\frac{t}{n}\right)^r \right) \]

- if \( p = 1 \),

- if \( t \gg n \),

\[ \mathbb{E}(Z \mid G) = \Theta\left(n^{k-\ell} \left(\frac{t}{n}\right)^\ell \right) \]

Proof (of the corollary). Since \( p \geq n^{-1/\nu} + \epsilon \gg n^{-1/m_0} \), the number of copies of \( H \) in \( G \) is \textit{whp} asymptotically equivalent to its expectation (see for example [8, Remark 3.7]) which is \( \Theta(n^k p^\ell) \).

The result then follows from Lemma 2.1 and the linearity of expectation. \( \square \)

Our goal now is to prove Lemma 2.1. In what follows, \( \varepsilon, \gamma > 0 \) are fixed constants, \( p \geq n^{-1+\varepsilon} \), \( G \sim G(n, p) \), \( X_0, X_1, \ldots, X_t \) is a (lazy, simple) random walk on \( G \) starting at a uniformly chosen vertex, \( \Gamma_t \) is its trace and \( t = O(n^{2-\gamma} p) \). Recall that the stationary distribution \( \pi \) of \( X \) is given by

\[ \pi_v = \frac{d(v)}{\sum_{u \in [n]} d(u)} = \frac{d(v)}{2|E|}, \]

and for every subset \( S \subseteq [n] \), denote

\[ \pi_S = \sum_{v \in S} \pi_v. \]

The following lemma about the edge distribution in \( G(n, p) \) can be easily proved using standard estimates for the tail of the binomial distribution. Let the \textit{edge boundary} of a vertex set \( S \) be the set of edges of \( G \) with exactly one endpoint in \( S \), and denote it by \( \partial S \). If \( f, g \) are functions of \( n \) we use the notation \( f \sim g \) to denote asymptotic equality. That is, \( f \sim g \) if and only if \( \lim_{n \to \infty} f/g = 1 \).

**Lemma 2.3.** With high probability, for every non-empty \( S \subseteq [n] \), \( |\partial S| \sim |S||S^c|p \). In particular, \textit{whp}, for every \( v \in [n] \), \( d(v) \sim np \).

**Corollary 2.4.** With high probability, the stationary distribution \( \pi \) of the random walk on \( G \) satisfies \( \pi_v \sim n^{-1} \) for every \( v \in [n] \). Thus, for every vertex set \( S \), \( \pi_S \sim |S|/n \).

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. To make things more formal, we need a couple of definitions. The \textit{total variation distance} between the distribution of \( X_t \) and the stationary distribution is

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

A parameter which measures the time required by the walk for the distance to stationarity to become small is the \textit{mixing time}, defined by

\[ \tau(\varepsilon) = \min\{t \geq 0 \mid \forall s \geq t, \ d_{TV}(X_s, \pi) < \varepsilon \}. \]
A theorem of Jerrum and Sinclair [9] implies that the mixing time of $X$ is indeed small. To use their bound we shall need the notion of conductance. The conductance of a cut $(S, S^c)$, with respect to $X$, is defined as

$$\varphi_X(S) = \frac{\sum_{v \in S, \ w \in S^c} \pi_v p_{vw}}{\min(\pi_S, \pi_{S^c})}.$$ 

In our case, according to Corollary 2.4,

$$\varphi_X(S) \sim \frac{|\partial S|}{\min\{|S|, |S^c|\} \cdot np}.$$ 

The conductance of $G$ is defined as

$$\Phi_X(G) = \min_{\sum_{v}^{\leq n} \pi_v \leq 1} \varphi_X(S).$$

**Claim 2.5.** With high probability (in $G(n, p)$), $\Phi_X(G) \sim 1/2$.

**Proof.** Recalling Lemma 2.3 we have that

$$\Phi_X(G) = \min_{\sum_{v}^{\leq n} \pi_v \leq 1/2} \varphi_X(S) = \min_{\sum_{v}^{\leq n} \pi_v \leq 1/2} \frac{|\partial S|}{\min\{|S|, |S^c|\} \cdot np} \sim \min_{\sum_{v}^{\leq n} \pi_v \leq 1/2} \frac{|S||S^c|p}{|S|np} \sim \frac{1}{2}.$$ 

\[\square\]

**Claim 2.6 ([6], Claim 2.8).** Let $\pi_{\min} = \min_v \pi_v$. For every $\varepsilon > 0$,

$$\tau(\varepsilon) \leq \frac{2}{\Phi_X(G)^2} \left(\ln\left(\frac{1}{\pi_{\min}}\right) + \ln\left(\frac{1}{2\varepsilon}\right)\right).$$

**Corollary 2.7.** For $\varepsilon > 0$, $\tau(\varepsilon) \leq 10 \ln(n/\varepsilon)$.

**Proof.** Plug Corollary 2.4 and Claim 2.5 into Claim 2.6. \[\square\]

Say that a vertex distribution $\pi'$ is almost stationary if $d_{TV}(\pi', \pi) = o(n^{-1})$. Set

$$B = 20 \ln n.$$ 

The last corollary practically means that regardless of the starting vertex, after $B$ steps, say, the distribution of the walk is almost stationary.

