Percolation of averages in the stochastic mean field model: the near-supercritical regime

Jian Ding∗ Subhajit Goswami∗
University of Chicago University of Chicago
January 20, 2015

Abstract

For a complete graph of size $n$, assign each edge an i.i.d. exponential variable with mean $n$. For $\lambda > 0$, consider the length of the longest path whose average weight is at most $\lambda$. It was shown by Aldous (1998) that the length is of order $\log n$ for $\lambda < 1/e$ and of order $n$ for $\lambda > 1/e$. In this paper, we study the near-supercritical regime where $\lambda = e^{-1} + \eta$ with $\eta > 0$ a small fixed number. We show that there exist two absolute constants $c^*, C^* > 0$ such that with high probability the length is in between $ne^{-C^*/\sqrt{\eta}}$ and $ne^{-c^*/\sqrt{\eta}}$. Our result corrects a non-rigorous prediction of Aldous (2005).

1 Introduction

Let $W_n$ be a complete undirected graph of $n$ vertices where each edge is assigned an independent exponential weight with mean $n$; this is referred to as the stochastic mean-field model. For a (self-avoiding) path $\pi = (v_0, v_1, \ldots, v_m)$, define its length $\text{len}(\pi)$ and average weight $A(\pi)$ by

$$\text{len}(\pi) = m, \text{ and } A(\pi) = \frac{1}{m} \sum_{i=1}^{m} W_{(v_{i-1}, v_i)},$$

where $W_{(u,v)}$ is the weight of the edge $(u, v)$. For $\lambda > 0$, let $L(n, \lambda)$ be the length of the longest path with average weight below $\lambda$, i.e.,

$$L(n, \lambda) = \max \{ \text{len}(\pi) : A(\pi) \leq \lambda, \pi \text{ is a path in SMF}_n \text{ model} \}.$$ 

In a non-rigorous paper of Aldous [2], it was predicted that $L(n, \lambda) \asymp n(\lambda - e^{-1})^\beta$ with $\beta = 3 \lambda \downarrow e^{-1}$. Our main result is the following theorem, which corrects Aldous’ prediction.

**Theorem 1.1.** Let $\lambda = 1/e + \eta$ where $\eta > 0$. Then there exist absolute constants $c^*, C^*, \eta^* > 0$ such that for all $\eta \leq \eta^*$,

$$\lim_{n \to \infty} P(ne^{-c^*/\sqrt{\eta}} \leq L(n, \lambda) \leq ne^{-C^*/\sqrt{\eta}}) = 1. \quad (1.1)$$

The study of the object $L(n, \lambda)$ was initiated by Aldous [1] where a phase transition was discovered at the threshold $e^{-1}$. It was shown that with high probability $L(n, \lambda)$ is of order $\log n$ for $\lambda < e^{-1}$ and $L(n, \lambda)$ is of order $n$ when $\lambda > e^{-1}$. The critical behavior was established in [4], where it was proved that with high probability $L(n, \lambda)$ is of order $(\log n)^3$ when $\lambda$ is around $e^{-1}$ within

*Partially supported by NSF grant DMS-1313596.
a window of order $(\log n)^{-2}$. Our Theorem 1.1 describes the behavior in the near-supercritical regime, and in particular states that $L(n, \lambda)/n$ is a stretched exponential in $\eta$ with $\eta = \lambda - e^{-1} \downarrow 0$.

A highly related question is the length for the cycle of minimal mean weight, which was studied by Mathieu and Wilson [9]. An interesting phase transition was found in [9] with critical threshold $e^{-1}$ on the mean weight. Further results on this problem will be proved in a forthcoming paper [5].

Another related question is the classical traveling salesman problem (TSP), where one minimizes the weight of the path subject to passing through every single vertex in the graph. For the TSP in the mean-field set up, Wästlund [12] established the sharp asymptotics for more general distributions on the edge weight, confirming the Krauth-Mézard-Parisi conjecture [10, 11, 8]. Indeed, it is an interesting challenge to give a sharp estimate on $L(n, \lambda)$ for $e^{-1} < \lambda < \lambda^*$ (here $\lambda^*$ is the asymptotic value for TSP), interpolating the critical behavior and the extremal case of TSP.

Another question of similar flavor is on the tree with average weight below a certain threshold, where a phase transition is proved in [1]. The extremal case of the question on the tree with minimal average weight is the well-known minimal spanning tree problem, where a $\zeta(3)$ limit is established by Frieze [7].

**Main ideas of our proofs.** A straightforward first moment computation as done in [1] implies that the critical threshold for the average weight is $1/e$, and that $\lim_{n \to \infty} \mathbb{P}(L(n, \lambda) = O(\log n)) = 1$ when $\lambda < 1/e$. For $\lambda > 1/e$, a sprinkling method was employed in [1] to show that with high probability $L(n, \lambda) = O(n)$, where the author first proved with high probability there exist a large number of paths with average weight slightly above $1/e$ and then used a certain greedy algorithm to connect these paths into a single long path with average weight slightly above $1/e$. However, the method in [1] was not able to describe the behavior at criticality. In [4] (see also [9] for the cycle with minimal average weight), a second moment computation was carried out restricted to paths of average weight below $1/e$ and with the maximal deviation (see [3.2] below for definition of the maximal deviation) at most $O(\log n)$, thereby yielding that with high probability $L(n, 1/e) = \Theta((\log n)^3)$. A crucial fact responsible for the success of the second moment computation is that the length of the target path is $\Theta((\log n)^3) \ll \sqrt{n}$. As such, a straightforward adaption of this method would not be able to succeed in the regime considered by this paper. For TSP, where one studies paths (cycles) that visit every single vertex, is in a sense analogous to the question of finding the minimal value $\lambda$ for which $L(n, \lambda) = n$ with high probability. Wästlund [12] showed that the minimum average cost of TSP converges in probability to a positive constant by relaxing it to a certain linear optimization problem. But it seems difficult to extend his method to incomplete TSP i.e. when the target object is the minimum cost cycle having at least $pn$ many edges for some $p \in (0, 1)$. Since our problem is in a sense dual to incomplete TSP in the regime we are interested in, the method of [12] does not seem to be suitable for our purpose either. In the current work, our method is inspired by the (first and) second moment method from [4] as well as the sprinkling method employed in [1].

In order to prove the upper bound, our main intuition is that we would have a larger number of short and light paths (a light path refers to a path with small average weight — slightly above $1/e$) than we would typically expect if $L(\lambda, n)$ were greater than $e^{-C}/\sqrt{n}$. Formally, let $\ell = \frac{c_1}{\eta}$ where $c_1$ is a small positive constant, and consider the number of paths (denoted by $N_\ell$) with length $\ell$ and total weight no more than $\lambda \ell - c_2\sqrt{\ell}$ for a positive constant $c_2$. We call such a path a downcrossing. A straightforward computation should give $\mathbb{E}N_\ell = O(1)nle^{-c_3/\sqrt{n}}$ for a positive constant $c_3$ depending on $c_1$ and $c_2$. Now we consider the number of paths (denoted by $N_\delta$) of length $\delta(\lambda)n$ and average weight at most $\lambda$. Such paths have two possibilities: (1) The path contains substantially more than $\mathbb{E}N_\ell$ many downcrossings, which is unlikely by Markov’s inequality. (2) The path does not have substantially more than $\mathbb{E}N_\ell$ many downcrossings. This is also unlikely for the following reasons: (a) A straightforward first moment computation gives that $\mathbb{E}N_\delta = O(n)e^{-c_4n\eta}$ for
a constant $c_4 > 0$; (b) The number of down crossings in a typical path of this kind should dominate a Binomial random variable $\text{Bin}(\delta n/\ell, c_5)$ where $c_5 > 0$ is an absolute constant (since in the random walk bridge, every sub-path of size $\ell$ has a positive chance to have such a downcrossing). If we choose $\delta$ suitably large as in Theorem [1], we are suffering a probability cost for the constraint on the number of downcrossings (probability for a binomial much smaller than its mean) and this probability cost is of magnitude $e^{-c_6 \delta n / \ell}$ for a constant $c_6 > 0$ depending in $c_1$. If we choose $c_1$ small enough this probability cost kills the growth of $e^{c_4 \delta n}$ in $\mathbb{E}N_\delta$. Therefore, paths of this kind do not exist either. The detailed are carried out in Section 2.

For the lower bound, our proof consists of two steps. In light of the preceding discussion, we cannot hope to directly apply a second moment method from [4, 9] to show the existence of a light path that is of length linear in $n$. As such, in the first step of our proof we prove that with high probability there exists a linear (in $n$) number of disjoint paths, each of which has weight slightly below $\lambda$ and is of length $e^{c_7/\sqrt{n}}$ for an absolute constant $c_7 > 0$. This is achieved by two second moment computations, which are expected to succeed as the length of the path under consideration is $\ll \sqrt{n}$ (indeed remains bounded as $n \to \infty$). In the second step, we propose an algorithm which, with probability going to 1, strings together a suitable collection of these short light paths to form a light path of length $e^{-c_8/\sqrt{n}n}$ for an absolute constant $c_8 > 0$. Our algorithm is similar to the greedy algorithm (or in a different name exploration process) employed in [1]. But in order to ensure that the additional weight introduced by these connecting bridges only increases the average weight of the final path by an amount of $O(\eta)$, we have to use a more delicate algorithm. The details are carried out in Section 3.

