On the trace of random walks on random graphs

Alan Frieze * Michael Krivelevich † Peleg Michaeli ‡ Ron Peled §

June 15, 2017

Abstract

We study graph-theoretic properties of the trace of a random walk on a random graph. We show that for any $\varepsilon > 0$ there exists $C > 1$ such that the trace of the simple random walk of length $(1+\varepsilon)n \ln n$ on the random graph $G \sim G(n,p)$ for $p > C \ln n/n$ is, with high probability, Hamiltonian and $\Theta(\ln n)$-connected. In the special case $p = 1$ (i.e. when $G = K_n$), we show a hitting time result according to which, with high probability, exactly one step after the last vertex has been visited, the trace becomes Hamiltonian, and one step after the last vertex has been visited for the k'th time, the trace becomes $2k$-connected.

1 Introduction

Since the seminal study of Erdős and Rényi [14], random graphs have become an important branch of modern combinatorics. It is an interesting and natural concept to study for its own sake, but it has also proven to have numerous applications both in combinatorics and in computer science. Indeed, random graphs have been a subject of intensive study during the last 50 years: thousands of papers and at least three books [6, 16, 19] are devoted to the subject. The term *random graph* is used to refer to several quite different models, each of which is essentially a distribution over all graphs on n labelled vertices. Perhaps the two most famous models are the classical models $G(n,m)$, obtained by choosing m edges uniformly at random among the $\binom{n}{2}$ possible edges, and $G(n,p)$, obtained by selecting each edge independently with probability p. Other models are discussed in [16].

In this paper, we study a different model of random graphs – the (random) graph formed by the trace of a simple random walk on a finite graph. Given a base graph and a starting vertex, we select a vertex uniformly at random from its neighbours and move to this neighbour, then independently select a vertex uniformly at random from that vertex's neighbours and move to it, and so on. The sequence of vertices this process yields is a *simple random walk* on that graph. The set of vertices in this sequence is called the *range* of the walk, and the set of edges

*Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA15213, USA. Email: `alan@random.math.cmu.edu`. Research supported in part by NSF Grant ccf0502793.

†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`.

§School of Mathematical Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 6997801, Israel. E-mail: `peledron@post.tau.ac.il`. Research supported in part by an IRG grant and by a grant from Israel Science Foundation.

traversed by this walk is called the *trace* of the walk. The literature on the topic of random walks is vast; however, most effort was put into answering questions about the range of the walk, or about the distribution of the position of the walk at a fixed time. Examples include estimating the *cover time* (the time it takes a walk to visit all vertices of the graph) and the *mixing time* of graphs (see Lovász [22] for a survey). On the other hand, to the best of our knowledge, there are almost no works addressing explicitly questions about the trace of the walk. One paper of this type is that of Barber and Long [2], discussing the quasirandomness of the trace of long random walks on dense quasirandom graphs. We mention here that there are several papers studying the subgraph induced by vertices that are *not* visited by the random walk (see for example Cooper and Frieze [11], and Černý, Teixeira and Windisch [8]). We also mention that on infinite graphs, several properties of the trace have been studied (for an example see [4]).

Our study focuses on the case where the base graph G is random and distributed as $G(n, p)$. We consider the graph Γ on the same vertex set ($[n] = \{1, \dots, n\}$), whose edges are the edges traversed by the random walk on G. A natural graph-theoretic question about Γ is whether it is connected. A basic requirement for that to happen is that the base graph is itself connected. It is a well-known result (see [13]) that in order to guarantee that G is connected, we must take $p > (\ln n + \omega(1))/n$. Given that our base graph is indeed connected, for the trace to be connected, the walk must visit all vertices. An important result by Feige [15] states that for connected graphs on n vertices, this happens on average after at least $(1 - o(1))n \ln n$ steps. Cooper and Frieze [10] later gave a precise estimation for the average cover time of (connected) random graphs, directly related to how large p is, in comparison to the connectivity threshold. In fact, it can be derived from their proof that if $p = \Theta(\ln n/n)$ and the length of the walk is at most $n \ln n$, then the trace is typically not connected.

It is thus natural to execute a random walk of length $(1+\varepsilon)n \ln n$ on a random graph which is above the connectivity threshold by at least a large constant factor (which may depend on ε), and to ask what other graph-theoretic properties the trace has. For example, is it highly connected? Is it Hamiltonian? The set of visited vertices does not reveal much information about the global structure of the graph, so the challenge here is to gain an understanding of that structure by keeping track of the traversed edges. What we essentially show is that the trace is typically Hamiltonian and $\Theta(\ln n)$-vertex-connected. Our method of proof will be to show that the set of traversed edges typically forms an expander.

In the boundary case where $p = 1$, i.e. when the base graph is K_n, we prove a much more precise result. As the trace becomes connected exactly when the last vertex has been visited, and at least one more step is required for that last visited vertex to have degree 2 in the trace, one cannot hope that the trace would contain a Hamilton cycle beforehand. It is reasonable to expect however that this degree requirement is in fact the bottleneck for a typical trace to be Hamiltonian, as is the case in other random graph models. In this paper, we show a hitting time result according to which, with high probability, one step after the walk connects the subgraph (that is, one time step after the cover time), the subgraph becomes Hamiltonian. This result implies that the bottleneck to Hamiltonicity of the trace lies indeed in the minimum degree, and in that sense, the result is similar in spirit to the results of Bollobás [5], and of Ajtai, Komlós and Szemerédi [1]. We also extend this result for k-cover-time vs. minimum degree $2k$ vs. $2k$-vertex-connectivity, obtaining a result similar in spirit to the result of Bollobás and Thomason [7].

Let us try to put our result in an appropriate context. The accumulated research experience in several models of random graphs allows to predict rather naturally that in our model (of the trace of a random walk on a random graph, or on the complete graph) the essential bottleneck for the Hamiltonicity would be existence of vertices of degree less than two. It is not hard to prove that typically the next step after entering the last unvisited vertex creates a graph of minimum degree two, and thus the appearance of a Hamilton cycle should be tightly bundled with the cover time of the random walk. Thus our results, confirming the above paradigm, are perhaps expected and not entirely surprising. Recall however that it took a fairly long and quite intense effort of the random graphs community, to go from the very easy, in modern terms at least, threshold for minimum degree two to the much more sophisticated result on Hamiltonicity, still considered one of the pinnacles of research in the subject. Here too we face similar difficulties; our proof, while having certain similarities to the prior proofs of Hamiltonicity in random graphs, still has its twists and turns to achieve the goal. In particular, for the case of a random walk on a random graph, we cast our argument in a pseudo-random setting and invoke a sufficient deterministic criterion for Hamiltonicity due to Hefetz, Krivelevich and Szabó [17]. The hitting time result about a random walk on the complete graph is still a delicate argument, as typical for hitting time results, due to the necessity to pinpoint the exact point of the appearance of a Hamilton cycle. There we look at the initial odd steps of the walk, forming a random multigraph with a given number of edges, and then add just the right amount of some further odd and even steps to hit the target exactly when needed.

1.1 Notation and terminology

Let G be a (multi)graph on the vertex set $[n]$. For two vertex sets $A, B \subseteq [n]$, we let $E_G(A, B)$ be the set of edges having one endpoint in A and the other in B. If $v \in [n]$ is a vertex, we may write $E_G(v, B)$ when we mean $E_G(\{v\}, B)$. We denote by $N_G(A)$ the *external* neighbourhood of A, i.e., the set of all vertices in $[n] \smallsetminus A$ that have a neighbour in A. Again, we may write $N_G(v)$ when we mean $N_G(\{v\})$. We also write $N_G^+(A) = N_G(A) \cup A$. The degree of a vertex $v \in [n]$, denoted by $d_G(v)$, is its number of incident edges, where a loop at u contributes 2 to its degree. The *simplified graph* of G is the simple graph G' obtained by replacing each multiedge with a single edge having the same endpoints, and removing all loops. The *simple degree* of a vertex is its degree in the simplified graph; it is denoted by $d'_G(v) = d_{G'}(v)$. We let $\delta(G)$ and $\Delta(G)$ be the minimum and maximum *simple* degrees (d') of G. Let the *edge boundary* of a vertex set S be the set of edges of G with exactly one endpoint in S, and denote it by $\partial_G S$ (thus $\partial_G(S)$ is the set of edge of the cut (S, S^c) in G). If v, u are distinct vertices of a graph G, the *distance from v to u* is defined to be the minimum length (measured in edges) of a path from v to u (or ∞ if there is no such path); it is denoted by $d_G(v, u)$. If v is a vertex, the *ball of radius r around v* is the set of vertices of distance at most r from v; it is denoted by $B_G(v, r)$. In symbols:

$$B_G(v, r) = \{u \in [n] \mid d_G(v, u) \leq r\}.$$

We also write $N_G(v, r) = B_G(v, r) \smallsetminus B_G(v, r-1)$ for the (inner) *vertex boundary* of that ball. We will sometimes omit the subscript G in the above notations if the graph G is clear from the context.

A *simple random walk* of length t on a graph G, starting at a vertex v, is denoted $(X_i^v(G))_{i=0}^t$. When the graph is clear from the context, we may omit it and simply write $(X_i^v)_{i=0}^t$. When the starting vertex is irrelevant, we may omit it as well, writing $(X_i)_{i=0}^t$. In Sections 2, 3 and 5 we shall define "lazy" versions of a simple random walk, for which we will use the same notation. The *trace* of a simple random walk on a graph G of length t, starting at a vertex v, is the subgraph of G having the same vertex set as G, whose edges are all edges traversed by the walk (including loops), counted with multiplicity (so it is in general a multigraph). We denote it by $\Gamma_t^v(G)$, Γ_t^v or Γ_t, depending on the context.

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 $[n]$ where the probability of each such a graph, $G = ([n], E)$, to be chosen is $p^{|E|}(1-p)^{\binom{n}{2}-|E|}$. 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 $\lim_{n\to\infty} f/g = 1$.

For the sake of simplicity and clarity of presentation, we do not make a particular effort to optimize the constants obtained in our proofs. We also omit floor and ceiling signs whenever these are not crucial. Most of our results are asymptotic in nature and whenever necessary we assume that n is sufficiently large. We say that an event holds *with high probability* (**whp**) if its probability tends to 1 as n tends to infinity.

1.2 Our results

Our first theorem states that if $G \sim G(n, p)$ with p above the connectivity threshold by at least some large constant factor, and the walk on G is long enough to traverse the expected number of edges required to make a random graph connected, then its trace is with high probability Hamiltonian and highly connected.

Theorem 1. *For every $\varepsilon > 0$ there exist $C = C(\varepsilon) > 0$ and $\beta = \beta(\varepsilon) > 0$ such that for every edge probability $p = p(n) \geq C \cdot \frac{\ln n}{n}$ and for every $v \in [n]$, a random graph $G \sim G(n, p)$ is **whp** such that for $L = (1 + \varepsilon)n \ln n$, the trace $\Gamma_L^v(G)$ of a simple random walk of length L on G, starting at v, is **whp** Hamiltonian and $(\beta \ln n)$-vertex-connected.*

Our proof strategy will be as follows. First we prove that **whp** $G \sim G(n, p)$ satisfies some pseudo-random properties. Then we show that **whp** the trace of a simple random walk on *any* given graph, which satisfies these pseudo-random properties, has good expansion properties. Namely, it has two properties, one ensures expansion of small sets, the other guarantees the existence of an edge between any two disjoint large sets.

In the next two theorems we address the case of a random walk X executed on the complete graph K_n, and we assume that the walk starts at a uniformly chosen vertex. Denote the number of visits of the random walk X to a vertex v by time t (including the starting vertex) by $\mu_t(v)$. For a natural number k, denote by τ_C^k the *k-cover time* of the graph by the random walk; that is, the first time t for which each vertex in G has been visited at least k times. In symbols,

$$\tau_C^k = \min\{t \mid \forall v \in [n],\ \mu_t(v) \geq k\}. \tag{1}$$

When $k = 1$ we simply write τ_C and call it the *cover time* of the graph. The objective of the following theorems is to prove that the minimal requirements for Hamiltonicity and k-vertex-connectivity are in fact the bottleneck for a typical trace to have these properties.

Theorem 2. *For a simple random walk on K_n, denote by $\tau_{\mathcal{H}}$ the hitting time of the property of being Hamiltonian. Then **whp** $\tau_{\mathcal{H}} = \tau_C + 1$.*

Corollary 3. *Assume n is even. For a simple random walk on K_n, denote by $\tau_{\mathcal{PM}}$ the hitting time of the property of admitting a perfect matching. Then **whp** $\tau_{\mathcal{PM}} = \tau_C$.*

Theorem 4. *For every $k \geq 1$, for a simple random walk on K_n, denote by τ_δ^k the hitting time of the property of being spanning with minimum simple degree k, and denote by τ_κ^k the hitting time of the property of being spanning k-vertex-connected. Then* ***whp***

$$\begin{aligned} \tau_C^k &= \tau_\delta^{2k-1} &= \tau_\kappa^{2k-1}, \\ \tau_C^k + 1 &= \tau_\delta^{2k} &= \tau_\kappa^{2k}. \end{aligned}$$

1.3 Organization

The organization of the paper is as follows. In the next section we present some auxiliary results, definitions and technical preliminaries. In Section 3 we explore important properties of the random walk on a pseudo-random graph. In Section 4 we prove the Hamiltonicity and vertex-connectivity results for the trace of the walk on $G(n,p)$. In Section 5 we prove the hitting time results of the walk on K_n. We end by concluding remarks and proposals for future work in Section 6.