For a vertex $v$, let $n_v$ 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,

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

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 2.8.** For every $\ell > 0$ 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 $o(n^{-\ell})$. 

5
Proof. First note that we may assume that \( \gamma \leq \varepsilon \); 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 \( t_{\varepsilon} \), and conclude that the result holds for the walk of length \( t \).

Fix \( v \in [n] \) and let \( s = n^{1-\gamma/2} \). First note 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 \( O(1/(np)) \). It follows that \( \text{whp} \)

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

To estimate the probability that \( \eta(v, s) \geq a \) for \( a > 0 \), we observe that after exiting \( v \) for the first time, in order to exit it \( a - 1 \) more times the walk must either exit it again in the next \( B \) steps, or exit it at least \( a - 1 \) times after \( B \) steps. Formally, let

\[
P_{\mu}(a) = \mathbb{P}_{\mu}(\eta(v, s) \geq a \mid G).
\]

We therefore have that for an almost stationary distribution \( \pi' \),

\[
P_{\pi'}(1) \sim P_{\pi}(1) = O(s/n) = O\left( n^{-\gamma/2} \right);
\]

and for \( a > 1 \),

\[
P_{\pi'}(a) \leq P_{\pi'}(1) \cdot (Q + P_{\pi''}(a - 1)),
\]

for some other almost stationary distribution \( \pi'' \). Applying (1) recursively one obtains

\[
P_{\pi'}(a) \leq (1 + o(1)) \cdot \left( Q \sum_{i=1}^{a-1} P_{\pi}(1)^i + P_{\pi}(1)^a \right) = O\left( n^{-a\gamma/2} \right),
\]

which is \( o(n^{-(\ell+2)}) \), for large enough \( a = a(\gamma) \). Now, let \( L = \lfloor t/(s + B) \rfloor = \Omega(n^{1-\gamma/2} p) \).

Consider dividing \([t] \) into \( L \) segments of length at most \( s \), with “buffers” of length \( B \) between them. Noting that the distribution of the first vertex is also almost stationary, it follows from (2) that (conditioning on \( G \)) with probability \( o(n^{-(\ell+1)}) \) 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 \( o(n^{-(\ell+1)}) \) the walk exits \( v \) more than \( n^{1-\gamma'} p \) times by time \( t \), for some \( \gamma' > 0 \). The union bound over all vertices yields the desired result.

\( \square \)

Lemma 2.9. For every \( \ell > 0 \) there exists \( K > 0 \) such that \( \text{whp} \) (over the distribution of \( G \)), the probability that the random walk (of length \( t \)) traverses at least one of the edges more than \( K \) times is \( o(n^{-\ell}) \).

Proof. For a vertex \( v \) and integer \( i \geq 0 \), let \( x^i_v \sim n_v \), 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 \( t \geq 0 \), \( X_{t+1} \) is determined as follows: with probability \( 1/(d(X_t) + 1) \) it equals \( X_t \), and with the remaining probability it equals \( x^m_{X_t} \). We think of \( x^i_v \) as being sampled before the walk is performed, and the walk, when it exits \( v \) for the \( i \)th time, simply reveals \( x^i_v \).

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

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

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

Choosing $\eta = 2n^{1-\gamma'}p$, with the right $\gamma'$, Lemma 2.8 tells us that the second term is $o(n^{-\ell})$, and standard concentration results for the binomial distribution tell us that for large enough $K$ the first term is $o(n^{-\ell})$, concluding the proof.

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

$$r(W) = |\{1 \leq i \leq t \mid i \in W \land i + 1 \notin W\}|.$$

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

$$W = \{t_1, t_1 + 1, \ldots, t_1 + a_1 - 1, t_2, t_2 + 1, \ldots, t_2 + a_2 - 1, t_3, \ldots, t_r, t_r + 1, \ldots, t_r + a_r - 1\},$$

where $t_i - 1 \notin W$ for $i \in [r]$ and $t_i + a_i < t_j$ for $1 \leq i < j \leq r$. If $t_{i+1} - (t_i + a_i) < B$, we say that the $(i + 1)'th$ run is defective, and we denote by $q(W) = |\{i \in [r-1] \mid t_{i+1} - (t_i + a_i) < B\}|$ 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 2.10.** For every $1 \leq r \leq w$,