**Notation convention.** For a graph $G$, denote by $V(G)$ and $E(G)$ the set of vertices and edges of $G$, respectively. For a path $\pi$, we use the notation $\pi$ to denote the sequence of vertices in the path as well as the underlying graph (which is a path). This should be clear from the context. The weight of an edge $e$ in $W_n$ is $W_e$ and we denote the total weight of a path $\pi$ by $W(\pi)$. $\Pi_\ell$ is the collection of all paths $\pi$ of length $\ell \in [n]$. We let $\lambda = 1/e + \eta$ where $\eta$ is a positive number. A path is called $\lambda$-light if its average weight is at most $\lambda$, and a path is called $(\lambda, C)$-light if its total weight is at most $\lambda \ell - C\sqrt{\ell}$ where $\ell$ is length of the path. For any non-negative integer or real valued variable $v$, the phrases “for large $v$” or “when $v$ is large” would mean that there exists an absolute positive constant $v_0$ such that the corresponding statement holds whenever $v \geq v_0$. In cases where some other variable $v'$ (or a list of several variables) is involved and the statement holds for any fixed value of $v'$ and $v \geq v_0$ where $v_0$ is a function of $v$, we will add “(given $v'$)” after the above phrases. In a similar vein we use “for small $v$” or “when $v$ is small” if the corresponding statement holds for $0 < v \leq v_0$. Throughout this paper the order notations $O(\cdot), \Theta(\cdot)$ and $o(\cdot)$ are assumed to be with respect to $n \to \infty$ while keeping all the other involved parameters (such as $\ell$, $\eta$ etc.) fixed. We will use $C_1, C_2, \ldots$ to denote for constants, and each $C_i$ will denote the same number throughout of the rest of the paper.

**Acknowledgements.** We thank David Aldous for very useful discussions.

### 2 Proof of the upper bound

Let $\eta'$ be a constant multiple of $\eta$ whose precise value is to be selected, and let $C$ be a constant to be selected later. Set $\ell = [1/\eta']$ and let $N_{\eta', C}$ be the number of “$(\lambda, C)$-light” paths of length $\ell$. As outlined in the introduction, we shall first control $N_{\eta', C}$.

It is clear that the distribution of the total weight of a path of length $k$ follows a Gamma distribution $\Gamma(k, 1/n)$, where the density $f_{\theta, k}(z)$ of Gamma$(k, \theta)$ is given by

$$f_{\theta, k}(z) = \theta^k z^{k-1} e^{-\theta z} / (k - 1)!$$

for all $z \geq 0, \theta > 0$ and $k \in \mathbb{N}$. (2.1)
By (2.1) and the Stirling’s formula, we carry out a straightforward computation and get that
\[
\mathbb{E}(N_{\eta',C}) = (1 + o(1)) \times n^{\ell+1} \times P\left(\text{Gamma}(\ell, 1/n) \leq \lambda \ell - C\sqrt{\ell}\right)
\]
\[
= (1 + o(1)) \times n^{\ell+1} \times \frac{e^{-(\lambda \ell - C\sqrt{\ell})/n}(\lambda \ell - C\sqrt{\ell})^\ell}{\ell!n^\ell}
\]
\[
= (1 + o(1))C_0(\eta)C_0(\eta')n^{\ell}e^{\eta/\eta'}n^\ell e^{-Ce/\sqrt{\eta}} ,
\]
(2.2)

where \(C_0(\eta) \to 1\) as \(\eta \to 0\), and \(C_0(C)\) is a positive constant depending on \(C\). Furthermore both of \(1 + o(1), C_0(\eta)\) are less than or equal to 1.

We also need a bound on the second moment of \(N_{\eta',C}\) to control its concentration around \(\mathbb{E}(N_{\eta',C})\).

For \(\gamma \in \Pi_\ell\), define \(F_\gamma\) to be the event that \(\gamma\) is \((\lambda, C)\)-light. Then clearly we have \(N_{\eta',C} = \sum_{\gamma \in \Pi_\ell} 1_{F_\gamma}\). In order to compute \(\mathbb{E}N_{\eta',C}^2\), we need to estimate \(P(F_\gamma \cap F_{\gamma'})\) for \(\gamma, \gamma' \in \Pi_\ell\). In the case \(E(\gamma) \cap E(\gamma') = \emptyset\), we have \(F_\gamma\) and \(F_{\gamma'}\) independent of each other and thus \(P(F_{\gamma'}|F_\gamma) = P(F_{\gamma'})\). In the case \(|E(\gamma) \cap E(\gamma')| = j > 0\), we have
\[
P(F_{\gamma'}|F_\gamma) \leq P\left(\text{Gamma}(\ell - j, 1/n) \leq \lambda \ell \right) \leq \frac{1}{(\ell - j)!} \frac{\lambda^{\ell - j}}{n^{\ell - j}}.
\]
(2.3)

Further notice that if \(|E(\gamma) \cap E(\gamma')| = j\), then \(|V(\gamma) \cap V(\gamma')|\) is at least \(j + 1\) as \(\gamma \cap \gamma'\) is acyclic. So given any \(\gamma \in \Pi_\ell\), the number of paths \(\gamma'\) such that \(|E(\gamma) \cap E(\gamma')| = j\) is at most \(O(n^{\ell - j})\).

Altogether, we obtain that
\[
\mathbb{E}(N_{\eta',C}^2) = \sum_{\gamma, \gamma' \in \Pi_\ell} P(F_\gamma \cap F_{\gamma'}) = \sum_{\gamma \in \Pi_\ell} P(F_\gamma) \sum_{\gamma' \in \Pi_\ell} P(F_{\gamma'}) = \mathbb{E}(N_{\eta',C}) \sum_{\gamma' \in \Pi_\ell} \mathbb{E}(F_{\gamma'}) = \mathbb{E}(N_{\eta',C}) \sum_{\gamma' \in \Pi_\ell} \left(P(F_{\gamma'}) + \sum_{1 \leq j \leq \ell, \gamma : |E(\gamma \cap \gamma')| = j} \frac{1}{(\ell - j)!} \frac{\lambda^{\ell - j}}{n^{\ell - j}} \right)
\]
\[
\leq \mathbb{E}(N_{\eta',C}) \left(\mathbb{E}(N_{\eta',C}) + O(1)\right).
\]
(2.4)

Combined with (2.2), it yields that
\[
\mathbb{E}(N_{\eta',C}^2) = \mathbb{E}(N_{\eta',C})^2 (1 + o(1)).
\]
(2.5)

As a consequence of Markov’s inequality (applied to \(|N_{\eta',C} - \mathbb{E}N_{\eta',C}|^2\)), we get that
\[
P\left(N_{\eta',C} \geq 2\mathbb{E}(N_{\eta',C})\right) = o(1).
\]
(2.6)

Next, we show that any long \(\lambda\)-light path should have a large number of sub-paths which are \((\lambda, C)\)-light. Let \(\pi\) be a path of length \(\delta n\) for some \(\delta > 0\). Denote its successive edge weights by \(X_1, X_2, \ldots, X_{\delta n}\) and let \(S_k = \sum_{i=1}^{k} X_i\) for \(1 \leq k \leq \delta n\). Throughout this section probabilities of events involving edge weights of \(\pi\) will be assumed to be conditioned on \(\{A(\pi) \leq \lambda\}\) although we will omit that phrase except in formal expressions and lemmas. Now divide \(\pi\) into edge disjoint sub-paths of length \(\ell\) (with the last subpath of length possibility less than \(\ell\) in the case \(\ell\) does not divide \(\delta n\)) and call each such sub-path a \(C\)-downcrossing if it is \((\lambda, C)\)-light. The following well-known result about exponential random variables (see, e.g., [3, Theorem 6.6]) will be very useful.

**Lemma 2.1.** Let \(W_1, W_2, \ldots, W_N\) be i.i.d. exponential random variables with mean \(1/\theta\), and let \(S_k = \sum_{i=1}^{k} W_i\) for \(1 \leq k \leq N\). Then the random vector \((\frac{W_1}{S_N}, \ldots, \frac{W_{N-1}}{S_{N-1}})\) follows Dirichlet(1\(_N\)) distribution, \(S_N\) follows Gamma\((N; \theta)\) distribution, and they are independent of each other. Here \(1_N\) is the \(N\)-dimensional vector whose all entries are 1.
Lemma 2.2. Let $Z_1, Z_2, \ldots, Z_N$ be i.i.d. exponential random variables with mean 1 and let $S_N = \sum_{i=1}^{N} S_i$. Then
\[
\begin{align*}
\mathbb{P}(S_N \geq N + \alpha) &\leq e^{-\alpha^2/4N} \text{ for all } 0 < \alpha \leq (2 - \sqrt{2})N, \quad (2.7) \\
\mathbb{P}(S_N \leq N - \alpha) &\leq e^{-\alpha^2/2N}, \text{ for all } \alpha > 0. \quad (2.8)
\end{align*}
\]

Proof. By Markov’s inequality, we get that for any $\alpha > 0$ and $0 < \theta < 1$,
\[
\mathbb{P}(S_N \geq N + \alpha) = \mathbb{P}(e^{\theta S_N} \geq e^{\theta(N+\alpha)}) \leq e^{\theta N - \alpha \theta / (1-\theta)^2}.
\]

When $\theta \leq 1 - 1/\sqrt{2}$, the right hand side is bounded above by $e^{\theta^2 - \alpha \theta}$. So setting $\theta = \alpha / 2N$ yields (2.7) as long as $0 < \alpha / 2N \leq 1 - 1/\sqrt{2}$. We can prove (2.8) in the same manner. 