2 Preliminaries

In this section we provide tools to be used by us in the succeeding sections. We start by stating two versions of known bounds on large deviations of random variables, due to Chernoff [9] and Hoeffding [18], whose proofs can be found, e.g., in Chapter 2 of [19]. Define

$$\varphi(x) = \begin{cases} (1+x)\ln(1+x) - x & x \geq -1 \\ \infty & \text{otherwise.} \end{cases}$$

Theorem 2.1 ([19, Theorem 2.1])**.** *Let $X \sim \mathrm{Bin}(n,p)$, $\mu = np$, $a \geq 0$. Then the following inequalities hold:*

$$\mathbb{P}(X \leq \mu - a) \leq \exp\left(-\mu\varphi\left(\frac{-a}{\mu}\right)\right) \leq \exp\left(-\frac{a^2}{2\mu}\right), \tag{2}$$

$$\mathbb{P}(X \geq \mu + a) \leq \exp\left(-\mu\varphi\left(\frac{a}{\mu}\right)\right) \leq \exp\left(-\frac{a^2}{2(\mu + a/3)}\right). \tag{3}$$

Theorem 2.2 ([19, Theorem 2.10])**.** *Let $N \geq 0$, and let $0 \leq K, n \leq N$ be natural numbers. Let $X \sim \mathrm{Hypergeometric}(N,K,n)$, $\mu = \mathbb{E}(X) = nKN^{-1}$. Then inequalities (2) and (3) hold.*

Observe that for every $c > 0$, letting $\ell_c(x) = -cx + 1 - c - e^{-c}$ we have that $\varphi(x) \geq \ell_c(x)$ for every x, as $\ell_c(x)$ is the tangent line to $\varphi(x)$ at $x = e^{-c} - 1$, and $\varphi(x)$ is convex. We thus obtain the following bound:

Corollary 2.3. *Let $X \sim \mathrm{Bin}(n,p)$, $\mu = np$, $0 < \alpha < 1 < \beta$. Then the following inequalities hold for every $c > 0$:*

$$\begin{aligned} \mathbb{P}(X \leq \alpha\mu) &\leq \exp\big(-\mu(1 - e^{-c} - \alpha c)\big), \\ \mathbb{P}(X \geq \beta\mu) &\leq \exp\big(-\mu(1 - e^{-c} - \beta c)\big). \end{aligned}$$

The following is a trivial yet useful bound:

Claim 2.4. *Suppose $X \sim \mathrm{Bin}(n,p)$. The following bound holds:*

$$\mathbb{P}(X \geq k) \leq \binom{n}{k} p^k \leq \left(\frac{enp}{k}\right)^k.$$

Proof. Think of X as $X = \sum_{i=1}^{n} X_i$, where X_i are i.i.d. Bernoulli tests with probability p. For any set $A \subseteq [n]$ with $|A| = k$, let E_A be the event "X_i have succeeded for all $i \in A$". Clearly, $\mathbb{P}(E_A) = p^k$. If $X \geq k$, there exists $A \subseteq [n]$ for which E_A. Thus, the union bound gives

$$\mathbb{P}(X \geq k) \leq \binom{n}{k}\mathbb{P}(E_A) = \binom{n}{k}p^k \leq \left(\frac{enp}{k}\right)^k. \qquad \square$$

2.1 (R, c)-expanders

Let us first define the type of expanders we intend to use.

Definition 2.5. *For every $c > 0$ and every positive integer R we say that a graph $G = (V, E)$ is an (R, c)*-expander *if every subset of vertices $U \subseteq V$ of cardinality $|U| \leq R$ satisfies $|N_G(U)| \geq c|U|$.*

Next, we state some properties of (R, c)-expanders.

Claim 2.6. *Let $G = (V, E)$ be an (R, c)-expander, and let $S \subseteq V$ of cardinality $k < c$. Denote the connected components of $G \setminus S$ by $S_1, \ldots, S_t$, so that $1 \leq |S_1| \leq \ldots \leq |S_t|$. It follows that $|S_1| > R$.*

Proof. Assume otherwise. Since any external neighbour of a vertex from S_1 must be in S, we have that

$$c > k = |S| \geq |N(S_1)| \geq c|S_1| \geq c,$$

which is a contradiction. $\square$

The following is a slight improvement of [3, Lemma 5.1].

Lemma 2.7. *For every positive integer k, if $G = ([n], E)$ is an (R, c)-expander such that $c \geq k$ and $R(c + 1) \geq \frac{1}{2}(n + k)$, then G is k-vertex-connected.*

Proof. Assume otherwise; let $S \subseteq [n]$ with $|S| = k - 1$ be a disconnecting set of vertices. Denote the connected components of $G \setminus S$ by $S_1, \ldots, S_t$, so that $1 \leq |S_1| \leq \ldots \leq |S_t|$ and $t \geq 2$. It follows from Claim 2.6 that $|S_1| > R$.

Take $A_i \subseteq S_i$ for $i \in [2]$ with $|A_i| = R$. Since any common neighbour of A_1 and A_2 must lie in S, it follows that

$$\begin{aligned} n &\geq |S_1 \cup S_2 \cup N(S_1) \cup N(S_2)| \\ &\geq |N^+(A_1) \cup N^+(A_2)| \\ &= |N^+(A_1)| + |N^+(A_2)| - |N(A_1) \cap N(A_2)| \\ &\geq 2R(c + 1) - |S| \geq n + 1, \end{aligned}$$

which is a contradiction. $\square$

The reason we study (R, c)-expanders is the fact that they entail some pseudo-random properties, from which we will infer the properties that are considered in this paper, namely, being Hamiltonian, admitting a perfect matching, and being k-vertex-connected.

2.2 Properties of random graphs

In the following technical lemma we establish properties of random graphs to be used later to prove Theorem 1.

Theorem 2.8. *Let $C = 150$ and let $C \leq \alpha = \alpha(n) \leq \frac{n}{\ln n}$. Let $p = p(n) = \alpha \cdot \frac{\ln n}{n}$, and let $G \sim G(n,p)$. Then,* ***whp****,*

(P1) *G is connected,*

(P2) *For every $v \in [n]$, $|d(v) - \alpha \ln n| \leq 2\sqrt{\alpha} \ln n$; in particular, $\frac{5\alpha}{6} \ln n \leq d(v) \leq \frac{4\alpha}{3} \ln n$,*

(P3) *For every non-empty set $S \subseteq [n]$ with at most $0.8n$ vertices, $|\partial S| > \frac{|S||S^c|\alpha \ln n}{2n}$,*

(P4) *For every large enough constant $K > 0$ (which does not depend on α) and for every non-empty set $A \subseteq [n]$ with $|A| = a$, the following holds:*

- *If $\frac{n}{\alpha \ln n} \leq a \leq \frac{n}{\ln n}$ then*
$$|E(A, \{u \in N(A) \mid |E(u,A)| \geq K a \alpha \ln n / n\})| \leq a\alpha \ln n / K;$$
- *If $a < \frac{n}{\alpha \ln n}$ then*
$$|E((A, \{u \in N(A) \mid |E(u,A)| \geq K\})| \leq a\alpha \ln n / \ln K.$$

(P5) *For every set A with $|A| = n(\ln \ln n)^{1.5}/\ln n$ there exist at most $|A|/2$ vertices v not in A for which $|E(v,A)| \leq \alpha(\ln \ln n)^{1.5}/2$,*

(P6) *If $\alpha < \ln^2 n$ then for every $v \in [n]$, $0 \leq r \leq \frac{\ln n}{15 \ln \ln n}$, and $w \in N(v,r)$ we have that $|E(w, B(v,r))| \leq 5$.*

Property (P1) is well-known (see, e.g., [16]), so we omit the proof here.

Proof of (P2). We note that $d(v) \sim \mathrm{Bin}(n-1, p)$. Denote $\mu = \mathbb{E}(d(v)) = (n-1)p$. Fix a vertex $v \in [n]$. Using Chernoff bounds (Theorem 2.1) we have that

$$\begin{aligned}\mathbb{P}\big(d(v) \leq \alpha \ln n - 2\sqrt{\alpha} \ln n\big) &\leq \mathbb{P}\Big(d(v) \leq \mu - \sqrt{3\alpha} \ln n\Big) \\ &\leq \exp\left(-\frac{3\alpha \ln^2 n}{2\mu}\right) = o\big(n^{-1}\big),\end{aligned}$$

and that

$$\begin{aligned}\mathbb{P}\big(d(v) \geq \alpha \ln n + 2\sqrt{\alpha} \ln n\big) &\leq \mathbb{P}\big(d(v) \geq \mu + 2\sqrt{\alpha} \ln n\big) \\ &\leq \exp\left(-\frac{4\alpha \ln^2 n}{2\big(\mu + \frac{2}{3}\sqrt{\alpha} \ln n\big)}\right) \leq \exp\left(-\frac{6}{5} \ln n\right) = o\big(n^{-1}\big).\end{aligned}$$

The union bound over all vertices $v \in [n]$ yields the desired result. Since $\alpha \geq 150$ we also ensure that for every $v \in [n]$,

$$\frac{5\alpha}{6} \ln n \leq d(v) \leq \frac{4\alpha}{3} \ln n. \qquad \square$$

Proof of (P3). Fix a set $S \subseteq [n]$ with $1 \le |S| = s \le 0.8n$. We note that $|\partial S| \sim \mathrm{Bin}(s(n-s), p)$, thus by Theorem 2.1 we have that

$$\mathbb{P}\Big(|\partial S| \le \frac{1}{2}s(n-s)p\Big) \le \exp\Big(-\frac{1}{8}s(n-s)p\Big).$$

Let F be the event "$\exists S$ such that $|\partial S| \le \frac{1}{2}s(n-s)p$". The union bound gives

$$\begin{aligned}
\mathbb{P}(F) &\le \sum_{s=1}^{0.8n} \binom{n}{s} \exp\Big(-\frac{1}{8}s(n-s)p\Big) \\
&\le \sum_{s=1}^{0.8n} \exp\Big(s\Big(1 + \ln n - \ln s - \frac{1}{8}(n-s)p\Big)\Big) \\
&\le \sum_{s=1}^{0.8n} \exp\Big(s\Big(1 + \ln n - \ln s - \frac{\alpha}{40}\ln n\Big)\Big) = o(1),
\end{aligned}$$

since $\alpha > 80$. □

Proof of (P4). Fix A with $|A| = a$, and suppose first that $\frac{n}{\alpha \ln n} \le a \le \frac{n}{\ln n}$. Let

$$B_0 = \{u \in N(A) \mid |E(b, A)| \ge Kap\},$$

for large K to be determined later. For a vertex $u \notin A$, the random variable $|E(u, A)|$ is binomially distributed with a trials and success probability p, and these random variables are independent for different vertices u. Thus, using Claim 2.4 we have that for large enough K,

$$\mathbb{P}(|E(b, A)| \ge Kap) \le \Big(\frac{e}{K}\Big)^{Kap} \le e^{-K}.$$

Thus $|B_0|$ is stochastically dominated by a binomial random variable with n trials and success probability e^{-K}. It follows again by Claim 2.4 that $\mathbb{P}\big(|B_0| > 3e^{-K}n\big) \le c^n$ for some $0 < c < 1$. Since $a \le n/\ln n$, $n\binom{n}{a} = o(c^{-n})$. Thus by the union bound,

$$\begin{aligned}
\mathbb{P}\Big(\exists A,\ |A| = a:\ |E(A, B_0)| > \frac{anp}{K}\Big) &\le \binom{n}{a}\Big(c^n + \mathbb{P}\Big(|E(A, B_0)| > \frac{anp}{K} \mid |B_0| \le 3e^{-K}n\Big)\Big) \\
&\le o\big(n^{-1}\big) + \binom{n}{a}\binom{n}{3e^{-K}n}\binom{3ae^{-K}n}{anp/K}p^{anp/K} \\
&\le o\big(n^{-1}\big) + \binom{n}{4e^{-K}n}\binom{3ae^{-K}n}{anp/K}p^{anp/K} \\
&\le o\big(n^{-1}\big) + e^{4Ke^{-K}n} \cdot \big(9Ke^{-K}\big)^{anp/K} \\
&= o\big(n^{-1}\big) + \Big(e^{4Ke^{-K}} \cdot \big(9Ke^{-K}\big)^{ap/K}\Big)^n = o\big(n^{-1}\big),
\end{aligned}$$

for large enough K. Now suppose $a \le \frac{n}{\alpha \ln n}$. Let

$$B_0 = \{u \in N(A) \mid |E(b, A)| \ge K\}.$$

From (P2) we know that the number of edges going out from A is **whp** at most $4anp/3$. Given that, $|B_0| \leq 2anp/K$. Let F_a be the event "there exists A, $|A| = a$, such that $|E(A, B_0)| > \frac{anp}{\ln K}$". Thus,

$$\begin{aligned}\mathbb{P}(F_a \mid \Delta(G) \leq 4np/3) &\leq \binom{n}{a}\binom{n}{2anp/K}\binom{2a^2np/K}{anp/\ln K} p^{anp/\ln K} \\ &\leq \left[n^{1/(np)} \left(\frac{eK}{2ap}\right)^{2/K} \left(\frac{2eap\ln K}{K}\right)^{1/\ln K}\right]^{anp} \\ &\leq \left[e^{1/\alpha} \left(\frac{eK}{2}\right)^{2/K} \left(\frac{2e\ln K}{K}\right)^{1/\ln K} (ap)^{1/\ln K - 2/K}\right]^{anp} = o\big(n^{-1}\big),\end{aligned}$$

for large enough K. Taking the union bound over all cardinalities $1 \leq a \leq n/\ln n$ implies that the claim holds **whp** in both cases. $\square$

Proof of (P5). Fix a set $A \subseteq [n]$ with $|A| = \Lambda = n(\ln\ln n)^{1.5}/\ln n$. We say that a vertex $v \notin A$ is *bad* if $|E(v, A)| \leq \Lambda p/2$. Since $|E(v, A)| \sim \mathrm{Bin}(\Lambda, p)$, Chernoff bounds (Theorem 2.1) give that the probability that v is bad with respect to A is at most $\exp(-\Lambda p/8)$.

Let U_A be the set of bad vertices with respect to A. We now show that U_A is typically not too large. To this end, note that $|U_A|$ is stochastically dominated by a binomial random variable with n trials and success probability $\exp(-\Lambda p/8)$. Thus, using Claim 2.4, we have that

$$\mathbb{P}(|U_A| \geq \Lambda/2) \leq \binom{n}{\Lambda/2} \exp\big(-\Lambda^2 p/16\big).$$

The probability that there exists A of cardinality Λ whose set of bad vertices is of cardinality at least $\Lambda/2$ is thus at most

$$\begin{aligned}\mathbb{P}(\exists A :\ |A| = \Lambda,\ |U_A| \geq \Lambda/2) &\leq \binom{n}{\Lambda}\binom{n}{\Lambda/2}\exp\big(-\Lambda^2 p/16\big) \\ &\leq \left(\frac{en}{\Lambda}\right)^{2\Lambda} \exp\big(-\Lambda^2 p/16\big) \\ &\leq \exp\big(3\Lambda \ln(n/\Lambda) - \Lambda^2 p/16\big) \\ &\leq \exp\left(3 \cdot \frac{n}{\ln n}(\ln\ln n)^{2.5} - \frac{n}{\ln n}(\ln\ln n)^3\right) \\ &\leq \exp\left(-\frac{n}{\ln n}(\ln\ln n)^{2.9}\right) = o(1),\end{aligned}$$

here we used $\alpha > 16$. Noting that $\Lambda p = \alpha(\ln\ln n)^{1.5}$, the claim follows. $\square$

Proof of (P6). Assume $\alpha < \ln^2 n$ and let $\lambda = \frac{\ln n}{15\ln\ln n}$. Fix $v \in [n]$, $0 \leq r \leq \lambda$, expose a BFS tree T, rooted at v, of depth r, and fix a vertex $w \in N(v, r)$. Note that in T, the vertex w is a leaf, and thus has a single neighbour in $B(v, r)$. We now expose the rest of the edges between w and $B(v, r)$. Note that by definition the neighbours of w must be in $N(v, r-1) \cup N(v, r)$, and by (P2) the cardinality of that set is at most $(c\ln^3 n)^r$ for some $c > 1$, hence by Claim 2.4,

$$\mathbb{P}(|E(w, B(v, r))| \geq 6) \leq \left(\frac{e(c\ln^3 n)^r p}{6}\right)^6 \leq \left(\frac{(c\ln^3 n)^{\lambda+1}}{2n}\right)^6 = o\big(n^{-3}\big).$$

It follows by the union bound that

$$\mathbb{P}(\exists v \in [n], 0 \leq r \leq \lambda, w \in N(v, r) :\ |E(w, B(v, r))| \geq 6) \leq n \cdot \lambda \cdot n \cdot o\big(n^{-3}\big) = o(1). \quad \square$$

For $\alpha = \alpha(n) > 0$, a graph for which (P1),…,(P6) hold will be called α-*pseudo-random*.