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

**Proof.** For every $a = (a_i)_{i=1}^r$ with $a_i > 0$ and $\sum_{i=1}^r a_i = w$, let $\mathcal{W}_a$ be the set of $W$’s in $\mathcal{W}_{w,r}$ with run lengths $a_1, \ldots, a_r$. The number of $W$’s in $\mathcal{W}_a$ 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 $a$, this number is the number of integer solutions to the equation

$$\sum_{i=0}^r b_i = t - w, \quad \begin{cases} b_0, b_r \geq 0 \\ b_i \geq 1 \quad 1 \leq i \leq r - 1, \end{cases}$$

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

$$|\mathcal{W}_a| = \binom{t - w + 1}{r}.$$
Since the number of a’s with $a_i > 0$ and $\sum_{i=1}^{r} a_i = w$ is the number of integer solutions to the equation
\[ \sum_{i=1}^{r} a_i = w, \quad \forall 1 \leq i \leq r, \quad a_i > 0, \]
it follows that
\[ |W_w,r| = \binom{w-1}{r-1} \binom{t-w+1}{r}. \]

**Lemma 2.11.** Let $K > 0$ be fixed, let $r \leq w \leq K$ and suppose $t \gg 1$. Sample $W$ uniformly from $W_{w,r}$. Then,
\[ \mathbb{P}(q(W) \geq q) = O((Bt^{-1})^q). \]

**Proof.** Given a set $J \subseteq [r-1]$ with $|J| = q$, $I = [r-1] \setminus J$ and $b = (b_j)_{j \in J}$ with $1 \leq b_j < B$ for $j \in [q]$, let $A_{J,b}$ be the set of $W \in W_{w,r}$ for which for every $j \in J$, $t_j + 1 - (t_j + a_j) = b_j$. The cardinality of $A_{J,b}$ is the number of solutions to the integer equation
\[ b_0 + b_r + \sum_{i \in I} b_i = t - w - \sum_{j \in J} b_j, \quad \begin{cases} b_0, b_r \geq 0 \\ b_i \geq 1 \end{cases} \quad i \in I, \]
which is clearly at most the number of integer solutions to the equation
\[ b_0 + b_r + \sum_{i \in I} b_i = t, \quad \begin{cases} b_0, b_r \geq 0 \\ b_i \geq 1 \end{cases} \quad i \in I. \]

It was shown in Claim 2.10 that $|W_{w,r}| = \Theta(t^r)$. By a similar argument, $|A_{J,b}| = \Theta(t^{r-q})$. The union bound over all choices of $J$ and $b$ yields
\[ \mathbb{P}(q(W) \geq q) \leq \left( \frac{r - 1}{q} \right) B^q \cdot \frac{\Theta(t^{r-q})}{\Theta(t^r)} = O((Bt^{-1})^q). \]

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

We are now ready to prove our key lemma.

**Proof of Lemma 2.1.** Let $\varepsilon, \gamma > 0$, $p \geq n^{-1+\varepsilon}$, $G \sim G(n,p)$ and $t = O(n^{2-\gamma}p)$. As promised in Remark 1.4, we assume that $G$ possesses the properties guaranteed w.h.p by Lemmas 2.3 and 2.9 and Corollaries 2.4 and 2.7. Let $H$ be a fixed graph with $\ell \geq 1$ edges, $k$ vertices and $\rho(H) = \rho$, and let $H_0$ be a copy of $H$ in $G$. Let $K > 0$ be such that the probability that any edge was traversed at least $\lfloor K/\ell \rfloor$ times is at most $n^{-3\varepsilon}$, as guaranteed by Lemma 2.9. Let $A$ be the event $H_0 \subseteq \Gamma_t$, and for any $W \subseteq [\ell]$ let $A_W$ be the event $A \wedge (W(H_0) = W)$. Note that for $W \neq W'$, the events $A_W$ and $A_{W'}$ are disjoint. If $\mathbb{P}(A_W) > 0$ we say that $W$ is feasible. Our
goal now is to estimate $\mathbb{P}(A)$. Let $W$ be with $|W| = w \leq K$, $r(W) = r \leq w$ and $q(W) = q < r$. In these settings,

$$\mathbb{P}(A_W) = O\left(n^{-r+q}(np)^{-w-q}\right) = O\left(n^{-r}(np)^{-w}\right), \quad (3)$$

as at the beginning of any non-defective run the probability that the walk will be at a vertex of $H_0$ is $O(1/n)$ (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 $O(1/(np))$, and at any time of $W$, the probability that the walk will traverse an edge of $H_0$ is $O(1/(np))$. It follows, using Claim 2.10, that

$$\sum_{W \in W_{w,r}} \mathbb{P}(A_W) = |W_{w,r}| \cdot O\left(n^{-r}(np)^{-w}\right) = O\left((np)^{-w}\left(\frac{t}{n}\right)^r\right). \quad (4)$$