Now consider the $k$-th sub-path $b_k^\pi$ along the path $\pi$ for $k \leq \delta N / 2\ell$. Denote by $\Lambda_k$ the average weight of the portion of the path $\pi$ starting from $b_k^\pi$, i.e., $\Lambda_k = (\delta n_k - S_{\ell(n_k - (k-1)\ell)})/(\delta n_k - (k-1)\ell)$. By Lemma 2.1, we know that as $n \to \infty$, conditional mean and variance of $S_k$ given $S_{\delta n_k} = \mu \delta n_k$ are $\mu \ell$ and $\mu^2 \ell (1 + o(1))$ respectively for all $\mu > 0$. So it is plausible to expect that the probability of the event $\{W(b_k^\pi) \leq \Lambda_k(\ell - 6\sqrt{\ell})\}$ to be bounded away from 0 uniformly for small $\eta$ and large $n$ (given $\eta$) regardless of the weights of first $(k-1)\ell$ edges of $\pi$. The formal statement is in the next lemma.

Lemma 2.3. Let $N_{\eta', \delta, n}^\pi$ be the number of sub-paths $b_k^\pi$ satisfying $W(b_k^\pi) \leq \Lambda_k(\ell - 6\sqrt{\ell})$ where $1 \leq k \leq \delta n / 2\ell$ and $\ell = [1/\eta']$. Then for any $0 < \eta' < \eta_0$ where $\eta_0$ is a positive, absolute constant and any $0 < \delta_0 < 1$ there exists a positive integer $n_0 = n_0(\delta_0, \eta')$ and an absolute constant $c > 0$ such that for all $\delta \geq \delta_0$ and $n \geq n_0$ the conditional distribution of $N_{\eta', \delta, n}^\pi$ given $\{A(\pi) \leq \lambda\}$ stochastically dominates the binomial distribution $\text{Bin}(\delta n / 2\ell, c)$.

Proof. Notice that it suffices to prove that there exists an absolute constant $c > 0$ such that uniformly for $\mu > 0$ and large $L$ (given $\ell$)
\[
\mathbb{P}(S_L \leq \frac{\delta n}{L} (\ell - 6\sqrt{\ell}) | S_L = \mu L) \geq c.
\]

To this end, we see that for $L > \ell$
\[
\mathbb{P}(S_L \leq \frac{\delta n}{L} (\ell - 6\sqrt{\ell}) | S_L = \mu L) = \mathbb{P}(\frac{S_L}{\ell} \leq \frac{\delta n}{L} (\ell - 6\sqrt{\ell}) / \ell | S_L = \mu L) \mathbb{P}(\frac{S_L}{\ell} \leq (\ell - 6\sqrt{\ell}) / L), \quad (2.9)
\]

where the last equality follows from Lemma 2.1. Since distribution of $\frac{S_L}{\ell} / L$ does not depend on the mean of the underlying $X_j$’s, we can in fact assume that $X_j$’s are i.i.d. exponential variables with mean 1 for purpose of computing (2.9). By (2.8), we have
\[
\mathbb{P}(S_L / L \leq 1 - 1/2\sqrt{\ell}) \leq e^{-L / 8\ell}.
\]

So for $\ell - 6\sqrt{\ell} > 0$, we get
\[
\mathbb{P}(S_L \leq \frac{\delta n}{L} (\ell - 6\sqrt{\ell})) \geq \mathbb{P}(S_L \leq \ell - 6\sqrt{\ell}) - e^{-L / 8\ell}. \quad (2.10)
\]

By central limit theorem there exist absolute numbers $\ell_0, c > 0$ such that $\mathbb{P}(S_L \leq \ell - 6\sqrt{\ell}) \geq c$ for $\ell \geq \ell_0$. Hence from (2.10) it follows that for any $\ell \geq \ell_0$ there exists $L_0 = L_0(\ell)$ such that the right hand side of (2.9) is at least $c$ for $L \geq L_0$. 


Lemma 2.3 motivates us to use a binomial approximation for the number of downcrossings that occur, say, up to first half of the path. One scenario that might undermine our estimate is when some $\Lambda_k$ is significantly above $\lambda$. But that would imply a substantial drop in $S_k$ for some $1 \leq k \leq \delta n/2$, which occurs with small probability.

**Lemma 2.4.** Denote by $E_{n, \eta'}$ the event that $\Lambda_k$ is more than $\lambda + \sqrt{\eta'}$ for some $1 \leq k \leq \frac{\delta n}{2\eta}$. Then for any $0 < \eta' < 1/4$ and $0 < \delta_0 < 1$ there exists a positive integer $n_0 = n_0(\delta_0, \eta')$ such that,

$$
\mathbb{P}(E_{n, \eta'}|A(\pi) \leq \lambda) \leq 2e^{-\delta n \eta'/16} \text{ for all } \delta \geq \delta_0 \text{ and } n \geq n_0.
$$

**Proof.** For $1 \leq k \leq \delta n/2\ell$, let $\ell_k = (k-1)\ell$ and let $E_{n,k, \eta'} = \{\Lambda_k \geq \lambda + \sqrt{\eta'}\}$. On $E_{n,k, \eta'}$, we have

$$
\frac{S_{\ell_k}}{S_{\delta n}} \leq \frac{\ell_k S_{\delta n}/\delta n - \sqrt{\eta'}(\delta n - \ell_k)}{S_{\delta n}} \leq \frac{\ell_k}{\delta n} - \sqrt{\eta'} (\delta n - \ell_k),
$$

where the last inequality holds since we are conditioning on $S_{\delta n} \leq \lambda \delta n$ and $\lambda \leq 1$ when $\eta' < 1/4$. Therefore, we get

$$
\mathbb{P}(E_{n,k, \eta'}|A(\pi) \leq \lambda) \leq \mathbb{P}(S_{\ell_k} \leq \frac{S_{\delta n}}{\delta n} (\ell_k - \sqrt{\eta'}(\delta n - \ell_k)))
$$

(2.12)

Now we evaluate the right hand side of (2.12). Analogous to (2.9) in the proof of Lemma 2.3, we can assume without loss of generality that $X_1, X_2, \ldots, X_L$ are i.i.d. exponential variables with mean 1. It is routine to check that

$$
(1 + \sqrt{\eta'}/2) \times (\ell_k - \sqrt{\eta'}(\delta n - \ell_k)) \leq \ell_k - \sqrt{\eta'} \delta n/4, \text{ for all } 1 \leq k \leq \delta n/2\ell.
$$

Thus, for all $1 \leq k \leq \delta n/2\ell$ we get

$$
\mathbb{P}
\left(S_{\ell_k} \leq \frac{S_{\delta n}}{\delta n} (\ell_k - \sqrt{\eta'}(\delta n - \ell_k))\right)
\leq \mathbb{P}
\left(S_{\ell_k} \leq \ell_k - \sqrt{\eta'} \delta n/4\right) + \mathbb{P}
\left(S_{\delta n}/\delta n \geq 1 + \sqrt{\eta'}/2\right)
$$

\[ \leq e^{-\delta n \eta'/16} + e^{-\delta n \eta'/16}, \]

where the second inequality follows from (2.8) and (2.7) respectively. Combined with (2.12), it gives that

$$
\mathbb{P}(E_{n,k, \eta'}|A(\pi) \leq \lambda) \leq 2e^{-\delta n \eta'/16}, \text{ for all } 1 \leq k \leq \delta n/2\ell.
$$

An application of a union bound over $k$ completes the proof of the lemma.