2.3 Properties of random walks

Throughout this section, G is a graph with vertex set $[n]$, having properties (P1), (P2) and (P3) for some $\alpha > 0$, and X is a $\frac{1}{2}$-*lazy* simple random walk on G, starting at some arbitrary vertex v_0. By $\frac{1}{2}$-lazy we mean that it stays put with probability $\frac{1}{2}$ at each time step, and moves to a uniformly chosen random neighbour otherwise. Our purpose in this section is to show that X mixes well, in a sense that will be further clarified below. To this end, we shall need some preliminary definitions and notations.

The *transition probability* of X from u to v is the probability

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

For $k \in \mathbb{N}$ we similarly denote

$$p_{uv}^{(k)} = \mathbb{P}(X_{t+k} = v \mid X_t = u) = \mathbb{P}(X_k = v \mid X_0 = u).$$

We note that the stationary distribution 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 $S \subseteq [n]$,

$$\pi_S = \sum_{v\in S} \pi_v.$$

The *total variation distance* between X_t and the stationary distribution is

$$d_{\mathrm{TV}}(X_t, \pi) = \sup_{S\subseteq[n]} |\mathbb{P}(X_t \in S) - \pi_S|,$$

and as is well-known, we have that

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

Now, let $(Y_t)_{t\geq 0}$ be the *stationary walk* on G; that is, the $\frac{1}{2}$-lazy simple random walk for which for every $v \in [n]$, $\mathbb{P}(Y_0 = v) = \pi_v$. We note for later use that there exists a standard coupling of X, Y under which for every t,

$$\mathbb{P}(\exists s \geq t \mid X_s \neq Y_s) \leq d_{\mathrm{TV}}(X_t, \pi).$$

Our goal is therefore to find not too large t's for which the total variation distance is very small. That is, we wish to bound the ξ-*mixing time* of X, which is given by

$$\tau(\xi) = \min\{t \geq 0 \mid \forall s \geq t,\ d_{\mathrm{TV}}(X_s, \pi) < \xi\}.$$

A theorem of Jerrum and Sinclair [20] will imply that the ξ-mixing time of X is indeed small. Their bound uses the notion of *conductance*: the conductance of a cut (S, S^c) with respect to X is defined as

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

which can be equivalently written in our case as

$$\phi_X(S) = \frac{|\partial S|}{2\min\left(\sum_{v\in S} d(v), \sum_{w\in S^c} d(w)\right)}.$$

The conductance of G with respect to X is defined as

$$\Phi_X(G) = \min_{\substack{S\subseteq[n]\\ 0<\pi_S\le 1/2}} \phi_X(S).$$

Claim 2.9. *Let $\pi_{\min} = \min_v \pi_v$. For every $\xi > 0$,*

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

Proof. Let

$$\tau'(\xi) = \min\left\{ t \ge 0 \mid \forall s \ge t,\ u, v \in [n],\ \frac{\left|p_{uv}^{(s)} - \pi_v\right|}{\pi_v} < \xi \right\}$$

be the ξ-uniform mixing time of X. Noting that the laziness of the walk implies $p_{uu} \ge \frac{1}{2}$ for every vertex u, Corollary 2.3 in [20] states that

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

Let $t = \tau'(\xi)$; thus, for all $s \ge t$, $u, v \in [n]$, $\left|p_{uv}^{(s)} - \pi_v\right| < \xi\pi_v$. Fix $s \ge t$. We have that

$$d_{\mathrm{TV}}(X_s, \pi) = \frac{1}{2}\sum_{v\in[n]} |\mathbb{P}(X_s = v) - \pi_v| \le \frac{\xi}{2}\sum_{v\in[n]} \pi_v = \frac{\xi}{2},$$

thus $\tau(\xi) \le \tau(\xi/2) \le t = \tau'(\xi)$ and the claim follows. $\square$

Corollary 2.10. *For $\xi > 0$, $\tau(\xi) \le 1800 \ln(2n/\xi)$.*

Proof. We note that due to (P2), for every $v \in [n]$,

$$\pi_v \ge \frac{5}{8n},$$

and thus for every $S \subseteq [n]$ with $0 < \pi_S \le 1/2$ we have that

$$\frac{1}{2} \ge \pi_S \ge |S| \cdot \frac{5}{8n},$$

hence $0 < |S| \le \frac{4}{5}n$. Thus, according to (P2),(P3),

$$\begin{aligned}
\Phi_X(G) &= \min_{\substack{S\subseteq[n]\\ 0<\pi_S\le 1/2}} \phi_X(S) \\
&\ge \min_{\substack{S\subseteq[n]\\ 0<|S|\le 4n/5}} \frac{|\partial S|}{2\sum_{v\in S} d(v)} \\
&\ge \min_{\substack{S\subseteq[n]\\ 0<|S|\le 4n/5}} \frac{\frac{|S||S^c|\alpha\ln n}{2n}}{2|S|\cdot\frac{4}{3}\alpha\ln n} \ge \frac{1}{30}.
\end{aligned}$$

Plugging this into Claim 2.9 we have

$$\tau(\xi) \leq 1800\left(\ln\left(\frac{8n}{5}\right) + \ln\left(\frac{1}{\xi}\right)\right) \leq 1800 \ln\left(\frac{2n}{\xi}\right). \qquad \square$$

The following is an immediate corollary of the above discussion:

Corollary 2.11. *Let $b = \tau(1/n) \leq 3601 \ln n$. Conditioned on $X_0, \ldots, X_t$, there exists a coupling of $(X_{t+b+s})_{s\geq 0}$ and $(Y_s)_{s\geq 0}$ under which*

$$\mathbb{P}(\exists s \geq 0 \mid X_{t+b+s} \neq Y_s) \leq \frac{1}{n}.$$

3 Walking on a pseudo-random graph

In order to prove Theorem 1, we will prove that the trace $\Gamma = \Gamma_L^v(G)$ is **whp** a good expander, in the sense that it satisfies the following two properties:

(E1) There exists $\beta > 0$ such that every set $A \subseteq [n]$ of cardinality $|A| \leq \frac{n}{\ln n}$ satisfies $|N_\Gamma(A)| \geq |A| \cdot \beta \ln n$;

(E2) There is an edge of Γ between every pair of disjoint subsets $A, B \subseteq [n]$ satisfying $|A|, |B| \geq \frac{n(\ln \ln n)^{1.5}}{\ln n}$.

Theorem 3.1. *For every $\varepsilon > 0$ there exist $C = C(\varepsilon) > 0$ and $\beta = \beta(\varepsilon) > 0$ such that for every edge probability $p = p(n) \geq C \cdot \frac{\ln n}{n}$ and for every $v_0 \in [n]$, a random graph $G \sim G(n, p)$ is **whp** such that for $L = (1+\varepsilon)n \ln n$, the trace $\Gamma_L^{v_0}(G)$ of a simple random walk of length L on G, starting at v_0, has the properties (E1) and (E2) **whp**.*

It will be convenient the consider a slight variation of this theorem, in which the random walk is *lazy* and the base graph is pseudo-random:

Theorem 3.2. *For every sufficiently small $\varepsilon > 0$, if $\alpha \geq 1500\varepsilon^{-2}$ and G is a α-pseudo-random graph on the vertex set $[n]$, $v_0 \in [n]$ and $L_2 = (2+\varepsilon)n \ln n$, then the trace $\Gamma_{L_2}^{v_0}(G)$ of a $\frac{1}{2}$-lazy random walk of length L_2 on G, starting at v_0, has the properties (E1) and (E2) **whp**.*

Before proving this theorem, we show that Theorem 3.1 is a simple consequence of it.

Proof of Theorem 3.1. Since (E1) and (E2) are monotone, we may assume ε is sufficiently small. Let $L = (1+\varepsilon)n \ln n$, $L_2 = (2+\varepsilon)n \ln n$, $C = 1500\varepsilon^{-2}$ and $p = \alpha \ln n/n$ for $\alpha \geq C$. Let $(X_t^{v_0})_{t=0}^{L_2}$ be the $\frac{1}{2}$-lazy random walk of length L_2 on G, starting at v_0, and define

$$R = |\{0 < t \leq L_2 \mid X_t \neq X_{t-1}\}|.$$

Since $\mathbb{P}(X_t = X_{t-1}) = 1/2$ for every $0 < t \leq L_2$, by Chernoff bounds (Theorem 2.1) we have that $\mathbb{P}(R > L) = o(1)$. Denote by $\Gamma_{L_2}^\ell$ the trace of that walk, ignoring any loops, and let P be a monotone graph property which $\Gamma_{L_2}^\ell$ satisfies **whp**. Given R, the trace $\Gamma_{L_2}^\ell$ has the same distribution as the trace of the non-lazy walk Γ_R. Thus:

$$\mathbb{P}(\Gamma_L \in P) \geq \mathbb{P}\left(\Gamma_{L_2}^\ell \in P, R \leq L\right) = 1 - o(1).$$

As (E1) and (E2) are both monotone, and since G is α-pseudo-random **whp** by Theorem 2.8, the claim holds using Theorem 3.2. $\square$

Thus, in what follows, $\varepsilon > 0$ is sufficiently small, G is a α-pseudo-random graph on the vertex set $[n]$ for $\alpha \geq C = 1500\varepsilon^{-2}$, $v_0 \in [n]$ is some fixed vertex, $L_2 = (2+\varepsilon)n \ln n$, $(X_t^{v_0})_{t=0}^{L_2}$ is a $\frac{1}{2}$-lazy simple random walk of length L_2 on G, starting at v_0 (which we may simply refer to as X), and $(Y_t)_{t=0}^{\infty}$ (or simply Y) is the $\frac{1}{2}$-lazy simple random walk on G, starting at a random vertex sampled from the stationary distribution of X (the *stationary walk*).

The rest of this section is organised as follows. In the first subsection we show that **whp**, every vertex is visited at least a logarithmic number of times. In the second and third subsections we use this fact to conclude that small vertex sets typically expand well, and that large vertex sets are typically connected, by that proving that the trace satisfies (E1) and (E2) **whp**.

3.1 Number of visits

Define

$$\nu(v) = |\{0 < t \leq L_2 \mid X_t = v,\ X_{t+1} \neq v\}|, \quad v \in [n].$$

Theorem 3.3. *There exists $\rho > 0$ such that **whp**, for every $v \in [n]$, $\nu(v) \geq \rho \ln n$.*

In order to prove this theorem, we first introduce a number of definitions and lemmas. Recall that a *supermartingale* is a sequence $M(0), M(1), \dots$ of random variables such that each conditional expectation $\mathbb{E}(M(t+1) \mid M(0), \dots, M(t))$ is at most $M(t)$. Given such a sequence, a *stopping rule* is a function from finite histories of the sequence into $\{0,1\}$, and a *stopping time* is the minimum time at which the stopping rule is satisfied (that is, equals 1). For two integers s, t, let $s \wedge t = \min\{s, t\}$. Let $\lambda = \frac{\ln n}{15 \ln \ln n}$, and for every $v \in [n]$ let F_t^v be the event "$Y_t = v$ or $d_G(Y_t, v) > \lambda$" (recall that for two vertices u, v, $d_G(u,v)$ denotes the distance from u to v in G). Note that if $\alpha < \ln^2 n$ it follows from (P2) and (P6) that the diameter of G is larger than λ.

Lemma 3.4. *Suppose $\alpha < \ln^2 n$. For $v \in [n]$, define the process*

$$\mathcal{M}^v(t) = \left(\frac{10}{\alpha \ln n}\right)^{d_G(Y_t, v)}.$$

Let $S = \min\{t \geq 0 : F_t^v\}$ be a stopping time; then $\mathcal{M}^v(t \wedge S)$ is a supermartingale. In particular, for every $u \in [n]$ the stationary walk Y_t satisfies

$$\mathbb{P}(Y_S = v \mid Y_0 = u) \leq \left(\frac{10}{\alpha \ln n}\right)^{d_G(u,v)}.$$

Proof. For a vertex $w \in [n]$, denote

$$p_{\leftarrow}(w) = \mathbb{P}(d_G(Y_1, v) < d_G(Y_0, v) \mid Y_0 = w),$$
$$p_{\rightarrow}(w) = \mathbb{P}(d_G(Y_1, v) > d_G(Y_0, v) \mid Y_0 = w).$$

We note that for $0 < y \leq x \leq 1$, $\frac{y}{x} + x - y \leq 1$. Thus, for $q_1, q_2 > 0$ for which $\frac{p_{\leftarrow}(w)}{p_{\rightarrow}(w)} \leq \frac{q_1}{q_2} \leq 1$,

$$\begin{aligned}
&\mathbb{E}\left(\left(\frac{q_1}{q_2}\right)^{d_G(Y_1,v)} \mid Y_0 = w\right) - \left(\frac{q_1}{q_2}\right)^{d_G(w,v)} \\
&= \left(\frac{q_1}{q_2}\right)^{d_G(w,v)} \left(\left(p_{\leftarrow}(w)\frac{q_2}{q_1} + p_{\rightarrow}(w)\frac{q_1}{q_2} + (1 - p_{\leftarrow}(w) - p_{\rightarrow}(w))\right) - 1\right) \\
&= \left(\frac{q_1}{q_2}\right)^{d_G(w,v)} p_{\rightarrow}(w)\left(\frac{bp_{\leftarrow}(w)}{ap_{\rightarrow}(w)} + \frac{q_1}{q_2} - \frac{p_{\leftarrow}(w)}{p_{\rightarrow}(w)} - 1\right) \leq 0.
\end{aligned}$$

Let w be such that $0 < d_G(v,w) \leq \lambda$. Since $\alpha < \ln^2 n$, G satisfies (P6). Considering that and (P2), and since $\alpha > 25$, we have that

$$q_{\leftarrow} := \frac{5}{2(\alpha - 2\sqrt{\alpha})\ln n} \geq \frac{5}{2d_G(w)} \geq p_{\leftarrow}(w),$$
$$q_{\rightarrow} := \frac{\alpha \ln n}{4(\alpha - 2\sqrt{\alpha})\ln n} \leq \frac{1}{2} - \frac{5}{2(\alpha - 2\sqrt{\alpha})\ln n} \leq \frac{d_G(w) - 5}{2d_G(w)} \leq p_{\rightarrow}(w),$$

and as $\frac{p_{\leftarrow}(w)}{p_{\rightarrow}(w)} \leq \frac{q_{\leftarrow}}{q_{\rightarrow}} = \frac{10}{\alpha \ln n} \leq 1$ we have that $\mathcal{M}^v(t \wedge S)$ is a supermartingale. In addition, for every $t \geq 0$, almost surely,

$$\begin{aligned}
\left(\frac{10}{\alpha \ln n}\right)^{d_G(Y_0,v)} &= \mathcal{M}^v(0) \\
&\geq \mathbb{E}(\mathcal{M}^v(t \wedge S) \mid Y_0) \\
&= \sum_{w \in [n]} \left(\frac{10}{\alpha \ln n}\right)^{d_G(w,v)} \cdot \mathbb{P}(Y_{t \wedge S} = w \mid Y_0) \\
&\geq \mathbb{P}(Y_{t \wedge S} = v \mid Y_0).
\end{aligned}$$

As this is true for every $t \geq 0$, and since S is almost surely finite, it follows that for every $u \in [n]$,

$$\mathbb{P}(Y_S = v \mid Y_0 = u) \leq \left(\frac{10}{\alpha \ln n}\right)^{d_G(u,v)}$$

whp. ∎

Let

$$T = \ln^2 n.$$

For a walk W on G, let $I_W(v,t)$ be the event "$W_t = v$ and $W_{t+1} \neq v$". Our next goal is to estimate the probability that $I_Y(v,t)$ occurs for some $1 \leq t < T$, given that $I_Y(v,0)$ has occurred.