We therefore have

$$\mathbb{P}(A) = \sum_{w=\ell}^K \sum_{r=\rho}^w \sum_{W \in W_{w,r}} \mathbb{P}(A_W)$$

$$= \sum_{w=\ell}^K \sum_{r=\rho}^w \sum_{W \in W_{w,r}} \mathbb{P}(A_W) + o\left(n^{-3\ell}\right)$$

$$= \sum_{w=\ell}^K \sum_{r=\rho}^w O\left((np)^{-w}\left(\frac{t}{n}\right)^r\right) + o\left(n^{-3\ell}\right).$$

Now note that since $np \gg 1$ and $np \gg t/n$, if $w > \ell$ then

$$(np)^{-\ell} \sum_{r=\rho}^\ell \left(\frac{t}{n}\right)^r \gg (np)^{-w} \sum_{r=\rho}^w \left(\frac{t}{n}\right)^r,$$

and

$$(np)^{-\ell} \sum_{r=\rho}^\ell \left(\frac{t}{n}\right)^r \geq n^{-\ell} \sum_{r=\rho}^\ell \left(\frac{1}{n}\right)^r \gg n^{-3\ell},$$

thus

$$\mathbb{P}(A) = O\left((np)^{-\ell} \sum_{r=\rho}^\ell \left(\frac{t}{n}\right)^r\right). \quad (5)$$

This concludes the first part of the lemma. We now proceed to prove the special cases where $p = 1$ and where $t \gg n$. First note that the probability that the walk traverses edges from $H_0$ during times not in $W$ (for $W \in W_{z,r}$) can be bounded from above by the probability that there exist a run, a vertex of $H_0$ and a time $i \in [B]$ such that the walk has visited that vertex $i$ steps after that run (and that probability is at most $\sim (np)^{-1}$), in addition to the probability that there exist a time in which the walk is mixed (that is, at least $B$ steps the end of the last run) and an edge of $H_0$ such that at this time, that edge was traversed and that probability is at most $\sim (n^2p)^{-1}$. All in all, the probability that the walk traverses edges from $H_0$ during times not in $W$ is at most

$$O\left(rkB(np)^{-1} + t\ell(n^2p)^{-1}\right) = o(1),$$

as $B \ll np$ and $t \ll n^2p$. 


The case $p = 1$. If $q = 0$ and $W$ is feasible,
\[
P(A_W) = \Theta(n^{-r-w}),
\]
as at the beginning of every run the probability that the walk will be at a vertex of $H_0$ is $\Theta(1/n)$, and at any time of $W$, the probability that the walk will traverse an edge of $H_0$ is $\Theta(1/n)$, given its location at one of the vertices of $H_0$. If $r \geq \rho$, the number of $W$’s in $W_{w,r,0}$ which are feasible is $\Omega(t^r)$, thus if $p = 1$,
\[
P(A) = \Omega\left(\sum_{r=\rho}^{\ell} \sum_{W \in W_{w,r,0}} P(A_W)\right) = \Omega\left(n^{-\ell} \sum_{r=\rho}^{\ell} \left(\frac{t}{n}\right)^r\right),
\]
and by (5),
\[
P(A) = \Theta\left(n^{-\ell} \sum_{r=\rho}^{\ell} \left(\frac{t}{n}\right)^r\right).
\]