**Proof of Theorem 1.1** upper bound. Let $N_{\eta', \delta, n}$ denote the number of 1-downcrossings in the path $\pi$. Let $\eta_0$ be given in the statement of Lemma 2.3 and let $n_0$ be the maximum value of the $n_0$'s as stated in Lemmas 2.3 and 2.4 for some fixed $\delta_0$ in $(0, 1)$. Assume that $\eta' < 1/4 \wedge \eta_0$. Roughly, the argument goes as follows: By Lemma 2.4, we see that with large probability $\Lambda_k \leq \lambda + \sqrt{\eta'}$ for all $k$ between 1 and $\delta n/2\ell$. On $E_{n, \eta'}$, it holds that the event $W(b_{\delta n}^c) \leq \Lambda_k (\ell - \sqrt{\ell})$ implies that $W(b_{\delta n}^c) \leq \ell - \sqrt{\ell}$ for small $\eta'$. Thus, we can use Lemma 2.3 to obtain a binomial approximation for $N_{\eta', \delta, n}$ on $E_{n, \eta'}$. Formally,

$$
\mathbb{P}(\tilde{N}_{\eta', \delta, n} \leq 2\mathbb{E}(N_{\eta', 1})|A(\pi) \leq \lambda) \leq \mathbb{P}
\left(N_{\eta', \delta, n} \leq 2\alpha_1 e^{\eta' n/\sqrt{\eta} e^{-e/\sqrt{\eta}}} |A(\pi) \leq \lambda\right) + 2ne^{-\delta n \eta'/16},
$$

where $\alpha_1$ is an absolute constant (see (2.2)) and the inequality follows from Lemma 2.4. Therefore, by Lemma 2.3 we get that

$$
\mathbb{P}(\tilde{N}_{\eta', \delta, n} \leq 2\mathbb{E}(N_{\eta', 1})|A(\pi) \leq \lambda) \leq \mathbb{P}\left(\text{Bin}(\delta n/2\ell, c) \leq 2\alpha_1 e^{\eta' n/\sqrt{\eta} e^{-e/\sqrt{\eta}}} + 2ne^{-\delta n \eta'/16},
\right)
$$

(2.13)
Choose \( \delta_0 = \delta_0(\eta') \) such that
\[
\delta_0 n\eta'c/4 = 2\alpha_1 e^{\eta'\eta}/n\sqrt{\eta' e^{-\eta'/\sqrt{\eta'}}}, \tag{2.14}
\]
we get from Binomial concentration that
\[
P\left( \text{Bin}(\delta n/2, \ell, c) \leq \alpha_1 e^{\eta'\eta}/n\sqrt{\eta' e^{-\eta'/\sqrt{\eta'}}} \right) \leq e^{-\delta_0 n c^2/16}.
\]
Next let us define a new event \( \Xi_{\eta,\delta_0,n} \) by
\[
\Xi_{\eta,\delta_0,n} = \bigcup_{k \geq \delta_0 n} \bigcup_{\pi \in \Pi_k} \left\{ N_{\pi,k/n,n} \geq 2\mathbb{E}(N_{\pi,k/n,n}), A(\pi) \leq \lambda \right\}.
\]
Then,
\[
P\left( L(\lambda, n) \geq \delta_0 n \right) = P\left( \{ L(\lambda, n) \geq \delta_0 n \} \cap \Xi_{\eta,\delta_0,n} \right) + P\left( \{ L(\lambda, n) \geq \delta_0 n \} \setminus \Xi_{\eta,\delta_0,n} \right)
\leq P\left( N_{\pi,k/n,n} \geq 2\mathbb{E}(N_{\pi,k/n,n}) \right) + \sum_{k \geq \delta_0 n} \left( 2ne^{-k\eta}/16 + e^{-k\eta}/2^{16} \right) \sum_{\pi \in \Pi_k} P\left( A(\pi) \leq \lambda \right).
\tag{2.15}
\]
By (2.1), we see that \( \sum_{\pi \in \Pi_k} P\left( A(\pi) \leq \lambda \right) \leq \frac{n}{\sqrt{2\pi k}} e^{k\eta} \). Set \( \eta' = 32\eta e^c/\sqrt{c} \). By (2.6) and (2.15),
\[
P\left( L(\lambda, n) \geq \delta_0 n \right) = o(1) \text{ as } n \to \infty.
\]
Combined with (2.14), this completes the proof of the upper bound.

\section{Proof of the lower bound}

\subsection{Existence of a large number of vertex-disjoint light paths}

For \( \eta, \zeta_1 \in (0, 1) \), let \( W_n^* \subseteq W_n \) be a complete graph of \( (1-\zeta_1\eta)n \) vertices. This subsection is devoted to prove that there exists a large number of vertex-disjoint \( \lambda \)-light paths with high probability. For \( \ell \in \mathbb{N} \), denote by \( \Pi_{\ell}^* \) the set of all paths of length \( \ell \) in \( W_n^* \). For \( \pi \in \Pi_{\ell}^* \) and some \( \zeta_2 > 0 \), define
\[
G_\pi = \left\{ \lambda\ell - 1 - W(\pi) \leq \lambda\ell, M(\pi) \leq (\zeta_2/\sqrt{\eta})(W(\pi)/\lambda\ell) \right\}, \tag{3.1}
\]
where \( M(\pi) \) is the maximum deviation of \( \pi \) away from the linear interpolation between the starting and ending points, formally given by
\[
M(\pi) = \sup_{1 \leq k \leq \ell} \left| \sum_{i=1}^k W_{e_i} - \frac{k}{\ell} W(\pi) \right| \tag{3.2}
\]
A similar class of events were considered in [4, 9] in order for second moment computation. As the authors mentioned in these papers, the factor \( W(\pi)/\lambda \ell \) provides some technical ease in view of the following property which is a consequence of Lemma 2.1
\[
P\left( M(\pi) \leq (\zeta_2/\sqrt{\eta})(W(\pi)/\lambda\ell) \mid W(\pi) = w \right) \equiv \text{constant for all } w > 0. \tag{3.3}
\]
We will call a path \( \pi \in \Pi_{\ell}^* \) \textit{good} if \( G_\pi \) occurs. Denote by \( N_{\ell}^* \) the total number of good paths, i.e., \( N_{\ell}^* = \sum_{\pi \in \Pi_{\ell}^*} 1_{G_\pi} \). In order to carry out second moment analysis we need to control the correlation between \( 1_{G_\pi} \) and \( 1_{G_{\pi'}} \), where \( \pi, \pi' \) are from \( \Pi_{\ell}^* \). It is plausible that such correlation depends on the number of common edges for \( \pi \) and \( \pi' \). But the number of common edges on its own is not...
Lemma 3.3. \([4, \text{Lemma } 3.2]\) Let \(1 \leq r \leq 1\). Consider \(c\) for any positive integers \(i \leq j\), define the set \(A_{i,j}\) as
\[
A_{i,j} \equiv A_{i,j}(\pi) = \{\pi' \in \Pi_k^\pi : \theta(\pi, \pi') = i, |E(\pi) \cap E(\pi')| = j\}. \tag{3.4}
\]
We need a number of lemmas from \([4]\).

Lemma 3.1. \([4, \text{Lemma } 2.10]\) For any \(1 \leq \ell \leq N \equiv (1 - \zeta \eta)n\) and any \(\pi \in \Pi_k^\pi\), we have that for any positive integers \(i \leq j\)
\[
|A_{i,j}(\pi)| \leq \left(\frac{\ell + 1}{2i}\right)\left(\frac{n - i - j}{\ell + 1 - i - j}\right)2^i(\ell + 1 - j)! \leq \ell^3iN^\ell+1-i-j.
\]

Lemma 3.2. \([4, \text{Lemma } 2.3]\) Let \(Z_i\) be i.i.d. exponential variables with mean \(\theta > 0\) for \(1 \leq i \leq n\). For \(1/4 \leq \rho \leq 4\), consider the variable
\[
M = M_n = \sup_{1 \leq k \leq n} |\sum_{i=1}^{k}Z_i - \rho k|, \tag{3.5}
\]
Then there exist absolute constants \(c^*, C^* > 0\) such that for all \(r \geq 1\) and \(n \geq r^2\),
\[
e^{-C^*n/r^2} \leq P(M \leq r | \sum_{i=1}^{n}Z_i = \rho n) \leq e^{-c^*n/r^2}
\]

Lemma 3.3. \([4, \text{Lemma } 3.2]\) Let \(Z_i\) be i.i.d. exponential variables with mean \(\theta > 0\) for \(i \in \mathbb{N}\). Consider \(1 \leq r \leq \sqrt{n}\) and \(1 \leq a_1 \leq b_1 \leq a_2 \leq \ldots \leq a_m \leq b_m \leq n\) such that \(q = \sum_{i=1}^{m}(b_i - a_i + 1) \leq n - 1\). Let \(1/4 \leq \rho \leq 1\) and \(M_n\) be defined as in the previous lemma. Then for all \(z_j\) such that
\[
\sum_{j=a_i}^{b_i}z_j - \rho(b_i - a_i + 1) \leq 2r,
\]
we have (write \(A = \bigcup_{i=1}^{m}[a_i, b_i] \cap \mathbb{N}\) and \(p_n = P(M \leq r | \sum_{i=1}^{n}Z_i = \rho n)\))
\[
P(M_n \leq r | \sum_{i=1}^{n}Z_i = \rho n, Z_j = z_j \text{ for all } j \in A) \leq C_3rq^{n}n^{1-10mr}e^{C^*q/r^2}, \tag{3.6}
\]
where \(C^*\) is the constant from Lemma 3.2 and \(C_3 > 0\) is an absolute constant.