Lemma 3.5. *For every vertex $v \in [n]$ we have that*

$$\mathbb{P}\left(\bigcup_{1 \leq t < T} I_Y(v,t) \mid I_Y(v,0)\right) \leq \ln^{-1/2} n.$$

Proof. Fix $v \in [n]$. Define the following sequence of stopping times: $U_0 = 0$, and for $i \geq 1$,

$$U_i = \min\{t > U_{i-1} \mid F_t^v\}.$$

Then,

$$\mathbb{P}\left(\bigcup_{1 \leq t < T} I_Y(v,t) \mid I_Y(v,0)\right) \leq \mathbb{P}(Y_{U_1} = v \mid I_Y(v,0)) + \sum_{i=2}^{T-1} \mathbb{P}\big(Y_{U_i} = v \mid Y_{U_{i-1}} \neq v,\ I_Y(v,0)\big).$$

Now, if $\alpha < \ln^2 n$, Lemma 3.4 and the Markov property imply that

$$\mathbb{P}(Y_{U_1} = v \mid I_Y(v,0)) \le \frac{10}{\alpha \ln n},$$
$$\mathbb{P}\big(Y_{U_i} = v \mid Y_{U_{i-1}} \ne v,\ I_Y(v,0)\big) \le \left(\frac{10}{\alpha \ln n}\right)^{\lambda} \qquad (i \ge 2),$$

so

$$\mathbb{P}\left(\bigcup_{1 \le i < T} I_Y(v,t) \mid I_Y(v,0)\right) \le \frac{10}{\alpha \ln n} + T \cdot \left(\frac{10}{\alpha \ln n}\right)^{\lambda}$$
$$= \frac{10}{\alpha \ln n} + \ln^2 n \cdot o\Big(n^{-1/20}\Big) \le \frac{20}{\alpha \ln n} \le \ln^{-1/2} n.$$

Now consider the case $\alpha \ge \ln^2 n$. The number of exits from v at times $1, \dots, T-1$ is at most the number of entries to v at times $1, \dots, T-2$ plus 1. Recalling (P2), at any time $i \in [T-3]$, the probability to enter v at time $i+1$ is at most $1/d_G(X_i) \le \ln^{-2.5} n$. Thus, the number of exits from v is stochastically dominated by a binomial random variable with $T-1$ trials and success probability $\ln^{-2.5} n$. Thus,

$$\mathbb{P}\left(\bigcup_{1 \le i < T} I_Y(v,t) \mid I_Y(v,0)\right) \le \mathbb{P}\left(\sum_{t=1}^{T-1} \mathbf{1}_{I_Y(v,t)} \ge 1\right)$$
$$\le (T-1)\ln^{-2.5} n \le \ln^{-1/2} n,$$

and the claim follows. $\square$

For a walk W on G and a vertex $v \in [n]$, let $M_W(v) = \sum_{t=0}^{T-1} \mathbf{1}_{I_W(v,t)}$.

Lemma 3.6. *For every vertex $v \in [n]$, $\mathbb{P}(M_Y(v) \ge 1) \ge \frac{T}{2n}(1 - 6\alpha^{-1/2})$.*

Proof. Fix $v \in [n]$. It follows from (P2) that

$$\mathbb{E}(M_Y(v)) = \sum_{t=0}^{T-1} \mathbb{P}(I_Y(v,t)) = T \cdot \mathbb{P}(Y_0 = v \wedge Y_1 \ne v) = T \cdot \mathbb{P}(Y_0 = v)\mathbb{P}(Y_1 \ne Y_0)$$
$$= \frac{T \cdot \pi_v}{2} = \frac{T \cdot d(v)}{4|E(G)|} \ge \frac{T \cdot \big(1 - 2\alpha^{-1/2}\big)}{2n \cdot \big(1 + 2\alpha^{-1/2}\big)} \ge \frac{T}{2n} \cdot \Big(1 - 5\alpha^{-1/2}\Big).$$

Thus by Lemma 3.5,

$$\frac{T}{2n}\Big(1 - 5\alpha^{-1/2}\Big) \le \mathbb{E}(M_Y(v))$$
$$= \sum_{i=1}^{\infty} i\mathbb{P}(M_Y(v) = i)$$
$$\le \mathbb{P}(M_Y(v) = 1) \sum_{i=1}^{\infty} i\Big(\ln^{-1/2} n\Big)^{i-1}$$
$$= \mathbb{P}(M_Y(v) = 1)\Big(1 - \Big(\ln^{-1/2} n\Big)\Big)^{-2}.$$

It follows that

$$\mathbb{P}(M_Y(v) \ge 1) \ge \mathbb{P}(M_Y(v) = 1) \ge \frac{T}{2n}\Big(1 - 6\alpha^{-1/2}\Big). \qquad \square$$

Let

$$b = 3601 \ln n.$$

Corollary 3.7. *Let $t \geq 0$. Conditioned on $X_0, \ldots, X_t$, for every vertex $v \in [n]$,*

$$\mathbb{P}\Big(M_{(X_{t+b+s})_{s\geq 0}}(v) \geq 1\Big) \geq \frac{T}{2n}(1 - 6\alpha^{-1/2}) - \frac{1}{n}.$$

Proof. According to Lemma 3.6 and Corollary 2.11,

$$\mathbb{P}\Big(M_{(X_{t+b+s})_{s\geq 0}}(v) \geq 1\Big) \geq \mathbb{P}\Big(M_{(Y_s)_{s\geq 0}}(v) \geq 1\Big) - \frac{1}{n} \geq \frac{T}{2n}\Big(1 - 6\alpha^{-1/2}\Big) - \frac{1}{n}. \qquad \square$$

Proof of Theorem 3.3. Consider dividing the L_2 steps of the walk X into segments of length $T+1$ with buffers of length b between them (and before the first). Formally, the k'th segment is the walk

$$\Big(X_s^{(k)}\Big)_{s=0}^{T} = \big(X_{(k-1)(T+1)+kb+s}\big)_{s=0}^{T}.$$

It follows from Corollary 3.7 that independently between the segments, for a given v,

$$\mathbb{P}(M_{X^{(k)}}(v) \geq 1) \geq \frac{T}{2n}\Big(1 - 6\alpha^{-1/2}\Big) - \frac{1}{n}.$$

Thus, $\nu(v)$ stochastically dominates a binomial random variable with $\lfloor L_2/(T+1+b)\rfloor$ trials and success probability $\frac{T}{2n}(1-6\alpha^{-1/2}) - \frac{1}{n}$. Let

$$\mu = \lfloor L_2/(T+1+b)\rfloor \cdot \left(\frac{T}{2n}\Big(1 - 6\alpha^{-1/2}\Big) - \frac{1}{n}\right),$$

and note that

$$\mu \sim \frac{2+\varepsilon}{2} \ln n \cdot \Big(1 - 6\alpha^{-1/2}\Big).$$

Thus, letting $\eta = (1-6\alpha^{-1/2})(1+\varepsilon/2)$, by Corollary 2.3,

$$\begin{aligned}\mathbb{P}(\nu(v) < \rho \ln n) = \mathbb{P}\left(\nu(v) < \frac{\rho}{\eta}\cdot \mu(1+o(1))\right) &\leq \exp\big(-\mu(1+o(1))(1-e^{-c} - \rho c/\eta)\big) \\ &\leq \exp\big(-\ln n(1+o(1))(\eta(1-e^{-c}) - \rho c)\big),\end{aligned}$$

for every $c > 0$. Since $\alpha \geq 1500\varepsilon^{-2}$ we have that $\eta \geq 1 + \varepsilon/3$, and taking $c > \ln(13/\varepsilon)$ gives $\eta(1-e^{-c}) > 1 + \varepsilon/4$, and finally taking $\rho < \varepsilon/20c$ gives

$$\mathbb{P}(\nu(v) < \rho \ln n) = o\Big(n^{-1+\varepsilon/5}\Big).$$

The union bound yields

$$\mathbb{P}(\exists v \in [n] : \ \nu(v) < \rho \ln n) \leq n \cdot \mathbb{P}(\nu(v) < \rho \ln n) = o(1),$$

and that concludes the proof. $\square$

3.2 Expansion of small sets in the trace graph

Let $\Gamma = \Gamma^{v_0}_{L_2}(G)$ be the trace of X.

Theorem 3.8. *There exists $\beta > 0$ such that **whp** for every set $A \subseteq [n]$ with $|A| = a \leq n/\ln n$, it holds that $|N_\Gamma(A)| \geq \beta \cdot a \ln n$.*

Proof. We prove the theorem assuming that for every $v \in [n]$, $\nu(v) \geq \rho \ln n$, and recall that this event occurs, according to Theorem 3.3, **whp**. Let $K > 0$ be large enough so that (P4) holds. Suppose first that A is such that $a \geq \frac{n}{\alpha \ln n}$. Let

$$B_0 = \{u \in N_G(A) \mid |E_G(u, A)| \geq Ka\alpha \ln n/n\},$$

and let

$$A_0 = \{v \in A \mid |E_G(v, B_0)| \geq 2\alpha \ln n/K\}.$$

According to (P4), $|A_0| \leq \frac{a}{2}$. Let $A_1 = A \setminus A_0$. For a vertex $v \in A_1$, let $\gamma(v)$ count the number of moves (if any) the walk has made from v to B_0 in the first $\rho \ln n$ exits from v. By (P2),

$$\mathbb{P}\left(\gamma(v) \geq \frac{\rho \ln n}{2}\right) \leq \binom{\rho \ln n}{\rho \ln n/2}\left(\frac{|E_G(v, B_0)|}{d_G(v)}\right)^{\rho \ln n/2} \leq \left(\frac{15}{K}\right)^{\rho \ln n/2}.$$

Let $A_2 \subseteq A_1$ be the set of vertices $v \in A_1$ with $\gamma(v) \geq \rho \ln n/2$. We have that

$$\mathbb{P}\left(|A_2| \geq \frac{a}{4}\right) \leq \binom{a}{a/4}\mathbb{P}\left(\gamma(v) \geq \frac{\rho \ln n}{2}\right)^{a/4} \leq \left(\frac{16}{K}\right)^{a\rho \ln n/8}.$$

Note that for large enough K, $\binom{n}{a}\left(\frac{16}{K}\right)^{a\rho \ln n/8} = o(n^{-1})$. Set $B_1 = N_G(A) \setminus B_0$, $B_2 = N^+_G(A) \setminus B_0$ and $A_3 = A_1 \setminus A_2$. Fix $B_3 \subseteq B_1$. For $v \in A_3$, let p_v be the probability that a walk which exits v and lands in B_2 will also land in $A \cup B_3$. By (P2) we have that for every $v \in A_3$,

$$p_v \leq \frac{|E_G(v, A \cup B_3)|}{|E_G(v, B_2)|} \leq \frac{|E_G(v, B_3)| + a}{\frac{5}{6}\alpha \ln n - |E_G(v, B_0)|} \leq \frac{|E_G(v, B_3)| + a}{\frac{4}{5}\alpha \ln n},$$

for large enough K. Assuming that $|A_3| \geq \frac{a}{4}$, we have that $|E_G(A_3, B_3)| \leq |B_3| \cdot Ka\alpha \ln n/n$, hence

$$\frac{1}{|A_3|}\sum_{v \in A_3} p_v \leq \frac{4}{a} \cdot \frac{|E_G(A_3, B_3)| + a}{\frac{4}{5}\alpha \ln n} \leq \frac{5K|B_3|}{n} + \frac{5}{\alpha \ln n},$$

hence by the inequality between the arithmetic and geometric means,

$$\prod_{v \in A_3} p_v \leq \left(\frac{5K|B_3|}{n} + \frac{5}{\alpha \ln n}\right)^{|A_3|}.$$

In particular, if $N_\Gamma(A) \subseteq B$ for B with $|B| \leq a\beta \ln n$ then there exists $B_3 \subseteq B_1$ with $|B_3| \leq a\beta \ln n$ such that $N_\Gamma(A_3) \cap B_2 \subseteq A \cup B_3$, and conditioning on $|A_3| \geq a/4$ and making sure $\beta = \beta(K)$ is small enough, the probability of that event is at most

$$\prod_{v \in A_3} (p_v)^{\rho \ln n/2} \leq \left(\frac{5Ka\beta \ln n}{n} + \frac{5}{\alpha \ln n}\right)^{a\rho \ln n/8} \leq \left(\frac{6Ka\beta \ln n}{n}\right)^{a\rho \ln n/8}.$$

Thus, taking the union bound,

$$
\begin{aligned}
&\mathbb{P}(\exists A,\ |A| = a :\ |N_\Gamma(A)| \le a\beta \ln n) \\
&\le \sum_{\substack{A \\ |A|=a}} \mathbb{P}(|N_\Gamma(A)| \le a\beta \ln n) \\
&\le \sum_{\substack{A \\ |A|=a}} \Big[\mathbb{P}\Big(|A_2| \ge \frac{a}{4}\Big) + \mathbb{P}\Big(\exists B, |B| = a\beta \ln n, N_\Gamma(A) \subseteq B \mid |A_3| \ge \frac{a}{4}\Big)\Big] \\
&\le \binom{n}{a}\left(\frac{16}{K}\right)^{a\rho \ln n/8} + \sum_{\substack{A \\ |A|=a}} \sum_{\substack{B \\ |B|=a\beta \ln n}} \mathbb{P}\Big(N_\Gamma(A) \subseteq B \mid |A_3| \ge \frac{a}{4}\Big) \\
&\le o\big(n^{-1}\big) + \binom{n}{a}\binom{n}{a\beta \ln n}\left(\frac{6Ka\beta \ln n}{n}\right)^{a\rho \ln n/8} \\
&\le o\big(n^{-1}\big) + \left[\left(\frac{en}{a\beta \ln n}\right)^{2\beta}\left(\frac{6Ka\beta \ln n}{n}\right)^{\rho/8}\right]^{a \ln n},
\end{aligned}
$$

and we may take $\beta > 0$ to be small enough so that the above expression will tend to 0 faster than $1/n$.