The case $t \gg n$. If $q = 0$ and $r = w = \ell$, we can give a more accurate estimate on $P(A_W)$. Note that since the diameter of $G$ is whp constant (see for example [5, Chapter 7]), $W$ is feasible (it is possible to reach any vertex of $H_0$ within $B$ steps, starting from anywhere). As this is the case we have that
\[
P(A_W) \sim \ell! \cdot \left(\frac{2}{n^2 p}\right)^\ell,
\]
as there are $\ell!$ ways to order the edges of $H_0$ by their traversal times, and for each such ordering, the probability that the walk will traverse an edge at a prescribed time is roughly the inverse of the number of edges in $G$. Thus, using (3), Claim 2.10 and Lemma 2.11,
\[
\sum_{W \in W_{\ell,\ell}} P(A_W) = \sum_{q=0}^{\ell-1} \sum_{W \in W_{\ell,q}} P(A_W) = \sum_{W \in W_{\ell,0}} P(A_W) + O\left(|W_{\ell,0}| \cdot \sum_{q=1}^{\ell-1} P(q(W) \geq q) n^{-\ell+q} (np)^{-\ell-q}\right)
\]
\[
\sim (\ell - 1) \left(\frac{n}{\ell}\right)^{\ell - 1} \ell! \cdot \left(\frac{2}{n^2 p}\right)^\ell + O\left(\ell^\ell \cdot \sum_{q=1}^{\ell-1} (Bt^{-1}) q n^{-\ell+q} (np)^{-\ell-q}\right)
\]
\[
\sim \left(\frac{2t}{n^2 p}\right)^\ell + O\left(\left(\frac{t}{n^2 p}\right)^{\ell} \cdot \sum_{q=1}^{\ell-1} \left(\frac{B}{tp}\right)^q\right) \sim \left(\frac{2t}{n^2 p}\right)^\ell,
\]
as $B \ll tp$. Using (4),
\[
\sum_{r=\rho}^{\ell-1} \sum_{W \in W_{\ell,r}} P(A_W) = \sum_{r=\rho}^{\ell-1} O\left((np)^{-\ell} \left(\frac{t}{n}\right)^r\right) = O\left((np)^{-\ell} \left(\frac{t}{n}\right)^{\ell-1}\right),
\]
and using similar logic as in (5),
\[
\sum_{w=\ell+1}^{K} \sum_{r=\rho}^{w} P(A_W) = \sum_{w=\ell+1}^{K} \sum_{r=\rho}^{w} O\left((np)^{-w} \left(\frac{t}{n}\right)^r\right) = O\left(\left(\frac{t}{n}\right)^{\ell+1} (np)^{-(\ell+1)}\right).
\]
Combining everything, we have
\[
P(A) = \sum_{w=\ell}^{K} \sum_{r=\rho}^{w} \sum_{W\in W_{r,w}} \mathbb{P}(A_{W}) + o\left(n^{-3\ell}\right)
\]
\[
= \sum_{W\in W_{\ell,t}} \mathbb{P}(A_{W}) + \sum_{r=\rho}^{\ell-1} \sum_{W\in W_{r,r}} \mathbb{P}(A_{W}) + \sum_{w=\ell+1}^{K} \sum_{r=\rho}^{w} \sum_{W\in W_{w,r}} \mathbb{P}(A_{W}) + o\left(n^{-3\ell}\right)
\]
\[
\sim \left(\frac{2t}{n^2p}\right)^{\ell} + O\left(\left(\frac{t}{n}\right)^{\ell-1}(np)^{-\ell}\right) + O\left(\left(\frac{t}{n}\right)^{\ell+1}(np)^{-(\ell+1)}\right) + o\left(n^{-3\ell}\right) \sim \left(\frac{2t}{n^2p}\right)^{\ell}.
\]
\[\square\]

2.1 Proof of Theorem 1.3

Throughout this subsection \(H\) is a fixed graph with \(k\) vertices, \(\ell\) edges and \(m_0(H) = m_0 \geq 1\), \(\varepsilon > 0\), \(p \geq n^{-1/m_0+\varepsilon}\) and \(G\) is sampled according to \(G(n,p)\). We also denote by \(Y\) the number of copies of \(H\) in \(G\), write \(y = \mathbb{E}(Y)\), and recall that whp \(Y \sim y\). Let \(\mathcal{H} = \{H_1, \ldots, H_Y\}\) be the set of copies of \(H\) in \(G\). By \(H_i \cup H_j\) we denote the graph whose vertex set is \(V(H_i) \cup V(H_j)\) and whose edge set is \(E(H_i) \cup E(H_j)\) (where multiple edges are ignored). If \(H_i, H_j\) are non-vertex-disjoint we say they intersect and denote it by \(H_i \sim H_j\).

2.1.1 Proof of the negative part

Assume \(t \ll n^{2-1/m_0}\). Let \(H' \subseteq H\) with \(k_0\) vertices and \(\ell_0\) edges be such that \(\ell_0/k_0 = m_0\), and write \(\rho = \rho(H')\). Let \(Z\) be the number of copies of \(H'\) in \(\Gamma_t\). From Corollary 2.2 it follows that whp
\[
\mathbb{E}(Z) = O\left(n^{\ell_0-\ell} \sum_{r=\rho}^{\ell_0} \left(\frac{t}{n}\right)^{r}\right).
\]

Now, if \(m_0 = 1\) then \(k_0 = \ell_0\) and \(t \ll n\) and thus \(\mathbb{E}(Z) = o(1)\). If \(m_0 > 1\) then \(k_0 - \ell_0 \leq -1\); in that case, if \(t < n\) then \(\mathbb{E}(Z) = O\left(n^{-1}\right) = o(1)\), and if \(t \geq n\) we have that
\[
\mathbb{E}(Z) = O\left(n^{\ell_0-2\ell_0t}\right) = o\left(n^{k_0-2\ell_0n^{\ell_0-1}}\right) = o(1).
\]