Remark 3.4. (1) Notice that the bounds in Lemma 3.2 and 3.3 do not depend on the particular mean of \(Z_i\)’s due to Lemma 2.1. (2) Lemma 3.3 is same as Lemma 3.2 in \([4]\) except that in the latter \(q\) is restricted to be at most \(n - 10r\). But we can easily extend this to all \(q \leq n - 10r\). To see this assume \(n - 1 \geq q \geq n - 10r\). Then the right hand side in \((3.6)\) becomes at least \(C_3pe^{C^*q/r^2}e^{-10C^*/r}\). Now from Lemma 3.2 we get \(p_n e^{C^*q/r^2} \geq 1\). So the right hand side in \((3.6)\) is bigger than \(C_3pe^{-10C^*/r}\) whenever \(n - 1 \geq q \geq n - 10r\). Increasing \(C_3\) if necessary we can make this number bigger than 1 and thus Lemma 3.3 follows.
By second moment computation, we can hope to show that $N_\ell \sim \mathbb{E}N_\ell$ with high probability. Then the main challenge is to prove that a large fraction of the good paths are mutually vertex-disjoint with high probability. To this end, we consider a graph $G_n$ where each vertex corresponds to a good path and an edge is present when the corresponding good paths intersect at one vertex at least. In this context, it is equivalent to the existence of a large independent subset (i.e., a subset that has no edge among them) in the graph. The following simple lemma is sometimes referred to as Turán's theorem, and can be proved simply by employing a greedy algorithm (see, e.g., [6]).

**Lemma 3.5.** Let $G = (V, E)$ be a finite, simple graph with $V \neq \emptyset$. Then $G$ contains an independent subset of size at least $\frac{|V|^2}{2|E| + |V|}$. Notice $2|E|$ is the total degree of vertices in $G$.

In light of the preceding lemma, we wish to show that with high probability the total degree of vertices in $G_n$ is not big. For this purpose, it is desirable to show that the typical number of good paths that intersects with a fixed good path $\pi \in \Pi_\ell^*$ is not big. Thus, we need to estimate $\sum_{\pi' \in \Pi_{\ell,\pi}^*} \mathbb{P}(G_{\pi'}) |G_\pi| \leq C_{\ell} (1 + o(1)) \mathbb{P}(G_\pi) \sqrt{n} \eta e^{-j\eta} e^{1000\eta^2}/\sqrt{\eta}$.

**Proof.** Denote by $S$ and $S'$ the sets $E(\pi) \cap E(\pi')$ and $E(\pi') \setminus E(\pi)$ respectively. By standard calculus, there exists $0 < \eta_1 \leq 1$ such that $1 + e\eta \geq e^{(1+e/2)\eta}$ for all $0 < \eta < \eta_1$.

Note that $\mathbb{P}(G_{\pi'} | G_\pi) = p_1 \cdot p_2$, where

\[
\begin{align*}
p_1 &= \mathbb{P}(\lambda \ell - 1 \leq W(\pi') \leq \lambda \ell | G_\pi), \\
p_2 &= \mathbb{P}(M(\pi') \leq (\zeta_2/\sqrt{\eta}), W(\pi')/\lambda \ell | G_\pi, \lambda \ell - 1 \leq W(\pi') \leq \lambda \ell)
\end{align*}
\]

Since the maximum deviation of a good path from its linear interpolation between starting and ending edges is at most $\zeta_2/\sqrt{\eta}$, conditioned on $G_\pi$ we have (recall that $\pi' \in A_{i,j}$)

\[
\sum_{e \in S} W_e \geq j\zeta_2 - 2i\zeta_2/\sqrt{\eta}.
\]

Consequently when $j \leq \ell - 1$,

\[
\begin{align*}
p_1 &\leq \mathbb{P}(\sum_{e \in S'} W_e \leq \lambda | S'| + 1 + 2i\zeta_2/\sqrt{\eta} | G_\pi) \\
&= \mathbb{P}(\text{Gamma}(\ell - j, 1/n) \leq \lambda(\ell - j) + 1 + 2i\zeta_2/\sqrt{\eta}) \\
&\leq C'_4 \frac{n^{-(\ell-j)}}{n} (\ell - j)^{-1/2} (1 + e\eta)^{\ell-j} e^{2i\zeta_2/\sqrt{\eta}(1+e\eta)},
\end{align*}
\]

where $C'_4 > 0$ is an absolute constant and the last inequality used (2.1). For the second term in the right hand side of (3.8), we can apply (3.3) and Lemma 3.3 to obtain

\[
\begin{align*}
p_2 &\leq C_3 \mathbb{P}(M(\pi) \leq \zeta_2/\sqrt{\eta} | W(\pi) = \lambda \ell) \sqrt{j} \wedge (\ell - j)/\eta 10^{1000\zeta_2/\sqrt{\eta}} e^{Cn e^{10\eta^2}/\sqrt{\eta}},
\end{align*}
\]

when $j \leq \ell - 1$ and $\ell \geq \zeta_2/\eta$ (see the conditions in Lemma 3.3). Using (3.3) again, we get that

\[
\begin{align*}
\mathbb{P}(M(\pi) \leq \zeta_2/\sqrt{\eta} | W(\pi) = \lambda \ell) &= \mathbb{P}(M(\pi) \leq (\zeta_2/\sqrt{\eta}), W(\pi)/\lambda \ell | \lambda \ell - 1 \leq W(\pi) \leq \lambda \ell) \\
&= \mathbb{P}(G_\pi) \mathbb{P}(\lambda \ell - 1 \leq W(\pi) \leq \lambda \ell) \\
&= \mathbb{P}(G_\pi) \mathbb{P}(\lambda \ell - 1 \leq \text{Gamma}(\ell, 1/n) \leq \lambda \ell) \\
&\leq C_4'' (1 + o(1)) \mathbb{P}(G_\pi) \ell! (n/\lambda \ell)^\ell.
\end{align*}
\]
where $C_4'' > 0$ is an absolute constant the last inequality follows from (2.1). Plugging the preceding inequality into (3.10) and using the fact $\ell! \leq \sqrt{\ell(\ell/e)^{\ell}}$ (Stirling’s approximation)

$$p_2 \leq eC_3C_4''(1 + o(1))\mathbb{P}(G_\pi)n^\ell \sqrt{\ell(\ell - j)/\eta(1 + \eta)}^{-\ell}10^{1000\zeta_2}/\eta e^{\zeta_2n}\zeta_2^2.$$

Combined with (3.9), it yields that

$$\mathbb{P}(G_{\pi'}|G_\pi) \leq eC_3C_4''\zeta_2(1 + o(1))\mathbb{P}(G_\pi)\sqrt{\ell/\eta n^j(1 + \eta)}^{-j}10^{1000\zeta_2}/\eta e^{\zeta_2n}\zeta_2^2.$$

Since $\zeta_2 \geq \sqrt{2C^*/e}$ and $\eta < \eta_1$ we have

$$\mathbb{P}(G_{\pi'}|G_\pi) \leq eC_3C_4''(1 + o(1))\mathbb{P}(G_\pi)n^\ell\sqrt{\ell/\eta e^{-j\eta}e^{1000\zeta_2}}.$$

Provided $j \leq \ell - 1$. The case $j = \ell$ can also be easily accommodated. To this end let us first compute $\mathbb{P}(G_\pi)$. It follows from (2.1) and Lemma 3.2 that

$$\mathbb{P}(G_\pi) \geq (1 + o(1))(1 - e^{-1/\lambda})(\lambda \eta/n)^\ell(1/\ell)e^{-C^*n\eta}/\zeta_2^2.$$

Applying Stirling’s formula again, we get that for $\zeta_2 \geq \sqrt{2C^*/e}$ and $\eta < \eta_1$,

$$\mathbb{P}(G_\pi) \geq C_4'''(1 + o(1))n^{-\ell-1/2}e^{\ell\eta},$$

for an absolute constant $C_4''' > 0$. Hence, with the choice of $C_4 = 1/C_4'' \lor eC_3C_4''$ the right hand side of (3.7) is at least 1, and thus (3.7) holds in this case.

**Lemma 3.7.** Let $0 < \zeta_1 < 1/4$ and let $\zeta_2, \ell, \eta$ be the same as stated in Lemma 3.6. Then there exists an absolute constant $C_5 > 0$ such that,

$$\sum_{\pi' \in \Pi_{\ell,\pi}} \mathbb{P}(G_{\pi'}|G_\pi) \leq C_5(1 + o(1))e^{1000\zeta_2}/\eta e^{\zeta_2n},$$

$$\sum_{\pi' \in \Pi_{\ell,\pi}} \mathbb{P}(G_{\pi'}|G_\pi) \leq (1 + o(1))\mathbb{E}N_\ell.$$  

**Proof.** By Lemmas 3.6 and 3.1 we get that for $1 \leq i \leq j \leq \ell$,

$$\sum_{\pi' \in A_{i,j}} \mathbb{P}(G_{\pi'}|G_\pi) \leq (1 + o(1))N^{\ell+1}\mathbb{P}(G_\pi)\xi(\eta,\ell,\pi^i)\mathbb{E}(N_\ell)\xi(\eta,\ell,\pi^i)/\eta^{J(i,j)},$$

where $\xi(\eta,\ell,\pi^i, j)$ is a number depending only on $(\eta, \ell, i, j)$ (so in particular, $\xi(\eta, \ell, i, j)$ does not depend on $n$). It is also clear that

$$\sum_{\pi' \in A_{i,0}} \mathbb{P}(G_{\pi'}|G_\pi) \leq \sum_{\pi' \in A_{i,0}} \mathbb{P}(G_{\pi'}) \leq \mathbb{E}N_\ell.$$