Now consider the case where $a < \frac{n}{\alpha \ln n}$. Let

$$
B_0 = \big\{u \in N_G(A) \;\big|\; |E_G(u, A)| \ge K\big\},
$$

and

$$
A_0 = \big\{v \in A \;\big|\; |E_G(v, B_0)| \ge 2\alpha \ln n / \ln K\big\}.
$$

According to (P4), $|A_0| \le \frac{a}{2}$. Let $A_1 = A \setminus A_0$. Define $\gamma(v)$ for vertices from A_1 as in the first case. It follows that

$$
\mathbb{P}\left(\gamma(v) \ge \frac{\rho \ln n}{2}\right) \le \binom{\rho \ln n}{\rho \ln n/2}\left(\frac{|E_G(v, B_0)|}{d_G(v)}\right)^{\rho \ln n/2} \le \left(\frac{15}{\ln K}\right)^{\rho \ln n/2}.
$$

Let $A_2 \subseteq A_1$ be the set of vertices v with $\gamma(v) \ge \rho \ln n/2$. We have that

$$
\mathbb{P}\Big(|A_2| \ge \frac{a}{4}\Big) \le \binom{a}{a/4}\mathbb{P}\left(\gamma(v) \ge \frac{\rho \ln n}{2}\right)^{a/4} \le \left(\frac{16}{\ln K}\right)^{a\rho \ln n/8}.
$$

Note that for large enough K, $\binom{n}{a}\big(\frac{16}{\ln K}\big)^{a\rho \ln n/8} = o\big(n^{-1}\big)$. Set $B_1 = N_G(A) \setminus B_0$, $B_2 = N_G^+(A) \setminus B_0$ and $A_3 = A_1 \setminus A_2$. Fix $B_3 \subseteq B_1$. For $v \in A_3$, let p_v be the probability that a walk which exits v and lands in B_2 will also land in $A \cup B_3$. We have that for every $v \in A_3$,

$$
p_v \le \frac{|E_G(v, A \cup B_3)|}{|E_G(v, B_2)|} \le \frac{|E_G(v, B_3)| + a}{\frac{5}{6}\alpha \ln n - |E_G(v, B_0)|} \le \frac{|E_G(v, B_2)| + a}{\frac{4}{5}\alpha \ln n},
$$

For large enough K. Assuming that $|A_3| \ge \frac{a}{4}$, we have that $|E_G(A_3, B_3)| \le K|B_3|$, and hence

$$
\frac{1}{|A_3|}\sum_{v \in A_3} p_v \le \frac{4}{a} \cdot \frac{|E_G(A_3, B_3)| + a}{\frac{5}{4}\alpha \ln n} \le \frac{5K|B_3|}{a\alpha \ln n} + \frac{5}{\alpha \ln n},
$$

and by the inequality between the arithmetic and geometric means,

$$\prod_{v\in A_3} p_v \le \left(\frac{5K|B_3|}{a\alpha\ln n} + \frac{5}{\alpha\ln n}\right)^{|A_3|}.$$

In particular, if $N_\Gamma(A) \subseteq B$ for B with $|B| \le a\beta\ln n$ then there exists $B_3 \subseteq B_1$ with $|B_3| \le a\beta\ln n$ such that $N_\Gamma(A_3) \cap B_2 \subseteq A \cup B_3$, and conditioning on $|A_3| \ge a/4$ and making sure $\beta = \beta(K)$ is small enough, the probability of that event is at most

$$\prod_{v\in A_3} (p_v)^{\rho\ln n/2} \le \left(\frac{5K\beta}{\alpha} + \frac{5}{\alpha\ln n}\right)^{a\rho\ln n/8} \le \left(\frac{6K\beta}{\alpha}\right)^{a\rho\ln n/8}.$$

Recalling (P2), we notice that $|N_G(A)| \le \frac{4}{3}a\alpha\ln n$. Thus, taking the union bound,

$$\begin{aligned}
&\mathbb{P}(\exists A,\ |A| = a:\ |N_\Gamma(A)| \le a\beta\ln n)\\
&\le \sum_{\substack{A\\|A|=a}} \mathbb{P}(|N_\Gamma(A)| \le a\beta\ln n)\\
&\le \sum_{\substack{A\\|A|=a}} \left[\mathbb{P}\Big(|A_2| \ge \frac{a}{4}\Big) + \mathbb{P}\Big(\exists B, |B| = a\beta\ln n, N_\Gamma(A) \subseteq B \mid |A_3| \ge \frac{a}{4}\Big)\right]\\
&\le \binom{n}{a}\left(\frac{16}{\ln K}\right)^{a\rho\ln n/8} + \sum_{\substack{A\\|A|=a}}\ \sum_{\substack{B\subseteq N_G(A)\\|B|=a\beta\ln n}} \mathbb{P}\Big(N_\Gamma(A) \subseteq B \mid |A_3| \ge \frac{a}{4}\Big)\\
&\le o\big(n^{-1}\big) + \binom{n}{a}\binom{\frac{4}{3}a\alpha\ln n}{a\beta\ln n}\left(\frac{6K\beta}{\alpha}\right)^{a\rho\ln n/8}\\
&\le o\big(n^{-1}\big) + \left[\left(\frac{en}{a}\right)^{1/\ln n}\left(\frac{4\alpha}{\beta}\right)^{\beta}\left(\frac{6K\beta}{\alpha}\right)^{\rho/8}\right]^{a\ln n},
\end{aligned}$$

and we may take $\beta > 0$ to be small enough so that the last expression will tend to 0 faster than $1/n$. Taking the union bound over all cardinalities $1 \le a \le n/\ln n$ implies that the claim holds **whp** in both cases. □

3.3 Edges between large sets in the trace graph

Theorem 3.9. *With high probability, there is an edge of Γ between every pair of disjoint subsets $A, B \subseteq [n]$ satisfying $|A|, |B| \ge \frac{n(\ln\ln n)^{1.5}}{\ln n}$.*

Proof. For each vertex $v \in [n]$ and integer $k \ge 0$, let $x_v^k \sim \mathrm{Unif}(N_G(v))$, independently of each other. Let $\nu_t(v)$ be the number of exits from vertex v by the time t. Think of the random walk X_t as follows:

$$X_{t+1} = \mathrm{Unif}\Big(\Big\{X_t, x_{X_t}^{\nu_t(X_t)}\Big\}\Big).$$

That is, with probability 1/2, the walk stays, and with probability 1/2 it goes to a uniformly chosen vertex from $N_G(v)$, independently from all previous choices. Let $\Lambda = n(\ln\ln n)^{1.5}/\ln n$, and fix two disjoint A, B with $|A| = |B| = \Lambda$. Denote

$$B' = \left\{v \in B \;\middle|\; |E(v, A)| \ge \frac{\Lambda\alpha\ln n}{2n}\right\},$$

and recall that according to property (P5) of an α-pseudo-random graph, $|B'| \geq \frac{\Lambda}{2}$. Note also that according to (P2), for every $v \in B'$

$$\frac{|A \cap N_G(v)|}{|N_G(v)|} \geq \frac{\Lambda \alpha \ln n}{2nd_G(v)} \geq \frac{3\Lambda}{8n}.$$

Let $E_{B,A}$ be the event "the walk has exited each of the vertices in B at least $\rho \ln n$ times, but has not traversed an edge from B to A", and for $v \in B'$, $k > 0$, denote by $F_{v,k,A}$ the event "$\forall 0 \leq i < k$, $x_v^i \notin A$". Clearly, $E_{B,A} \subseteq \bigcap_{v \in B'} F_{v,\rho \ln n, A}$, and $F_{v,\rho \ln n, A}$ are mutually independent for distinct vertices v, hence

$$\mathbb{P}(E_{B,A}) \leq \prod_{v \in B'} \mathbb{P}(F_{v,\rho \ln n, A}) \leq \left(1 - \frac{3\Lambda}{8n}\right)^{\rho \ln n \cdot \Lambda/2}$$

Let E be the event "there exist two disjoint sets A, B of size Λ such that the walk X has exited each of the vertices of B at least $\rho \ln n$ times, but has not traversed an edge from B to A". We have that

$$\begin{aligned}
\mathbb{P}(E) &\leq \sum_{\substack{A \subseteq [n] \\ |A| = \Lambda}} \sum_{\substack{B \subseteq [n] \\ |B| = \Lambda}} \mathbb{P}(E_{B,A}) \\
&\leq \binom{n}{\Lambda}^2 \left(1 - \frac{3\Lambda}{8n}\right)^{\rho \ln n \cdot \Lambda/2} \\
&\leq \left(\frac{en}{\Lambda}\right)^{2\Lambda} \left(1 - \frac{3\Lambda}{8n}\right)^{\rho \ln n \cdot \Lambda/2} \\
&\leq \exp\left(2\Lambda \ln (en/\Lambda) - \frac{3\rho \Lambda^2 \ln n}{16n}\right) \\
&\leq \exp\left(3 \cdot \frac{n}{\ln n} (\ln \ln n)^{2.5} - \frac{n}{\ln n} (\ln \ln n)^{2.9}\right) = o(1).
\end{aligned}$$

Finally, let E' be the event "there exist two disjoint sets A, B of size Λ such that the walk X has not traversed an edge from B to A". It follows from Theorem 3.3 that

$$\begin{aligned}
\mathbb{P}\big(E'\big) &= \mathbb{P}\big(E',\ \forall v \in [n] : \nu(v) \geq \rho \ln n\big) + \mathbb{P}\big(E',\ \exists v \in [n] : \nu(v) < \rho \ln n\big) \\
&= \mathbb{P}(E) + o(1) = o(1),
\end{aligned}$$

and this completes the proof. □

4 Hamiltonicity and vertex connectivity

This short section is devoted to the proof of Theorem 1, which is a simple corollary of the results presented in the previous sections. In addition to these results, we will use the following Hamiltonicity criterion by Hefetz et al:

Lemma 4.1 ([17], Theorem 1.1)**.** *Let $12 \leq d \leq e^{\sqrt[3]{\ln n}}$ and let G be a graph on n vertices satisfying properties (Q1), (Q2) below:*

(Q1) *For every $S \subseteq [n]$, if $|S| \leq \frac{n \ln \ln n \ln d}{d \ln n \ln \ln \ln n}$, then $|N(S)| \geq d|S|$;*

(Q2) *There is an edge in G between any two disjoint subsets $A, B \subseteq [n]$ such that $|A|, |B| \geq \frac{n \ln \ln n \ln d}{4130 \ln n \ln \ln \ln n}$.*

Then G is Hamiltonian, for sufficiently large n.

Proof of Theorem 1. Noting that Theorem 3.1 follows from Theorems 3.8 and 3.9, and setting $d = \ln^{1/2} n$ in the above lemma, we see that its conditions are typically met by the trace Γ_L, $L = (1+\varepsilon)n \ln n$, with much room to spare actually. Hence Γ_L is **whp** Hamiltonian.

Theorem 3.1 also implies that **whp** Γ_L is an $(n/\ln n, \beta \ln n)$-expander, for some $\beta > 0$, and in addition, that there is an edge connecting every two disjoint sets with cardinality at least $n(\ln\ln n)^{1.5}/\ln n$. Set $k = \beta \ln n$, and suppose to the contrary that under these conditions, Γ_L^v is not k-connected. Thus, there is a cut $S \subseteq [n]$ with $|S| \le k-1$ such that $[n] \smallsetminus S$ can be partitioned into two non-empty sets, A, B, with no edge connecting them. Without loss of generality, assume $|A| \le |B|$. If $|A| \le n/\ln n$ then $k \le \beta \ln n |A| \le |N(A)| \le |S| < k$, a contradiction. If $n/\ln n < |A| < \beta n - k + 1$ then take $A_0 \subseteq A$ with $|A_0| = n/\ln n$, and then $\beta n \le |N(A_0)| \le |A \cup S| < \beta n$, again a contradiction. Finally, if $|A| \ge \beta n - k + 1$ then A, B are both of size at least $n(\ln\ln n)^{1.5}/\ln n$, thus there is an edge connecting the two sets, again a contradiction. $\square$

5 Hitting time results for the walk on K_n

From this point on, a *lazy* random walk on K_n is a walk which starts at a uniformly chosen vertex, and at any given step, stays at the current vertex with probability $1/n$. Of course, this does not change matters much, and the random walk of the theorem, including its cover time, can be obtained from the lazy walk by simply ignoring loops. Considering the lazy version makes things much more convenient; observe that for any $t \ge 0$, the modified random walk is equally likely to be located at any of the vertices of K_n after t steps, regardless of its history. Hence, for any t, if we look at the trace graphs Γ_t^{o} and Γ_t^{e} formed by the edges (including loops) traversed by the lazy walk at its odd, respectively even, steps, they are mutually independent, and the graphs formed by them are distributed as $\hat{G}(n,m)$ with $m = \lceil t/2 \rceil$ and $m = \lfloor t/2 \rfloor$, respectively, where $\hat{G}(n,m)$ is the random (multi)graph obtained by drawing independently m edges (with replacement) from all possible *directed* edges (and loops) of the complete graph K_n, and then ignoring the orientations. Note that whenever $m = o\big(n^2\big)$, the probability of a given non-loop edge to appear in $\hat{G}(n,m)$ is $\sim 2m/n^2$.