Markov’s inequality then yields the desired result. \(\square\)

2.1.2 Proof of the positive part

Assume \(t \gg n^{2-1/m_0}\). We also assume, without loss of generality, that \(t = O\left(n^{2-\gamma}p\right)\) for sufficiently small \(\gamma > 0\). For \(1 \leq i \leq Y\) let \(Z_i\) be the indicator of the event \(H_i \subseteq \Gamma_t\), and let \(Z = \sum_{i=1}^{Y} Z_i\). For \(1 \leq i \neq j \leq Y\) write \(i \sim j\) if \(H_i \sim H_j\) and \(i \sim j\) otherwise. Let \(H' = H_i \cup H_j\) for \(i \sim j\). Let \(Z'\) be the random variable counting the number of appearances of a copy of \(H'\) in \(\Gamma_t\). We now show that whp \(\mathbb{E}(Z') = \mathbb{E}(Z)\). Indeed, according to Corollary 2.2 whp
\[
\mathbb{E}(Z) = \Theta\left(n^{k_0-2\ell t}\right) = \omega(1),
\]

11
and thus

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

Let \( k_0, \ell_0 \) be the number of vertices and edges in the intersection \( H_i \cap H_j \), respectively, and note that \( H' \) has \( 2k - k_0 \) vertices and \( 2\ell - \ell_0 \) edges and that \( m_0 \geq \ell_0/k_0 \). We therefore have that

\[ \mathbb{E}(Z') = \Theta\left(n^{(2k-k_0)-2(2\ell-\ell_0)}t^{2\ell-\ell_0}\right), \]

and thus

\[ \frac{\mathbb{E}^2(Z)}{\mathbb{E}(Z')} = \Theta\left(n^{k_0-2\ell_0}t^{\ell_0}\right), \]

so, as \( H_i, H_j \) are intersecting, either \( \ell_0 = 0 \) and \( k_0 > 0 \), in which case the above expression is \( \omega(1) \), or \( \ell_0 > 0 \), in which case \( t^{\ell_0} \gg n^{2k_0-\ell_0/m_0} \) and the above expression is

\[ \omega\left(n^{k_0-\ell_0/m_0}\right) = \omega(1). \]

Now suppose \( i \sim j \). Note that in that case \( Z_i, Z_j \) are almost independent, in the sense that their covariance is very small. Indeed, according to Lemma 2.1 and since \( t \gg n \),

\[ \mathbb{E}(Z_i) = \mathbb{P}(H_i \subseteq \Gamma_t) \sim (2t)^{\ell}(n^2p)^{-\ell}, \]

thus

\[ \mathbb{E}(Z_i Z_j) = \mathbb{P}(H_i \cup H_j \subseteq \Gamma_t) \sim (2t)^{2\ell}(n^2p)^{-2\ell}, \]

and

\[ \mathbb{E}(Z_i)\mathbb{E}(Z_j) = (\mathbb{P}(H_i \subseteq \Gamma_t))^2 \sim (2t)^{2\ell}(n^2p)^{-2\ell}, \]

thus

\[ \text{Cov}(Z_i, Z_j) = o\left(t^{2\ell}n^{-4\ell}p^{-2\ell}\right). \]

Since \( \mathbb{E}(Z) = \omega(1) \), it follows that

\[
\begin{align*}
\text{Var}(Z) & = \sum_{i=1}^{Y} \sum_{j=1}^{Y} \text{Cov}(Z_i, Z_j) \\
& = \sum_{i=1}^{Y} \text{Var}(Z_i) + \sum_{i \sim j} \text{Cov}(Z_i, Z_j) + \sum_{i \sim j} \text{Cov}(Z_i, Z_j) \\
& \leq \sum_{i=1}^{Y} \mathbb{E}(Z_i) + \sum_{i \sim j} \mathbb{P}(H_i \cup H_j \subseteq \Gamma_t) + o\left(n^{2k}p^{2\ell} \cdot t^{2\ell}n^{-4\ell}p^{-2\ell}\right) \\
& = \mathbb{E}(Z) + o(\mathbb{E}^2(Z)) + o(\mathbb{E}^2(Z)) + o(\mathbb{E}^2(Z)).
\end{align*}
\]

Chebyshev’s inequality then yields the desired result. \( \square \)
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 $\Gamma_t$ be the trace of a random walk of length $t$ on $K_n$. Then,