Combined with (3.13), it yields (3.12). It remains to prove (3.11). To this end, we note that the major contribution to the term $\sum_{\pi' \in \Pi_{\ell,\pi}} \mathbb{P}(G_{\pi'}|G_\pi)$ comes from those paths $\pi'$ with $\theta(\pi, \pi') = 1$ or $|V(\pi') \cap V(\pi)| = 1$. Thus, we revisit (3.13) for the case of $i = 1$. By Lemmas 3.6 and 3.1 again, we get that

$$\sum_{1 \leq j \leq \ell} \sum_{A_{1,j}} \mathbb{P}(G_{\pi'}|G_\pi) \leq 2C_4(1 + o(1))e^{1000\zeta_2}/\eta e^{\zeta_2n}\mathbb{P}(G_\pi)n^{-j}\sum_{1 \leq j \leq \ell} e^{-j\eta}(1 - \zeta(1 - \zeta)^{-j}) \leq 2C_4(1 + o(1))e^{1000\zeta_2}/\eta e^{\zeta_2n}\mathbb{E}(N_\ell)(n(1 - e^{-2}))^{-1} \leq 8C_4(1 + o(1))e^{1000\zeta_2}/\eta e^{\zeta_2n}\mathbb{E}(N_\ell)n^{-1},$$

(3.14)
where the last two inequalities follow from the facts that $\zeta_i < 1/4$ and $e^{-\eta/2} \leq 1 - \eta/4$ whenever $0 < \eta < 1$. We still need to consider paths that share vertices with $\pi$ but no edges. For $1 \leq i \leq \ell$, define $B_i$ to be the collection of paths which shares $i$ vertices with $\pi$ but no edges, i.e.,

$$B_i = \{ \pi' \in \Pi_i^* : |V(\pi') \cap V(\pi)| = i, E(\pi') \cap E(\pi) = \emptyset \}.$$  

We need an upper bound on the size of $B_i$. To this end notice that there are $\binom{\ell+1}{i}$ many choices for $V(\pi') \cap V(\pi)$ as cardinality of the latter is $i$ and these vertices can be placed along $\pi'$ in at most $\binom{\ell+1}{i}!$ many different ways. Also the number of ways we can choose the remaining $\ell + 1 - i$ vertices is at most $N^{\ell+1-i}$. Multiplying these numbers we get

$$|B_i| \leq \binom{\ell+1}{i}^2 N^{\ell+1-i}.$$

Since the edge sets are disjoint, $\mathbb{P}(G_{\pi'}|G_{\pi}) = \mathbb{P}(G_{\pi})$ for all $\pi' \in B_i$ and $1 \leq i \leq \ell$. So we have

$$\sum_{\pi' \in B_i} \mathbb{P}(G_{\pi'}|G_{\pi}) \leq (1 + o(1)) \binom{\ell+1}{i}^2 i! (1 - \zeta_i \eta)^{-i} \frac{\mathbb{E}(N_i)}{n} \leq (8 + o(1)) \ell^2 \frac{\mathbb{E}(N_i)}{n}.$$  

Combined with (3.14), it completes the proof of (3.11). \hfill \Box

We will now proceed with our plan of finding a large independent subset of $\mathcal{G}_n$. For any two paths $\pi$ and $\pi'$ in $\Pi_i^*$, define an event

$$H_{\pi, \pi'} = \begin{cases} G_{\pi} \cap G_{\pi'} & \text{if } V(\pi) \cap V(\pi') \neq \emptyset, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Writing $N'_\ell = \sum_{\pi, \pi' \in \Pi_i^*} H_{\pi, \pi'}$, we see that $N'_\ell = 2|E(G_n)| + |V(G_n)|$. Also notice that $N_\ell = |V(G_n)|$. As an immediate consequence of Lemma 3.7, we can compute an upper bound of $\mathbb{E}(N'_\ell)$ as follows:

$$\mathbb{E}(N'_\ell) = \sum_{\pi \in \Pi_\ell} \mathbb{P}(G_\pi) \sum_{\pi' \in \Pi_i^*} \mathbb{P}(G_{\pi'}|G_{\pi}) \leq C_5 (1 + o(1)) e^{1000 \xi_2 \sqrt{\eta} \ell^2 / n^3}.$$  

In view of Lemma 3.5, we now wish to control the deviation of $N_\ell$ and $N'_\ell$. For this purpose, it suffices to show that $\mathbb{E}(N'_\ell)^2 = (\mathbb{E}(N'_\ell)^2 (1 + o(1)))$ and $\mathbb{E}(N'_\ell)^2 = (\mathbb{E}(N'_\ell)^2 (1 + o(1)))$. The latter has already been addressed by (3.12). For the former we need to estimate contributions from terms like $\mathbb{P}(H_{\pi_1, \pi_2} \cap H_{\pi_3, \pi_4})$ in the second moment calculation for $N'_\ell$. Our next lemma will be useful for this purpose.

**Lemma 3.8.** Let $\pi_1, \pi_2, \pi_3, \pi_4$ be paths in $\Pi_\ell^*$ such that $|E(\pi_3 \cup \pi_4)| = 2\ell - j$ and $|E(\pi_1 \cup \pi_2) \cap E(\pi_3 \cup \pi_4)| = j'$ where $0 \leq j \leq \ell$ and $1 \leq j' \leq 2\ell - j$. Also assume that $V(\pi_3) \cap V(\pi_4) \neq \emptyset$. Then,

$$|V(\pi_3) \cap V(\pi_4)| + |V(\pi_3 \cup \pi_4) \cap V(\pi_1 \cup \pi_2)| \geq j + j' + 2.$$  

**Proof.** We orient each path from left to right in the order of appearance of vertices in its defining sequence. Denote by $H$ the graph $(\pi_1 \cup \pi_2) \cap (\pi_3 \cup \pi_4)$ and by $S$ the graph $\pi_3 \cap \pi_4$. Suppose $S$ has exactly $k + 1$ components and let $v_1, v_2, \ldots, v_k+1$ be their right endpoints along $\pi_4$. Also let $v_{k+1}$ be the rightmost vertex of $\pi_4$ that belongs to $S$. It follows from these definitions that for all $1 \leq i \leq k$ (when $k > 0$) the edge $e_i$ having $v_i$ as its left endpoint along $\pi_4$ lies in $E(\pi_4) \setminus E(\pi_3)$. We also have $|V(S)| = j + k + 1$ since $E(S) = j$ and $S$ is acyclic with $k + 1$ components. Now define a new graph $\overline{H}$ with $V(\overline{H}) = V(H)$ and $E(\overline{H}) = E(H) \setminus \bigcup_{1 \leq i \leq k} \{ e_i \}$. Due to acyclicity of $\pi_3$ and $\pi_4$, any cycle $C$ in $\overline{H}$ must contain edges from both $E(\pi_4)/E(\pi_3)$ and $E(\pi_3)/E(\pi_4)$. Thus $C$ must also contain an $e_i$ for some $1 \leq i \leq k$. Consequently $\overline{H}$ is acyclic with at least $j' - k$ many edges. Thus $|V(\overline{H})| \geq j' - k + 1$. Adding this to $|V(S)| = j + k + 1$, we get (3.17). \hfill \Box
We will now use (3.17) and Lemma 3.7 to show that $N_\ell$ and $N'_\ell$ concentrate around their expected values.

**Lemma 3.9.** Assume the same conditions on $\zeta_1, \zeta_2, \ell$ and $\eta$ as in Lemma 3.7. Then there exists $g_{\ell, \eta} : \mathbb{N} \to [0, \infty)$ depending on $\ell$ and $\eta$ with $g_{\ell, \eta}(n) \to 0$ as $n \to \infty$ such that the following hold:

1. $P(|N_\ell - EN_\ell| \leq g_{\ell, \eta}(n)EN_\ell) \to 1$ as $n \to \infty$;
2. $P(|N'_\ell - EN'_\ell| \leq g_{\ell, \eta}(n)EN'_\ell) \to 1$ as $n \to \infty$.