Let now for $k \ge 1$,

$$t_-^{(k)} = n(\ln n + (k-1)\ln\ln n - \ln\ln\ln n), \tag{4}$$

$$t_+^{(k)} = n(\ln n + (k-1)\ln\ln n + \ln\ln\ln n). \tag{5}$$

We may as well just write t_- or t_+, when k is clear or does not matter. Recall the definition of τ_C^k from (1). The following is a standard result on the coupon collector problem:

Theorem 5.1 (Proved in [12]). *For every $k \ge 1$, **whp**, $t_-^{(k)} < \tau_C^k < t_+^{(k)}$.*

To ease notations, we shall denote $\Gamma_+ = \Gamma_{t_+^{(k)}}$ and similarly $\Gamma_- = \Gamma_{t_-^{(k)}}$. We add a superscript o or e to consider the odd, respectively even, steps only. We denote the edges of the walk by $\{e_i \mid i > 0\}$.

We note that the trace of our walk is typically not a graph, but rather a multigraph. However, that fact does not matter much, as the multiplicity of the edges of that multigraph is typically well bounded, as the following lemma shows:

Lemma 5.2. *With high probability, the multiplicity of any edge of Γ_+ is at most 4.*

Proof. Suppose the multiplicity of an edge e in Γ_+ is greater than 4; in that case, its multiplicity in Γ_+^{o} or in Γ_+^{e} is at least 3. As $\Gamma_+^{\mathrm{o}} \sim \hat{G}(n, \lceil t_+/2 \rceil)$, we have that the probability for that to happen is $O(t_+^3/n^6) = o(n^{-2})$. Applying the union bound over all possible edges gives the desired result for the odd case (and the even case is identical). $\square$

5.1 k-connectivity

Clearly, if a given vertex has been visited at most $k-1$ times, or has been visited k times without exiting the last time, its degree in the trace is below $2k-1$ or $2k$ respectively, hence $\tau_C^k \le \tau_\delta^{2k-1}$ and $\tau_C^k + 1 \le \tau_\delta^{2k}$; furthermore, if some vertex has a (simple) degree less than m, then removing all of its neighbours from the graph will disconnect it, hence it is not m-vertex-connected, thus $\tau_\delta^m \le \tau_\kappa^m$. To prove Theorem 4 it therefore suffices to prove the following two claims:

Claim 5.3. *For any constant integer $k \ge 1$, **whp** $\tau_C^k \ge \tau_\delta^{2k-1}$ and $\tau_C^k + 1 \ge \tau_\delta^{2k}$.*

Claim 5.4. *For any constant integer $m \ge 1$, **whp** $\tau_\delta^m \ge \tau_\kappa^m$.*

5.1.1 The set SMALL

To argue about the relation between the number of visits of a vertex and its degree, we would wish to limit the number of loops and multiple edges incident to a vertex. This can be easily achieved for small degree vertices, which are the only vertices that may affect the minimum degree anyway. This gives motivation for the following definition.

Denote $d_0 = \lfloor \delta_0 \ln n \rfloor$ for $\delta_0 = e^{-20}$.

$$\mathrm{SMALL} = \left\{ v \in [n] \mid d_{\Gamma_-^{\mathrm{o}}}(v) < d_0 \right\}$$

be the set of all small degree vertices of Γ_-^{o}. Note that the exact value of δ_0 is not important. We will simply need it to be small enough for the proof of Lemma 5.20.

Lemma 5.5. *Let $m \sim n \ln n/2$, $\hat{G} \sim \hat{G}(n,m)$ and $v \in [n]$. Then, $\mathbb{P}\big(d_{\hat{G}}(v) < d_0\big) \le n^{-0.9}$.*

Proof. Noting that $d_{\hat{G}}(v)$ is distributed binomially with $2m$ trials and success probability $1/n$, we invoke Corollary 2.3 with $c = 10$ to obtain

$$\mathbb{P}\big(d_{\hat{G}}(v) < d_0\big) \le \exp\big(-\ln n\big(1 - e^{-10} - 10\delta_0\big)(1 + o(1))\big) \le n^{-0.9}. \qquad \square$$

An application of Markov's inequality (since $t_- \sim n \ln n$) gives the following:

Corollary 5.6. *With high probability, $|\mathrm{SMALL}| \le n^{0.2}$.*

Lemma 5.7. *With high probability, no vertex in* SMALL *is incident to a loop or to a multiple edge in Γ_+.*

Proof. Let L_v^i be the event "v is incident to a loop in Γ_+ which is the i'th step of the random walk". Note that we consider loops in Γ_+, which need not be in Γ_+^{o}. Fix a vertex v and assume it is incident to a loop in Γ_+. Take i such that the i'th step of X_t, e_i, is a loop incident to v (that is, $X_{i-1} = X_i = v$). Let $\hat{G}$ be the graph obtained from Γ_+^{o} by removing e_j for every odd $i-1 \le j \le i+1$. It is clear then that $\hat{G}$ is distributed like $\hat{G}(n,m)$ with $t_+/2 - 2 \le m \le t_+/2 - 1$

(so $m \sim n\ln n/2$), and it is independent of the event L_v^i. Noting that $\mathbb{P}\big(L_v^i\big) = n^{-2}$ and using Lemma 5.5 we conclude that

$$\mathbb{P}\big(L_v^i, v \in \text{SMALL}\big) \le \mathbb{P}\big(L_v^i, d_{\hat{G}}(v) < d_0\big) = \mathbb{P}\big(L_v^i\big)\mathbb{P}\big(d_{\hat{G}}(v) < d_0\big) \le n^{-2.9},$$

and by applying the union bound over all vertices and over all potential times for loops at a vertex we obtain the following upper bound for the existence of a vertex from SMALL which is incident to a loop:

$$\mathbb{P}\big(\exists v \in [n],\ i \in [t_+] :\ L_v^i,\ v \in \text{SMALL}\big) \le n \cdot t_+ \cdot n^{-2.9} = o(1).$$

Using a similar method, we can show that **whp** there is no vertex in SMALL which is incident to a multiple edge in Γ_+, and this completes the proof. □

Lemma 5.8. *With high probability, for every pair of disjoint vertex subsets $U, W \subseteq [n]$ of size $|U| = |W| = n/\ln^{1/2} n$, Γ_-^{o} has at least $0.5n$ edges between U and W.*

Proof. We note that $\Big|E_{\Gamma_-^{\mathrm{o}}}(U,W)\Big|$ is distributed binomially with $\lceil t_-/2\rceil$ trials and success probability $p = \frac{n^2}{\ln n}\binom{n+1}{2}^{-1}$. As $p > 1.9/\ln n$, using the Chernoff bounds we have that

$$\begin{aligned}\mathbb{P}\Big(\Big|E_{\Gamma_-^{\mathrm{o}}}(U,W)\Big| < 0.5n\Big) &\le \mathbb{P}(\mathrm{Bin}(\lceil t_-/2\rceil, 1.9/\ln n) < 0.5n) \\ &\le \mathbb{P}(\mathrm{Bin}(n\ln n/1.9, 1.9/\ln n) \le n - 0.5n) \le e^{-0.1n},\end{aligned}$$

thus by the union bound

$$\begin{aligned}\mathbb{P}\Big(\exists U, W : \Big|E_{\Gamma_-^{\mathrm{o}}}(U,W)\Big| < 0.5n\Big) &\le \binom{n}{n/\ln^{1/2} n}^2 e^{-0.1n} \\ &\le \big(e^2 \ln n\big)^{n/\ln^{1/2} n} e^{-0.1n} \\ &\le \exp\bigg(\frac{n}{\ln^{1/2} n}(2 + \ln\ln n) - 0.1n\bigg) = o(1).\end{aligned}$$ □

5.1.2 Extending the trace

Define

$$\Gamma_* = \Gamma_-^{\mathrm{o}} + \Big\{e_i \mid 1 \le i \le \tau_C^k + 1, e_i \cap \text{SMALL} \ne \varnothing\Big\}.$$

Lemma 5.9. *With high probability, $\delta(\Gamma_*) \ge 2k$.*

Proof. Let v be a vertex. If $v \notin \text{SMALL}$ then $d(v) \ge d_0$ hence **whp** $d'(v) \ge (d_0 - 8)/4 \ge 2k$ (according to Lemma 5.2). On the other hand, if $v \in \text{SMALL}$, and is not the first vertex of the random walk, then it was entered and exited at least k times in the first $\tau_C^k + 1$ steps of the random walk. By the definition of Γ_*, all of these entries and exits are in $E(\Gamma_*)$. Since **whp** none of the vertices in SMALL is incident to loops or multiple edges (according to Lemma 5.7), the simple degree of each such vertex is at least $2k$.

Noting that **whp** the first vertex of the random walk is not in SMALL (according to Lemma 5.5), we obtain the claim. □

We note that by deleting the edge $e_{\tau_C^k+1}$ from Γ_* its minimum degree cannot drop by more than one, so Claim 5.3 follows from Lemma 5.9.

Lemma 5.10. *With high probability, $\Delta(\Gamma_*) \le 6\ln n$.*

Proof. Fix a vertex v. Noting that $d_{\Gamma_-^\circ}(v)$ is binomially distributed with mean $t_-/n \sim \ln n$, we invoke Corollary 2.3 with $c = -1$ to obtain

$$\mathbb{P}\Big(d_{\Gamma_-^\circ}(v) > 3\ln n\Big) \le \exp(-\ln n(1-e+3)(1+o(1))) = O\big(n^{-1.2}\big).$$

Similarly one can derive $\mathbb{P}\Big(d_{\Gamma_+^e}(v) > 3\ln n\Big) = O\big(n^{-1.2}\big)$. Since $d'_{\Gamma_*}(v) \le d_{\Gamma_-^\circ}(v) + d_{\Gamma_+^e}(v)$ we have that $\mathbb{P}\big(d'_{\Gamma_*}(v) > 6\ln n\big) = O\big(n^{-1.2}\big)$. The union bound over all vertices gives that $\mathbb{P}(\Delta(\Gamma_*) > 6\ln n) = o(1)$, as we have wished to show. $\square$

Lemma 5.11. *Fix $\ell \ge 1$. With high probability there is no path of length between 1 and ℓ in Γ_* such that both of its (possibly identical) endpoints lie in* SMALL*.*

Proof. For a set $T \subseteq [t_+]$ let $r(T)$ be the minimum number of integer intervals whose union is the set of elements from T. In symbols,

$$r(T) = |\{1 \le i \le t_+ \mid i \in T \wedge i+1 \notin T\}|.$$

Fix $\ell \ge 1$ and $P = (v_0, \dots, v_\ell)$, a path of length ℓ. Suppose first that $v_0 \ne v_\ell$. Let A be the event $P \subseteq E(\Gamma_+)$. For every set $T = \{s_1, \dots, s_\ell\} \subseteq [t_+]$ with $s_1 < s_2 < \dots < s_\ell$, let A_T be the event "$\forall j \in [\ell]$, $e_{s_j} = \{v_{j-1}, v_j\}$". We have that

$$\begin{aligned}
\mathbb{P}(A,\ v_0, v_\ell \in \text{SMALL}) &\le \sum_{T \in \binom{[t_+]}{\ell}} \mathbb{P}(A_T,\ v_0, v_\ell \in \text{SMALL}) \\
&= \sum_{r=1}^{\ell} \sum_{\substack{T \in \binom{[t_+]}{\ell} \\ r(T)=r}} \mathbb{P}(A_T,\ v_0, v_\ell \in \text{SMALL}).
\end{aligned}$$

For every set $T \in \binom{[t_+]}{\ell}$, let

$$I_T = \{i \in [t_-] \mid i \text{ is odd},\ \nexists s \in T:\ |i - s| \le 1\},$$

and for a vertex $v \in [n]$, let $d_{I_T}(v)$ be the degree of v in the graph formed by the edges $\{e_i \mid i \in I_T\}$. Let $D_T(v)$ be the event "$d_{I_T}(v) \le d_0$". It follows from the definition of I_T that $D_T(v_0)$ and $D_T(v_\ell)$ are independent of the event A_T. Moreover, if $v \in$ SMALL then $D_T(v)$ (since $d_{I_T}(v) \le d_{\Gamma_-^\circ}(v)$), and as there is exactly one edge of K_n connecting v_0 with v_ℓ, conditioning on the event $D_T(v_0)$ cannot increase the probability of the event $D_T(v_\ell)$ by much:

$$\begin{aligned}
\mathbb{P}(D_T(v_0), D_T(v_\ell)) &\le \mathbb{P}(D_T(v_0), D_T(v_\ell) \mid \{v_0, v_\ell\} \notin I_T) \\
&= \mathbb{P}(D_T(v_0) \mid \{v_0, v_\ell\} \notin I_T)\mathbb{P}(D_T(v_\ell) \mid \{v_0, v_\ell\} \notin I_T) \\
&\le \mathbb{P}(D_T(v_0))\mathbb{P}(D_T(v_\ell)) \cdot \frac{1}{(\mathbb{P}(\{v_0, v_\ell\} \notin I_T))^2} \\
&= \mathbb{P}(D_T(v_0))\mathbb{P}(D_T(v_\ell))(1+o(1)) \le n^{-1.7},
\end{aligned}$$

here we have used Lemma 5.5, and the fact that $|I_T| \sim n \ln n/2$. Thus, for a fixed T,

$$\begin{aligned}\mathbb{P}(A_T,\ v_0, v_\ell \in \text{SMALL}) &\leq \mathbb{P}(A_T,\ D_T(v_0),\ D_T(v_\ell)) \\ &= \mathbb{P}(A_T)\mathbb{P}(D_T(v_0), D_T(v_\ell)) \\ &\leq \mathbb{P}(A_T) \cdot n^{-1.7}.\end{aligned}$$

Similarly, if $v_0 = v_\ell$ we obtain $\mathbb{P}(A_T, v_0 \in \text{SMALL}) \leq \mathbb{P}(A_T) \cdot n^{-0.9}$. Now, given T with $r(T) = r$ ($1 \leq r \leq \ell$), the probability of A_T is at most $n^{-(\ell+r)}$. It may be 0, in case T is not feasible, and otherwise there are exactly $\ell + r$ times where the walk is forced to be at a given vertex (the walk has to start each of the r intervals at a given vertex, and to walk according to the intervals ℓ steps in total), and the probability for each such restriction is $1/n$. The number of T's for which $r(T) = r$ is $O((t_+)^r)$ (choose r points from $[t_+]$ to be the starting points of the r intervals; then for every $j \in [\ell]$ there are at most $r\ell$ options for the j'th element of T). Noting that the number of paths of length ℓ is no larger than $n^{\ell+1}$ if $v_0 \neq v_\ell$, or n^ℓ if $v_0 = v_\ell$, the union bound gives

$$\begin{aligned}&\mathbb{P}(\exists P = (v_0, \ldots, v_\ell) :\ A,\ v_0, v_\ell \in \text{SMALL}) \\ &\leq n^{\ell+1} \sum_{r=1}^{\ell} \frac{O((t_+)^r)}{n^{\ell+r}} \cdot n^{-1.7} + n^\ell \sum_{r=1}^{\ell} \frac{O((t_+)^r)}{n^{\ell+r}} \cdot n^{-0.9} \\ &\leq n^{-0.7} \sum_{r=1}^{\ell} O(\ln^r n) = o(1).\end{aligned}$$