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

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

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

**Proof.** If $\text{odd}(G) = 0$ then $G$ is Eulerian, thus $\rho(G) = 1$. Otherwise, let $\text{odd}(G) = 2k$, and let $v_1, v_2, \ldots, v_{2k}$ be the odd degree vertices. Create $G'$ by adding the edges \{${v_i, v_{i+1}}$\}. $G'$ 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 $\rho(G) \leq k$. On the other hand, every trail removed from $E(G)$ decreases $\text{odd}(G)$ by at most 2, hence $\rho(G) \geq k$. \hfill \Box

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

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

Since $t \ll n$, we have that

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

Markov’s inequality then yields the result. \hfill \Box

#### 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 $\rho(T_1) \leq \rho(T_2)$.
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 $e$ not in $S_3$, 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. \hfill \Box

Alternative proof. In view of Lemma 3.2 it suffices to show that $\text{odd}(T_1) \leq \text{odd}(T_2)$, 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. \hfill \Box

Lemma 3.4. Let $T_1, T_2$ be two intersecting labelled copies of $T$ in $K_n$. Let $k_0$ and $\ell_0$ denote the number of vertices and edges, respectively, of the intersection $T_1 \cap T_2$, and let $\rho_0 = \rho(T_1 \cup T_2)$. Then

$$k_0 - \ell_0 - 2 + \rho_0/\rho \geq 0.$$ 

Proof. Observe that $T_1 \cap T_2$ is a forest. If it is not a tree, then $k_0 - \ell_0 \geq 2$ and the claim follows. Consider now the case where $T_1 \cap T_2$ is a tree. In that case, $k_0 - \ell_0 = 1$, thus it suffices to show that $\rho_0 \geq \rho$. Note that in that case it also follows that $T_1 \cup T_2$ is a tree, since it is connected with $2k - k_0$ vertices and $2\ell - \ell_0$ edges, and

$$(2k - k_0) - (2\ell - \ell_0) = 2(k - \ell) - (k_0 - \ell_0) = 1.$$ 

It follows that $T$ is a subtree of $T_1 \cup T_2$, thus by Lemma 3.3, $\rho \leq \rho_0$. \hfill \Box

Lemma 3.5. Let $T_1, T_2$ be two intersecting labelled copies of $T$ in $K_n$, and let $H = T_1 \cup T_2$. Write $\rho_0 = \rho(H)$. Let $n^{1-1/\rho} \ll t \ll n$, let $Z$ be the random variable counting the number of appearances of $T$ in $T_1$, and let $Y$ be the random variable counting the number of appearances of $H$ in $T_1$. Then

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

Proof. Let $k_0, \ell_0$ 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, $k_0 > \ell_0$. From Corollary 2.2, and since $t \ll n$, we have that

$$\mathbb{E}(Y) = \Theta\left(n^{2k-k_0-(2\ell-\ell_0)}\left(\frac{t}{n}\right)^{\rho_0}\right) = \Theta\left(n^{2+\ell_0-k_0-\rho_0/\rho}\right),$$

and

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

Now, if $\rho_0 \geq 2\rho$, then $(t/n)^{2\rho-\rho_0} = \Omega(1)$ and

$$\frac{\mathbb{E}^2(Z)}{\mathbb{E}(Y)} = \Theta\left(n^{k_0-\ell_0+\rho_0-2\rho/2-\rho_0}\right) = \Omega\left(n^{k_0-\ell_0}\right) = \omega(1).$$

14
On the other hand, if $\rho_0 < 2\rho$, then $(t/n)^{2\rho - \rho_0} \gg n^{\rho_0/\rho - 2}$ and

$$\frac{E^2(Z)}{E(Y)} = \Theta\left( n^{k_0 - \rho_0 - 2} t^{2\rho - \rho_0} \right) = \omega\left( n^{k_0 - \rho_0 - 2 + \rho_0/\rho} \right),$$

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

**Lemma 3.6.** Let $H_1, H_2$ be two vertex-disjoint subgraphs of $K_n$. Let $A_i$ be the event “$H_i \subseteq \Gamma_t$”. Then $A_1, A_2$ are not positively correlated.

**Proof.** For every $W \subseteq [t]$, let

$$W^+ = \{ w + j \mid w \in W, \ j \in \{-1, 0, 1\} \},$$

and let $\Gamma[W]$ be the set of edges traversed by the walk on times not in $W^+$. Formally,

$$\Gamma[W] = \{ \{X_{i-1}, X_i\} \mid i \in [t] \setminus W^+ \}.$$

For $W \subseteq [t]$, let $A^W_1$ be the event

$$A_1 \land \{ \{i \in [t] \mid \{X_{i-1}, X_i\} \in E(H_1) \} = W \}.$$

Since $H_1, H_2$ are vertex-disjoint, and since the events “$H_2 \subseteq \Gamma[W]$” and $A^W_1$ are independent, we have that