**Proof.** The proof of (1) is rather straightforward. By (3.12), we see that

$$E(N_\ell^2) = \sum_{\pi \in \Pi_\ell^*} P(G_\pi) = \sum_{\pi' \in \Pi_\ell^*} P(G_{\pi'}) = E(N_\ell) \sum_{\pi' \in \Pi_\ell^*} P(G_{\pi'}) \leq E(N_\ell)^2 (1 + o(1)).$$

An application of Markov’s inequality then yields Part (1). In order to prove Part (2), we first argue that $E(N'_\ell) = \Theta(n)$. Similar to the computation of (2.2), we can show that $E(N_\ell)$ is $O(n)$. But then (3.16) tells us that same is also true for $E(N'_\ell)$. For the lower bound, notice that given any path $\pi_1$ in $\Pi_\ell^*$, there are $\Theta(n^\ell)$ many paths in $\Pi'_\ell$ that intersect $\pi_1$ in exactly one vertex. Furthermore for any such pair $(\pi_1, \pi_2)$ we have

$$P(H_{\pi_1, \pi_2}) = (P(G_\pi))^2 = \Theta(n^{-2\ell}),$$

where the last equality follows from (2.1) (see the computation in (2.2)) and Lemma 3.2. Therefore, we obtain that

$$EN'_\ell = \Theta(n^{\ell+1}) \sum_{\pi_2 \in \pi_\ell^*} P(G_{\pi_1} \cap G_{\pi_2}) \geq \Theta(n^{\ell+1}) \Theta(n^\ell) \Theta(n^{-2\ell}) = \Theta(n).$$

Next we estimate $E(N'_\ell^2)$. For this purpose, we first consider two fixed $\pi_1, \pi_2 \in \Pi_\ell^*$. For $0 \leq j \leq \ell$ and $1 \leq j' \leq 2\ell - j$, let $\Pi^{\ell, j, j'}_{\pi_1, \pi_2}$ be the collection of all pairs of paths $(\pi_3, \pi_4) \in \Pi_\ell^*$ such that $|E(\pi_1 \cup \pi_2) \cap E(\pi_3 \cup \pi_4)| = j'$ and $|E(\pi_3 \cup \pi_4)| = 2\ell - j$. For $(\pi_3, \pi_4) \in \Pi^{\ell, j, j'}_{\pi_1, \pi_2}$, we see that $|E(\pi_3 \cup \pi_4) \setminus E(\pi_1 \cup \pi_2)| = 2\ell - j - j'$ and thus by a similar reasoning as employed in (2.3) we get

$$P(H_{\pi_3, \pi_4} | H_{\pi_1, \pi_2}) = O(n^{j+\ell-j'-2\ell}).$$

Now let $\Pi^{\ell, j, j'}_{\pi_1, \pi_2}(n_1, n_2) \subseteq \Pi^{\ell, j, j'}_{\pi_1, \pi_2}$ contain all the pairs $(\pi_3, \pi_4)$ such that $n_1 = |V(\pi_3) \cap V(\pi_4)|$ and $n_2 = |V(\pi_3 \cup \pi_4) \setminus V(\pi_1 \cup \pi_2)|$. Then $|V(\pi_3 \cup \pi_4) \setminus V(\pi_1 \cup \pi_2)| = 2\ell + 2 - n_1 - n_2$ and consequently $|\Pi^{\ell, j, j'}_{\pi_1, \pi_2}(n_1, n_2)| = O(n^{2\ell+2-n_1-n_2})$. By Lemma 3.8, we know that for $n_1 + n_2 \geq j + j' + 2$ for $(\pi_3, \pi_4) \in \Pi^{\ell, j, j'}_{\pi_1, \pi_2}(n_1, n_2)$. Therefore,

$$\sum_{(\pi_3, \pi_4) \in \Pi^{\ell, j, j'}_{\pi_1, \pi_2}} P(H_{\pi_3, \pi_4} | H_{\pi_1, \pi_2}) = \sum_{1 \leq n_1, n_2 \leq \ell+1} \sum_{(\pi_3, \pi_4) \in \Pi^{\ell, j, j'}_{\pi_1, \pi_2}(n_1, n_2)} P(H_{\pi_3, \pi_4} | H_{\pi_1, \pi_2}) = O(1).$$

This implies that

$$\sum_{(\pi_1, \pi_2), (\pi_3, \pi_4)} P(H_{\pi_1, \pi_2} \cap H_{\pi_3, \pi_4}) = O(1)EN'_\ell,$$

where the sum is over all such pairs such that $|E(\pi_1 \cup \pi_2) \cap E(\pi_3 \cup \pi_4)| \neq \emptyset$. In addition,

$$\sum_{(\pi_1, \pi_2), (\pi_3, \pi_4)} P(H_{\pi_1, \pi_2} \cap H_{\pi_3, \pi_4}) = (1 + o(1))(EN'_\ell)^2,$$

where the sum is over all such pairs such that $|E(\pi_1 \cup \pi_2) \cap E(\pi_3 \cup \pi_4)| = \emptyset$ (thus in this case $H_{\pi_1, \pi_2}$ is independent of $H_{\pi_3, \pi_4}$). Combined with the fact that $EN'_\ell = \Theta(n)$, it gives that $E(N'_\ell)^2 = (1 + o(1))(EN'_\ell)^2$. At this point, another application of Markov’s inequality completes the proof of the lemma. □
We are now well-equipped to prove the main lemma of this subsection. For convenience of notation, write

\[ f(\ell, \eta) = e^{-1000 \zeta_2 / \sqrt{\eta}} \sqrt{\eta^3 / \ell^2}. \] (3.18)

**Lemma 3.10.** Assume the same conditions on \( \zeta_1, \zeta_2, \ell \) and \( \eta \) as in Lemma 3.7. Let \( S_{\eta, \ell} \) be a set with maximum cardinality among all subsets of \( \Pi_{\ell}^* \) containing only pairwise disjoint good paths. Then there exists an absolute constant \( C_6 > 0 \) such that,

\[ \mathbb{P}(|S_{\eta, \ell}| \geq C_6 f(\ell, \eta)n) \to 1 \text{ as } n \to \infty. \] (3.19)

**Proof.** Let \( h(\ell, \eta) = C_5 e^{1000 \zeta_2 / \sqrt{\eta}} \sqrt{\eta^3 / \ell^2} \). By Lemma 3.9 and (3.16), we assume without loss that

\[ |N_\ell - \mathbb{E}(N_\ell)| \leq \mathbb{E}(N_\ell)g_{\ell, \eta}(n) \text{ and } N'_\ell \leq h(\ell, \eta) \frac{\mathbb{E}(N_\ell)^2}{n}(1 + g_{\ell, \eta}(n)), \]

where \( g_{\ell, \eta}(n) \) is defined as in Lemma 3.9. Since \( N'_\ell = 2|E(G_n)| + |V(G_n)| \), by Lemma 3.5 we get that the graph \( G_n \) has an independent subset of size at least

\[ N^2_\ell / N'_\ell = n(1 + o(1))/h(\ell, \eta). \]

Therefore, with high probability \( |S_{\eta, \ell}| \geq n/2h(\ell, \eta) \) which leads to (3.19) for \( C_6 = 1/2C_5 \). \( \square \)

### 3.2 Connecting short light paths into a long one

By Lemma 3.19 throughout this subsection we assume without loss that there exist \( C_6 f(\ell, \eta)n \) many disjoint good paths in \( \Pi_{\ell}^* \). The remaining part of our scheme is to connect a fraction of these disjoint good paths in a suitable way to form a light and long path \( \gamma \). In order to describe our algorithm for the construction of \( \gamma \), we need a few more notations. Denote a collection of disjoint good paths from \( \Pi_{\ell}^* \) with maximum cardinality by \( \Pi_{G, \ell, \delta}^* \) and denote the vertex sets \( V(W_n^*) \) and \( V(W_n) \setminus V(W_n^*) \) by \( V_1 \) and \( V_2 \) respectively. Fix an integer \( U > 0 \). Let \( \delta > 0 \) be a number satisfying \( \delta n / \ell \leq |\Pi_{G, \ell, \delta}^*| \) and \( \delta n U / \ell \leq |V_2| \). Now label the paths in \( \Pi_{G, \ell, \delta}^* \) as \( \pi_1, \pi_2, \ldots \) in some arbitrary way. Our aim is to build up the path \( \gamma \) in step-by-step fashion starting from \( \pi_1 \). In each step we will connect \( \gamma \) to \( \pi_j \) by a path of length 2 whose middle vertex is in \( V_2 \). These paths will be referred to as *bridges*. To leverage additional flexibility we also demarcate two segments of length \( \lfloor \ell/4 \rfloor \) one on each end of the paths \( \pi_j \)'s which we call *end segments*. These end segments will allow us to “choose” endpoints of \( \pi_j \)'s while connecting them (as such, it is possible that we only keep half of the vertices of \( \pi_j \) in \( \gamma \)). A vertex \( v \) will be said to be adjacent to a path or an edge if it is an endpoint of that path or edge. If an edge \( e \) has exactly one endpoint in \( S \), we denote by \( v_{e, S} \) that endpoint.

**Initialization.** \( \gamma = \pi_1 \), \( P_1 \) is the set of all vertices which are in end segments of \( \pi_j \)'s for \( j \geq 2 \), \( P_2 = V_2 \), \( P_3 = \emptyset \) and designate an end segment of \( \gamma \) as the open end \( \gamma_O \) and let \( v \) be the endpoint of \( \gamma \) not in \( \gamma_O \). Vertices in \( P_2 \) will be the middle vertices of bridges.

Now repeat the following sequence of steps \( \lfloor \delta n / \ell \rfloor - 1 \) times:

**Step 1.** Find the lightest edge \( e \) between \( \gamma_O \) and \( P_2 \), remove \( v_{e, P_2} \) from \( P_2 \) and include it in \( P_3 \). Repeat this step \( U \) times. These edges will be called predecessor edges (so at the end of this step, \( |P_3| = U \)).

**Step 2.** Find the lightest edge between \( P_3 \) and \( P_1 \). Call it \( e' \). Then \( v_{e', P_1} \) comes from an end segment of some path in \( \Pi^*_{G, \ell} \) say \( \pi \).