□

Lemma 5.12. *With high probability, every vertex set U with $|U| \leq n/\ln^{1/2} n$ spans at most $2|U| \cdot \ln^{3/4} n$ edges (counting multiple edges and loops) in Γ_*.*

Proof. Fix $U \subseteq [n]$ with $|U| = u \leq n/\ln^{1/2} n$. Let $e^{\mathrm{o}}(U)$ and $e^{\mathrm{e}}(U)$ be the number of edges (including multiple edges and loops) spanned by U in Γ_+^{o} and Γ_+^{e} respectively. Note that $e^{\mathrm{o}}(U)$ is binomially distributed with $\lceil t_+/2 \rceil$ trials and success probability u^2/n^2. Thus, using Claim 2.4 we have that

$$\mathbb{P}\Big(e^{\mathrm{o}}(U) > u \ln^{3/4} n\Big) \leq \left(\frac{e t_+ u^2}{2n^2 u \ln^{3/4} n}\right)^{u \ln^{3/4} n} \leq \left(\frac{e \ln^{1/4} n u}{n}\right)^{u \ln^{3/4} n}.$$

The union bound over all choices of U yields

$$\begin{aligned}\mathbb{P}\left(\exists U,\ |U| \leq \frac{n}{\ln^{1/2} n},\ e^{\mathrm{o}}(U) \geq |U| \ln^{3/4} n\right) &\leq \sum_{u=1}^{n/\ln^{1/2} n} \binom{n}{u} \left(\frac{e \ln^{1/4} n u}{n}\right)^{u \ln^{3/4} n} \\ &\leq \sum_{u=1}^{n/\ln^{1/2} n} \left(\frac{en}{u} \cdot \left(\frac{e \ln^{1/4} n u}{n}\right)^{\ln^{3/4} n}\right)^u.\end{aligned}$$

We now split the sum into two:

$$\sum_{u=1}^{\ln n} \left(\frac{en}{u}\left(\frac{e \ln^{1/4} n u}{n}\right)^{\ln^{3/4} n}\right)^u \leq \ln n \cdot en \left(\frac{e \ln^{5/4} n}{n}\right)^{\ln^{3/4} n} = o(1),$$

and

$$\sum_{u=\ln n}^{n/\ln^{1/2} n}\left(\frac{en}{u}\left(\frac{e\ln^{1/4} n u}{n}\right)^{\ln^{3/4} n}\right)^{u} = \sum_{u=\ln n}^{n/\ln^{1/2} n}\left(e\left(\frac{u}{n}\right)^{\ln^{3/4} n-1}\left(e\ln^{1/4} n\right)^{\ln^{3/4} n}\right)^{u}$$
$$\leq n\left(e\left(\frac{1}{\ln^{1/2} n}\right)^{\ln^{3/4} n-1}\left(e\ln^{1/4} n\right)^{\ln^{3/4} n}\right)^{\ln n} = o(1).$$

As the same bound applies for $e^{\mathrm{e}}(U)$, the union bound concludes the claim (noting that $\Gamma_* \subseteq \Gamma_+$). $\square$

Lemma 5.13. *With high probability, for every pair of disjoint vertex sets U, W with $|U| \leq n/\ln^{1/2} n$ and $|W| \leq |U| \cdot \ln^{1/4} n$, it holds that $|E_{\Gamma_*}(U, W)| \leq 2|U|\ln^{0.9} n$.*

Proof. For $U, W \subseteq [n]$, $|U| \leq n/\ln^{1/2} n$, $|W| \leq |U|\ln^{1/4} n$, let $e^{\mathrm{o}}(U, W)$ ($e^{\mathrm{e}}(U, W)$) be the number of edges in Γ_+^{o} (in Γ_+^{e}) between U and W. For $\mathrm{x} \in \{\mathrm{o}, \mathrm{e}\}$, let $A^{\mathrm{x}}(U, W)$ be the event "$e^{\mathrm{x}}(U, W) \geq |U|\ln^{0.9} n$", and let A^{x} be the event "$\exists U, W,\ |U| \leq n/\ln^{1/2} n,\ |W| \leq |U|\ln^{1/4} n,\ A^{\mathrm{x}}(U, W)$".

Fix U, W with $|U| = u \leq n/\ln^{1/2} n$ and $|W| = w \leq u\ln^{1/4} n$. Note that $e^{\mathrm{o}}(U, W)$ is binomially distributed with $\lceil t_+/2\rceil$ trials and success probability $2uw/n^2$. Thus, using Claim 2.4 we have that

$$\mathbb{P}\left(e^{\mathrm{o}}(U, W) > u\ln^{0.9} n\right) \leq \left(\frac{et_+uw}{n^2u\ln^{0.9} n}\right)^{u\ln^{0.9} n} \leq \left(\frac{ew\ln^{0.1} n}{n}\right)^{\ln^{0.9} n}.$$

The union bound over all choices of U, W yields

$$\begin{aligned}
\mathbb{P}(A^{\mathrm{o}}) &\leq \sum_{u=1}^{n/\ln^{1/2} n}\sum_{w=1}^{u\ln^{1/4} n}\binom{n}{u}\binom{n}{w}\left(\frac{ew\ln^{0.1} n}{n}\right)^{\ln^{0.9} n}\\
&\leq \sum_{u=1}^{n/\ln^{1/2} n}\sum_{w=1}^{u\ln^{1/4} n}\left(\frac{en}{u}\left(\frac{en}{w}\right)^{w/u}\left(\frac{ew\ln^{0.1} n}{n}\right)^{\ln^{0.9} n}\right)^{u}\\
&\leq \sum_{u=1}^{n/\ln^{1/2} n} u\ln^{1/4} n\left(\frac{en}{u}\left(\frac{en}{u\ln^{1/4} n}\right)^{\ln^{1/4} n}\left(\frac{eu\ln^{0.35} n}{n}\right)^{\ln^{0.9} n}\right)^{u}\\
&\leq \sum_{u=1}^{n/\ln^{1/2} n} u\ln^{1/4} n\left(e\left(\frac{u}{n}\right)^{\ln^{0.9} n-\ln^{1/4} n-1}\left(e\ln^{0.35} n\right)^{\ln^{0.9} n}\left(e\ln^{-1/4} n\right)^{\ln^{1/4} n}\right)^{u}.
\end{aligned}$$

We now split the sum into two:

$$\sum_{u=1}^{\ln n} u\ln^{1/4} n\left(e\left(\frac{u}{n}\right)^{\ln^{0.9} n-\ln^{1/4} n-1}\left(e\ln^{0.35} n\right)^{\ln^{0.9} n}\left(e\ln^{-1/4} n\right)^{\ln^{1/4} n}\right)^{u}$$
$$\leq \ln^{9/4} n\cdot e\left(\frac{\ln n}{n}\right)^{\ln^{0.9} n-\ln^{1/4} n-1}\left(e\ln^{0.35} n\right)^{\ln^{0.9} n}\left(e\ln^{-1/4} n\right)^{\ln^{1/4} n} = o(1),$$

and

$$\sum_{u=\ln n}^{n/\ln^{1/2} n} u \ln^{1/4} n \left(e \left(\frac{u}{n} \right)^{\ln^{0.9} n - \ln^{1/4} n - 1} \left(e \ln^{0.35} n \right)^{\ln^{0.9} n} \left(e \ln^{-1/4} n \right)^{\ln^{1/4} n} \right)^{u}$$

$$\leq n^2 \left(e \left(\frac{1}{\ln^{1/2} n} \right)^{\ln^{0.9} n - \ln^{1/4} n - 1} \left(e \ln^{0.35} n \right)^{\ln^{0.9} n} \left(e \ln^{-1/4} n \right)^{\ln^{1/4} n} \right)^{\ln n} = o(1).$$

As the same bound applies for $\mathrm{x} = \mathrm{e}$, the union bound over $\mathrm{x} \in \{\mathrm{o}, \mathrm{e}\}$ concludes the claim (noting that $\Gamma_* \subseteq \Gamma_+$). □

We will need the following lemma, according to which not too many edges were added by extending the trace, when we will prove the Hamiltonicity of the trace:

Lemma 5.14. *With high probability,* $\left|E(\Gamma_*) \setminus E\left(\Gamma_-^{\mathrm{o}}\right)\right| \leq n^{0.4}$.

Proof. Recall from Corollary 5.6 that **whp** $|\mathrm{SMALL}| \leq n^{0.2}$. From Lemma 5.10 it follows that **whp** $\Delta(\Gamma_*) \leq 6 \ln n$. From Lemma 5.2 it follows that **whp** $d_{\Gamma_*}(v) \leq 4\Delta(\Gamma_*) \leq 24 \ln n$ for every $v \in \mathrm{SMALL}$. We conclude that the number of edges in Γ_* with at least one end in SMALL is **whp** at most $n^{0.2} \cdot 24 \ln n < n^{0.4}$, and the claim follows by the definition of Γ_*. □

5.1.3 Sparsifying the extension

We may use the results of Lemmas 5.8 to 5.13 to show that Γ_* is a (very) good expander. This, together with Lemma 2.7, will imply that Γ_* is $2k$-connected. However, in order to later show that Γ_* is Hamiltonian, we wish to show it contains a much *sparser* expander, which is still good enough to guarantee high connectivity.

To obtain this, we assume Γ_* has the properties guaranteed by these lemmas, and sparsify Γ_* randomly as follows: for each vertex v, if $v \in \mathrm{SMALL}$, define $E(v)$ to be all edges incident to v; otherwise let $E(v)$ be a uniformly chosen subset of size d_0 of all edges incident to v. Let Γ_0 be the spanning subgraph of Γ_* whose edge set is the union of $E(v)$ over all vertices v.

Lemma 5.15. *With high probability (over the choices of $E(v)$), for every pair of disjoint vertex sets $U, W \subseteq [n]$ of size $|U| = |W| = n/\ln^{1/2} n$, Γ_0 has at least one edge between U and W.*

Proof. Let $U, W \subseteq [n]$ with $|U| = |W| = n/\ln^{1/2} n$. From Lemma 5.8 it follows that in $\Gamma = \Gamma_-^{\mathrm{o}}$ there are at least $0.5n$ edges between U and W. If there is a vertex $v \in U \cap \mathrm{SMALL}$ with an edge into W, we are done, so we can assume that there is no such. Let $U' = U \setminus \mathrm{SMALL}$; thus, $|E_\Gamma(U', W)| \geq 0.5n$.

Fix a vertex $u \in U'$. Let X_u be the number of edges between u and W in Γ that fall into $E(u)$. X_u is a random variable, distributed according to $\mathrm{Hypergeometric}(d_\Gamma(u), |E_\Gamma(u, W)|, d_0)$. According to Theorem 2.2, the probability that $X_u = 0$ may be bounded from above by

$$\exp\left(-\frac{|E_\Gamma(u, W)| \cdot d_0}{2 d_\Gamma(u)} \right),$$

which, according to Lemmas 5.2 and 5.10, may be bounded from above by

$$\exp\left(-\frac{|E_\Gamma(u, W)| \cdot d_0}{50 \ln n} \right).$$

Hence, the probability that there is no vertex $u \in U$ from which there exists an edge to W can be bounded from above by

$$\prod_{u \in U'} \exp\left(-\frac{d_0}{50 \ln n} \cdot |E_\Gamma(u, W)|\right) = \exp\left(-\frac{d_0}{50 \ln n} \cdot \left|E_\Gamma(U', W)\right|\right) = \exp(-\Theta(n)).$$

Union bounding over all choices of U, W, we have that the probability that there exists such a pair of sets with no edge between them is at most

$$\binom{n}{n/\ln^{1/2} n}^2 e^{-\Theta(n)} \le \exp\left(\frac{n}{\ln^{1/2} n}(2 + \ln \ln n) - \Theta(n)\right) = o(1). \qquad \square$$

Lemma 5.16. *$\delta(\Gamma_0) \ge 2k$.*

Proof. This follows from Lemma 5.9, since we have not removed any edge incident to a vertex from SMALL and since any other vertex is incident to at least d_0 edges. $\square$

Recall the definition of (R, c)-expanders from Section 2.1.

Lemma 5.17. *With high probability (over the choices of $E(v)$) Γ_0 is a $\left(\frac{n}{2k+2}, 2k\right)$-expander, with at most $d_0 n$ edges.*

Proof. Since by definition $|E(v)| \le d_0$ for every $v \in [n]$, it follows that $|E(\Gamma_0)| \le d_0 n$. Let $S \subseteq [n]$ with $|S| \le n/(2k+2)$. Denote $S_1 = S \cap \text{SMALL}$ and $S_2 = S \setminus \text{SMALL}$. Consider each of the following cases:

Case $|S_2| \ge n/\ln^{1/2} n$: From Lemma 5.15 it follows that the set of all non-neighbours of S_2 (in Γ_0) is of cardinality less than $n/\ln^{1/2} n$. Thus

$$|N_{\Gamma_0}(S)| \ge n - n/\ln^{1/2} n - |S| \ge \frac{(2k+1)n}{2k+2} - n/\ln^{1/2} n \ge \frac{2kn}{2k+2} \ge 2k|S|.$$

Case $|S_2| < n/\ln^{1/2} n$: From Lemmas 5.11 and 5.16 it follows that $|N_{\Gamma_0}(S_1)| \ge 2k|S_1|$. From Lemma 5.12 it follows that S_2 spans at most $2|S_2| \cdot \ln^{3/4} n$ edges in Γ_0. Consequently,

$$|\partial_{\Gamma_0} S_2| \ge d_0|S_2| - 2|E_{\Gamma_0}(S_2)| > |S_2|(d_0 - 4\ln^{3/4} n) \ge 3|S_2| \cdot \ln^{0.9} n,$$

hence, by Lemma 5.13 it holds that $|N_{\Gamma_0}(S_2)| > |S_2| \cdot \ln^{1/4} n$. Finally, by Lemma 5.11 we obtain that for each $u \in S_2$, $\left|N_{\Gamma_0}(S_1) \cap N^+_{\Gamma_0}(u)\right| \le 1$, hence

$$\left|N_{\Gamma_0}(S_1) \cap N^+_{\Gamma_0}(S_2)\right| \le |S_2|,$$

and thus

$$\left|N_{\Gamma_0}(S_1) \setminus N^+_{\Gamma_0}(S_2)\right| \ge 2k|S_1| - |S_2|.$$

Similarly, for each vertex in S_2 has at most one neighbour in S_1, thus

$$|N_{\Gamma_0}(S_2) \setminus S_1| \ge |N_{\Gamma_0}(S_2)| - |S_2| > |S_2| \cdot \ln^{0.2} n.$$

To summarize, we have that

$$\begin{aligned} |N_{\Gamma_0}(S)| &= \left|N_{\Gamma_0}(S_1) \setminus N^+_{\Gamma_0}(S_2)\right| + |N_{\Gamma_0}(S_2) \setminus S_1| \\ &\ge 2k|S_1| - |S_2| + |S_2| \cdot \ln^{0.2} n \\ &\ge 2k(|S_1| + |S_2|) = 2k|S|. \end{aligned} \qquad \square$$

Since Γ_0 is **whp** an (R,c)-expander (with $R(c+1)=\frac{n(2k+1)}{2k+2}\geq\frac{n}{2}+k$), we have that Γ_* is such, and from Lemma 2.7 we conclude it is $2k$-vertex-connected. Claim 5.4 follows for even values of m.