$$\mathbb{P}(A_2 \cap A_1) = \sum_{W \subseteq [t]} \mathbb{P}(A_2 \cap A^W_1)$$

$$= \sum_{W \subseteq [t]} \mathbb{P}((H_2 \subseteq \Gamma[W]) \land A^W_1)$$

$$= \sum_{W \subseteq [t]} \mathbb{P}(H_2 \subseteq \Gamma[W]) \cdot \mathbb{P}(A^W_1)$$

$$\leq \sum_{W \subseteq [t]} \mathbb{P}(H_2 \subseteq \Gamma_t) \cdot \mathbb{P}(A^W_1)$$

$$= \mathbb{P}(H_2 \subseteq \Gamma_t) \cdot \sum_{W \subseteq [t]} \mathbb{P}(A^W_1)$$

$$= \mathbb{P}(A_2) \cdot \mathbb{P}(A_1). \quad \square$$

**Proof of the positive part of Theorem 3.1.** Assume $n^{1-1/\rho} \ll t < n$. Suppose $T$ has $k$ vertices, and let $Z$ count the number of copies of $T$ in $\Gamma_t$. Recall that

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

Since $t \ll n$, we have that

$$\mathbb{E}(Z) = \Theta\left( n \left( \frac{t}{n} \right)^\rho \right) = \Theta(n^{1-\rho} t^\rho) = \omega(n^{1-\rho} n^{\rho-1}) = \omega(1).$$
Let $\mathcal{T} = \{T_1, T_2, \ldots, T_y\}$ be the set of all labelled copies of $T$ in $K_n$, let $Z_i$ be the indicator of the event $T_i \subseteq \Gamma_t$, let $\mathcal{H} = \{H_1, H_2, \ldots\}$ be the set of all possible unions of two intersecting (distinct) copies of $T$, and for $H \in \mathcal{H}$, let $Y_H$ be the random variable counting the number of copies of $H$ in $\Gamma_t$. Write $i \sim j$ if $Z_i$ and $Z_j$ are positively correlated, and recall (from Lemma 3.6) that if $i \sim j$ then $T_i \sim T_j$ (that is, $T_i, T_j$ intersect). It follows that

$$\text{Var}(Z) = \sum_{i=1}^y \sum_{j=1}^y \text{Cov}(Z_i, Z_j)$$

$$= \sum_{i=1}^y \text{Var}(Z_i) + \sum_{i \sim j} \text{Cov}(Z_i, Z_j) + \sum_{i \not\sim j} \text{Cov}(Z_i, Z_j)$$

$$\leq \sum_{i=1}^y \mathbb{E}(Z_i) + \sum_{i \sim j} \mathbb{P}(T_i \cup T_j \subseteq \Gamma_t)$$

$$\leq \mathbb{E}(Z) + \sum_{T_i \sim T_j} \mathbb{P}(T_i \cup T_j \subseteq \Gamma_t)$$

$$= \mathbb{E}(Z) + 2 \sum_{H \in \mathcal{H}} \mathbb{E}(Y_H).$$

Since $|\mathcal{H}| = O(1)$, it follows from Lemma 3.5 that $\text{Var}(Z) = o(\mathbb{E}^2(Z))$, 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 $\text{odd}(T_i) = \theta$. By Theorem 1.5, whp $T_i$ is not a subgraph of $\Gamma_t$, and hence $F$ is not a subgraph of $\Gamma_t$.

Now suppose that $t \gg n^{1-2/\theta}$. Let $s = \lfloor t/z \rfloor$, and for $i \in [z]$ let $\Gamma_i$ 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 $\Gamma_i$ (if there exists one; let it be an arbitrary tree otherwise). Note that the events $A_i$ are mutually independent. Let

$$U_i = \bigcup_{1 \leq j < i} V(T_j'),$$

let $B_i$ be the event that an edge from $\Gamma_i$ intersects $U_i$, and let $C_i = A_i \land \overline{B_i}$. Observe that for $U \subseteq [n]$ with $|U| = O(1)$, the probability that an edge from $\Gamma_i$ 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 $F$.\qed

### 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 \textit{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 $t \geq \ell$, then $\Gamma_t$ contains a copy of $P$ \textit{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.

Another possible direction would be to study the trace of the walk on other expander graphs, such as $(n,d,\lambda)$-graphs (see [11] for a survey), or on other random graphs, such as random regular graphs. The small subgraph problem for random regular graphs 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 $d \geq n^{1-1/m_0(H)+\varepsilon}$, 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?

\textbf{Acknowledgement.} The authors wish to thank Alan Frieze and Asaf Nachmias for useful discussions.

\textbf{References}