**Step 3.** The edge \( e' \) and the *unique* predecessor edge adjacent to \( v_{e', P_3} \) defines a path \( b \) of length 2 (so \( b \) connects a vertex in \( \gamma_O \) to a vertex in \( \pi \)). Let \( w \) be the endpoint of \( \pi \) not in the end segment
that $v', P_1$ came from. Then there is a unique path $\gamma'$ in the tree $\gamma \cup b \cup \pi$ between $v$ and $w$. Set $\gamma = \gamma'$ and $\gamma_O$ = the end segment of $\pi$ containing $w$.

**Step 4.** Remove the vertices on the end segments of $\pi$ from $P_1$ and reset $P_3$ at $\emptyset$.

Figure 1: Illustrating an iteration of BRIDGE for $U = 2$ and $\ell = 4$

We set $\zeta_1 = 1/5$ and $\zeta_2 = 1 + \sqrt{2C^*}/e$ for our next lemma. Note that this satisfies the conditions in Lemma 3.10.

**Lemma 3.11.** For any $0 < \eta < \eta_2$ where $\eta_2 > 0$ is an absolute constant there exists positive integers $U = U(\eta)$, $\ell = \ell(\eta)$ and a positive number $\delta = \delta(\eta)$ such that with probability tending to 1 as $n$ tends to infinity, we have that BRIDGE$(U, \ell, \delta)$ produces a path of length at least $e^{-C_7}/\sqrt{n}$ having average weight at most $1/e + 12\eta$. Here $C_7 > 0$ is an absolute constant.

**Proof.** A crucial observation is that, at the beginning of each iteration the edges between $P_2$ and $\gamma_O$ are still unexplored. The same is true for the edges between $P_3$ and $P_1$ at the end of Step 1. Consequently their weights are i.i.d. $\text{Exp}(1/n)$ regardless of the outcomes from the previous iterations. Therefore, all the bridge weights are independent of each other. Now suppose the mean and variance of each bridge weight is bounded above by $2\ell \eta$ and $\sigma^2$ respectively and we emphasize that the latter does not depend on $n$. By Markov’s inequality it then follows that with probability tending to 1 as $n \to \infty$ the total contribution to the weight of $\gamma$ by the bridges does not exceed $3\ell \eta \times \lfloor \delta n/\ell \rfloor$. We are now ready to bound the average weight $A(\gamma)$ of $\gamma$. Let $\ell_i$ be the length of the segment selected by the algorithm in the $i$-th iteration. We see that its weight can be no more than $\lambda \ell_i + 2\zeta_2/\sqrt{\eta}$, since the segment is chosen from a good path of average weight at most
\(\lambda\) and maximum deviation from its linear interpolation at most \(\zeta_2/\sqrt{n}\) (see (3.1) as well as the proof for Lemma 3.10). Thus the total weight of edges in \(\gamma\) from the good paths is bounded by 
\[\lambda L + |\delta n/\ell|.(2\zeta_2/\sqrt{n})\]
where \(L = \sum_{i} \ell_i\). Adding this to the total weight of bridges we get with probability tending to 1 as \(n \to \infty\)
\[W(\gamma) \leq L(1/e + \eta) + |\delta n/\ell| \cdot (2\zeta_2/\sqrt{n}) + |\delta n/\ell|3\ell\eta.\]
Since the algorithm selects at least \(\ell/2\) edges from each of the \(|\delta n/\ell|\) good paths it connects, we have \(\ell_i \geq \ell/2\) for each \(i\) and thus \(L \geq |\delta n/\ell| \times \ell/2\). Therefore,
\[A(\gamma) \leq 1/e + \eta + |\delta n/\ell| \cdot (2\zeta_2/L\sqrt{n}) + |\delta n/\ell|3\ell\eta/L\]
\[\leq 1/e + \eta + 4\zeta_2/\ell\sqrt{n} + 6\eta\]
\[\leq 1/e + 12\eta,\]
provided \(\ell \geq \zeta_2/\eta^{3/2}\). We can assume this restriction on \(\ell\) since Lemma 3.10 remains valid as long as \(\ell \geq \zeta_2/\eta^{3/2}\). Indeed, later we will specify the value of \(\ell\) such that \(\zeta_2/\eta^{3/2}\)

It remains to bound the mean and variance of each bridge weight by \(2\zeta_2/\eta\). Since the algorithm selects at least \(\ell/2\) edges from each of the \(|\delta n/\ell|\) good paths it connects, we have \(\ell_i \geq \ell/2\) for each \(i\) and thus \(L \geq |\delta n/\ell| \times \ell/2\). Therefore,
\[A(\gamma) \leq 1/e + \eta + |\delta n/\ell| \cdot (2\zeta_2/L\sqrt{n}) + |\delta n/\ell|3\ell\eta/L\]
\[\leq 1/e + \eta + 4\zeta_2/\ell\sqrt{n} + 6\eta\]
\[\leq 1/e + 12\eta,\]
provided \(\ell \geq \zeta_2/\eta^{3/2}\). We can assume this restriction on \(\ell\) since Lemma 3.10 remains valid as long as \(\ell \geq \zeta_2/\eta^{3/2}\). Indeed, later we will specify the value of \(\ell\) such that \(\zeta_2/\eta^{3/2}\)

Assume for now that (3.21) holds. Since \(W_{e'}\) is minimum of \(U \times |P_1|\) many independent \(\text{Exp}(1/n)\) random variables, it is distributed as \(\text{Exp}(U/|P_1|/n)\). As for \(W_e\), it is bounded by the maximum weight of the \(U\) predecessor edges. From properties of exponential distributions and description of the algorithm it is not difficult to see that this maximum weight is distributed as \(E_1 + E_2 + \ldots + E_U\), where \(E_i\) is exponential with rate \((|P_2| - i) \times 1/n \times \lfloor \ell/4 \rfloor\).

By (3.20), we can then bound the expected weight of the bridge from above by
\[\frac{1}{C_6\lfloor \ell/4 \rfloor f(\ell,\eta)n} \times \frac{U}{n} \times n + \frac{U}{n} \times \frac{1}{C_6Uf(\ell,\eta)n} \leq \frac{5}{C_6Uf(\ell,\eta)n} + \frac{11U}{\zeta_1\eta^2},\]
where the last inequality holds for \(\ell \geq 20\) and large \(n\) (given \(\eta, U\)). By the same line of arguments, we get that the variance is bounded by a number that depends only on \(\eta, \ell\) and \(U\) (so in particular independent of \(n\)). By straightforward calculation, we see that the right hand side of (3.22) is bounded by \(2\ell\eta\) provided that
\[U \geq 5 \frac{1000\zeta_2/\sqrt{n}}{C_6} \eta^{3/2} + \zeta_1(\ell\eta)^2 \geq 11U.\]
Finally, set \(\ell = e^{C_9}/\sqrt{n}\), \(\delta = e^{-C_9}/\sqrt{n}\) and \(U = e^{C_{10}/\sqrt{n}}\). It is clear then that there exist absolute constants \(\eta_2, C_8, C_9\) and \(C_{10}\) such that \(\ell \geq \zeta_2/\eta^{3/2} \lor \zeta_2^2/\eta\) and (3.21), (3.23) hold whenever \(0 < \eta < \eta_2\). This completes the proof of the lemma.

Combining Lemmas 3.10 and 3.11 completes the proof of the lower bound in Theorem 1.1.
References

[1] D. Aldous. On the critical value for “percolation” of minimum-weight trees in the mean-field distance model. *Combin. Probab. Comput.*, 7(1):1–10, 1998.

[2] D. J. Aldous. Percolation-like scaling exponents for minimal paths and trees in the stochastic mean field model. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 461(2055):825–838, 2005.

[3] A. Dasgupta. *Probability for Statistics and Machine Learning: Fundamentals and Advanced Topics*. Springer Texts in Statistics, 2011.

[4] J. Ding. Scaling window for mean-field percolation of averages. *Ann. Probab.*, 41(6):4407–4427, 2013.

[5] J. Ding, N. Sun, and D. B. Wilson. Supercritical minimum mean-weight cycles. In preparation.

[6] P. Erdős. On the graph theorem of Turán. *Mat. Lapok*, 21:249–251 (1971), 1970.

[7] A. M. Frieze. On the value of a random minimum spanning tree problem. *Discrete Appl. Math.*, 10(1):47–56, 1985.

[8] W. Krauth and M. Mézard. The cavity method and the travelling salesman problem. *Europhys. Lett.*, 8(3):213–218, 1989.

[9] C. Mathieu and D. B. Wilson. The min mean-weight cycle in a random network. *Combin. Probab. Comput.*, 22(5):763–782, 2013.

[10] M. Mézard and G. Parisi. Mean-field equations for the matching and travelling salesman problems. *Europhys. Lett.*, 2:913–918, 1986.

[11] M. Mézard and G. Parisi. A replica analysis of travelling salesman problem. *Journal de Physique*, 47(3):1285–1296, 1986.

[12] J. Wästlund. The mean field traveling salesman and related problems. *Acta Math.*, 204(1):91–150, 2010.