We have already shown (in Claim 5.3) that $\tau_\delta^{2k-1}+1=\tau_C^k+1=\tau_\delta^{2k}$. Hence, using what we have just shown we have that $\tau_\delta^{2k-1}+1=\tau_\kappa^{2k}$. Since removing an edge may decrease connectivity by not more than 1, it follows that $\tau_\delta^{2k-1}\geq\tau_\kappa^{2k-1}$.

That concludes the proof of Claim 5.4 and of Theorem 4.

5.2 Hamiltonicity

We start by describing the tools needed for our proof.

Definition 5.18. *Given a graph G, a non-edge $e=\{u,v\}$ of G is called a* booster *if adding e to G creates a graph $\tilde{G}$, which is either Hamiltonian or whose maximum path is longer than that of G.*

Note that technically every non-edge of a Hamiltonian graph G is a booster by definition. Adding a booster advances a non-Hamiltonian graphs towards Hamiltonicity. Sequentially adding n boosters makes any graph with n vertices Hamiltonian.

Lemma 5.19. *Let G be a connected non-Hamiltonian $(R,2)$-expander. Then G has at least $(R+1)^2/2$ boosters.*

The above is a fairly standard tool in Hamiltonicity arguments for random graphs, based on the Pósa rotation-extension technique [23]. Its proof can be found, e.g., in [6, Chapter 8.2].

We have proved in Lemma 5.17, for $k=1$, that Γ_* (and thus Γ_{τ_C+1}) typically contains a sparse $\left(\frac{n}{4},2\right)$-expander Γ_0. We can obviously assume that Γ_0 does not contain loops or multiple edges. Expanders are not necessarily Hamiltonian themselves, but they are extremely helpful in reaching Hamiltonicity as there are many boosters relative to them by Lemma 5.19. We will thus start with Γ_0 and will repeatedly add boosters to it to bring it to Hamiltonicity. Note that those boosters should come from within the edges of the trace Γ_{τ_C+1}. This is taken care of by the following lemma.

Lemma 5.20. *With high probability, every non-Hamiltonian $\left(\frac{n}{4},2\right)$-expander $H\subseteq\Gamma_*$ with $|E(H)|\leq d_0n+n$ and $\left|E(H)\smallsetminus E(\Gamma_-^{\mathrm{o}})\right|\leq n^{0.4}$ has a booster in Γ_-^{o}.*

Proof. For a non-Hamiltonian $\left(\frac{n}{4},2\right)$-expander H let $H_{\mathrm{o}}=H\cap\Gamma_-^{\mathrm{o}}$ and $H_{\mathrm{e}}=H\smallsetminus H_{\mathrm{o}}$ be two subgraphs of H. Denote by $\mathcal{B}(H)$ the set of boosters with respect to H. At the first stage we choose H. For that, we first choose how many edges H has (at most d_0n+n) and call that quantity i, then we choose the edges from K_n. At the second stage we choose H_{e}. For that, we first choose how many of H's edges are not in Γ_-^{o} (at most $n^{0.4}$) and call that quantity j, then we choose the edges from H. At the third stage, we require all of H_{o}'s edges to appear in Γ_-^{o}. For that, we first choose for each edge of H_{o} a time in which it was traversed, then we actually require that edge to be traversed at that time.

Finally, we wish to bound the probability that given all of the above choices, Γ_-^{o} does not contain a booster with respect to H. For that, recall the definition of $t_-=t_-^{(1)}$ from (4), let T be the set of times in which edges from H were traversed, and as in the proof of Lemma 5.11, define

$$I_T=\{i\in[t_-]\mid i\text{ is odd},\ \nexists s\in T:\ |i-s|\leq 1\}.$$

Note that in view of Lemma 5.2, **whp**

$$|I_T| \geq \frac{t_-}{2} - 4|E(H)| \geq \frac{t_-}{2} - 5d_0 n \geq \frac{t_-}{3},$$

since $\delta_0 < 1/1000$, and observe that every edge traversed in Γ_-^{o} at one of the times in I_T is chosen uniformly at random, and independently of all previous choices, from all n^2 possible directed edges (including loops), so the probability of hitting a given (undirected) edge is $2n^{-2}$. Since H is a $\left(\frac{n}{4}, 2\right)$-expander, it is connected, hence by Lemma 5.19, $|\mathcal{B}(H)| \geq n^2/32$, and it follows that for $t \in I_T$,

$$\mathbb{P}(e_t \in \mathcal{B}(H)) \geq \frac{n^2}{32} \cdot 2n^{-2} = \frac{1}{16},$$

hence for every H

$$\prod_{t \in I_T} \mathbb{P}(e_t \notin \mathcal{B}(H)) \leq \left(\frac{15}{16}\right)^{t_-/3}.$$

To summarize,

$$\begin{aligned}
&\mathbb{P}\big(\exists H :\ \mathcal{B}(H) \cap E\big(\Gamma_-^{\text{o}}\big) = \varnothing\big)\\
&\leq \sum_{i \leq d_0 n + n} \binom{\binom{n}{2}}{i} \sum_{j \leq n^{0.4}} \binom{i}{j} \left\lceil \frac{t_-}{2} \right\rceil^{i-j} \left(\frac{2}{n^2}\right)^{i-j} \left(\frac{15}{16}\right)^{t_-/3}\\
&\leq \left(\frac{15}{16}\right)^{t_-/3} \sum_{i \leq 2d_0 n} \left(\frac{en^2}{2i}\right)^i \left[n^{0.4} (2d_0 n)^{n^{0.4}} \left(\frac{n^2}{t_-}\right)^{n^{0.4}}\right] \left(\frac{2t_-}{n^2}\right)^i\\
&\leq \left(\frac{15}{16}\right)^{t_-/3} n^{\sqrt{n}} \sum_{i \leq 2d_0 n} \left(\frac{et_-}{i}\right)^i\\
&\leq \left(\frac{15}{16}\right)^{t_-/4} \sum_{i \leq 2d_0 n} \left(\frac{et_-}{i}\right)^i.
\end{aligned}$$

Let $f(x) = (et_-/x)^x$. In the interval $(0, et_-)$, f attains its maximum at t_-, and is unimodal. Recalling that $d_0 = \lfloor \delta_0 \ln n \rfloor$ and that we chose $\delta_0 < 1/1000$, f is strictly increasing in the interval $(0, 2d_0 n)$. Thus

$$\begin{aligned}
\mathbb{P}\big(\exists H :\ \mathcal{B}(H) \cap E\big(\Gamma_-^{\text{o}}\big) = \varnothing\big) &\leq \left(\frac{15}{16}\right)^{t_-/4} 2d_0 n \left(\frac{et_-}{2d_0 n}\right)^{2d_0 n}\\
&\leq \exp\left(\frac{t_-}{4} \ln\left(\frac{15}{16}\right) + 2d_0 n \ln\left(\frac{e}{\delta_0}\right)\right)\\
&\leq \exp\left(n \ln n \left(\frac{\ln(15/16)}{4} + 2\delta_0 \ln\left(\frac{e}{\delta_0}\right)\right)\right),\\
&\leq \exp\left(n \ln n \left(-\frac{1}{64} + 42e^{-20}\right)\right) = o(1). \qquad \square
\end{aligned}$$

Now all ingredients are in place for our final argument. We first state that **whp** the graph Γ_* contains a sparse $\left(\frac{n}{4}, 2\right)$-expander Γ_0, as delivered by Lemma 5.17. We set $H_0 = \Gamma_0$, and as long as H_i is not Hamiltonian, we seek a booster from Γ_-^{o} relative to it; once such a booster b is found, we add it to the graph and set $H_{i+1} = H_i + b$. This iteration is repeated less than n times. It cannot get stuck as otherwise we would get graph H_i for which the following hold:

- H_i is a non-Hamiltonian $\left(\frac{n}{4}, 2\right)$-expander (as $H_0 \subseteq H_i$)
- $|E(H_i)| \le d_0 n + n$ (as $|E(\Gamma_0)| \le d_0 n$)
- $\left|E(H_i) \smallsetminus E\left(\Gamma^{\mathrm{o}}_{-}\right)\right| \le n^{0.4}$ (follows from Lemma 5.14)
- Γ^{o}_{-} does not contain a booster with respect to H_i

and by Lemma 5.20, with high probability, such H_i does not exist.

This shows that Γ_{τ_C+1} is **whp** Hamiltonian; since $\delta(\Gamma_{\tau_C}) = 1$, $\tau_{\mathcal{H}} = \tau_C + 1$, and the proof of Theorem 2 is complete.

5.3 Perfect Matching

Assume n is even. Since $\delta(\Gamma_{\tau_C - 1}) = 0$, in order to prove Corollary 3 it suffices to show that $\tau_{\mathcal{PM}} \le \tau_C$. Indeed, our proof above shows that **whp** Γ_{τ_C} contains a Hamilton path. Taking every second edge of that path, including the last edge, yields a matching of size $n/2$, thus **whp** Γ_{τ_C} contains a matching of that size, and Corollary 3 follows.

6 Concluding remarks

We have investigated several interesting graph properties (minimum degree, vertex-connectivity, Hamiltonicity) of the trace of a long-enough random walk on a dense-enough random graph, showing that in the relevant regimes, the trace behaves much like a random graph with a similar density. In the special case of a complete graph, we have shown a hitting time result, which is similar to standard results about random graph processes.

However, the two models are, in some aspects, very different. For example, an elementary result from random graphs states that the threshold for the appearance of a vertex of degree 2 is $n^{-3/2}$, whereas the expected density of the trace of the walk on K_n, at the moment the maximum degree reaches 2, is of order n^{-2} (as it typically happens after two steps). It is therefore natural to ask for which graph properties and in which regimes the two models are alike.

Further natural questions inspired by our results include asking for the properties of the trace of the walk in different random environments, such as random regular graphs, or in deterministic environments, such as (n, d, λ)-graphs and other pseudo-random graphs (see [21] for a survey). We have decided to leave these questions for a future research.

A different direction would be to study the directed trace. Consider the set of *directed* edges traversed by the random walk. This induces a random directed (multi)graph, and we may ask, for example: is it true that when walking on the complete graph, typically one step after covering the graph we achieve a *directed* Hamilton cycle?

Acknowledgement. The authors wish to thank Wojciech Samotij and Yinon Spinka for useful discussions, and the anonymous referee for careful reading and valuable comments and suggestions.

References

[1] Miklós Ajtai, János Komlós and Endre Szemerédi, *First occurrence of Hamilton cycles in random graphs*, Cycles in graphs (Burnaby, B.C., 1982), 1985, pp. 173–178. MR821516 ↑2

[2] Ben Barber and Eoin Long, *Random walks on quasirandom graphs*, Electronic Journal of Combinatorics **20** (2013), no. 4, Paper 25. MR3158264 ↑2

[3] Sonny Ben-Shimon, Asaf Ferber, Dan Hefetz and Michael Krivelevich, *Hitting time results for Maker-Breaker games*, Random Structures & Algorithms **41** (2012), no. 1, 23–46. MR2943425 ↑6

[4] Itai Benjamini, Ori Gurel-Gurevich and Russell Lyons, *Recurrence of random walk traces*, The Annals of Probability **35** (2007), no. 2, 732–738. MR2308594 ↑2

[5] Béla Bollobás, *The evolution of sparse graphs*, Graph theory and combinatorics, 1984, pp. 35–57. MR777163 ↑2

[6] Béla Bollobás, **Random graphs**, 2nd ed., Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001. MR1864966 ↑1, 29

[7] Béla Bollobás and Andrew Thomason, *Random graphs of small order*, Random graphs '83 (Poznań, 1983), 1985, pp. 47–97. MR860586 ↑2

[8] Jiří Černý, Augusto Teixeira and David Windisch, *Giant vacant component left by a random walk in a random d-regular graph*, Annales de l'Institut Henri Poincaré Probabilités et Statistiques **47** (2011), no. 4, 929–968. MR2884219 ↑2

[9] Herman Chernoff, *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations*, Annals of Mathematical Statistics **23** (1952), 493–507. MR0057518 ↑5

[10] Colin Cooper and Alan Frieze, *The cover time of sparse random graphs*, Random Structures & Algorithms **30** (2007), no. 1-2, 1–16. MR2283218 ↑2

[11] Colin Cooper and Alan Frieze, *Component structure of the vacant set induced by a random walk on a random graph*, Random Structures & Algorithms **42** (2013), no. 2, 135–158. MR3019395 ↑2

[12] Paul Erdős and Alfréd Rényi, *On a classical problem of probability theory*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **6** (1961), 215–220. MR0150807 ↑21

[13] Paul Erdős and Alfréd Rényi, *On random graphs. I*, Publicationes Mathematicae Debrecen **6** (1959), 290–297. MR0120167 ↑2

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

[15] Uriel Feige, *A tight lower bound on the cover time for random walks on graphs*, Random Structures & Algorithms **6** (1995), no. 4, 433–438. MR1368844 ↑2

[16] Alan Frieze and Michał Karoński, **Introduction to random graphs**, Cambridge: Cambridge University Press, 2016 (English). ↑1, 7

[17] Dan Hefetz, Michael Krivelevich and Tibor Szabó, *Hamilton cycles in highly connected and expanding graphs*, Combinatorica **29** (2009), no. 5, 547–568. MR2604322 ↑3, 20

[18] Wassily Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association **58** (1963), 13–30. MR0144363 ↑5

[19] Svante Janson, Tomasz Łuczak and Andrzej Ruciński, **Random graphs**, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000. MR1782847 ↑1, 5

[20] Mark Jerrum and Alistair Sinclair, *Approximating the permanent*, SIAM Journal on Computing **18** (1989), no. 6, 1149–1178. MR1025467 ↑10, 11

[21] Michael Krivelevich and Benny Sudakov, *Pseudo-random graphs*, More sets, graphs and numbers, 2006, pp. 199–262. MR2223394 ↑31

[22] László Lovász, *Random walks on graphs: a survey*, Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), 1996, pp. 353–397. MR1395866 ↑2

[23] Lajos Pósa, *Hamiltonian circuits in random graphs*, Discrete Mathematics **14** (1976), no. 4, 359–364. MR0389666 ↑29