Electronic Journal of Probability

Electron. J. Probab. **20** (2015), no. 124, 1–21.
ISSN: 1083-6489 DOI: 10.1214/EJP.v20-4111

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

Jian Ding[†] Subhajit Goswami[‡]

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 λ. It was shown by Aldous (1998) that the length is of order $\log n$ for $\lambda < 1/\mathrm{e}$ and of order n for $\lambda > 1/\mathrm{e}$. In this paper, we study the near- supercritical regime where $\lambda = \mathrm{e}^{-1} + \eta$ with $\eta > 0$ a small fixed number. We show that there exist two absolute constants $C_1, C_2 > 0$ such that with high probability the length is in between $n\mathrm{e}^{-C_1/\sqrt{\eta}}$ and $n\mathrm{e}^{-C_2/\sqrt{\eta}}$. Our result corrects a non-rigorous prediction of Aldous (2005).

Keywords: Combinatorial optimization; stochastic distance model; percolation.
AMS MSC 2010: 60C05; 60G70.
Submitted to EJP on February 14, 2015, final version accepted on November 1, 2015.

1 Introduction

Let $\mathcal{W}_n$ 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* (SMF$_n$) model. For a (self-avoiding) *path* $\pi = (v_0, v_1, \ldots, v_m)$, define its length $len(\pi)$ and average weight $A(\pi)$ by

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

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 λ, i.e.,

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

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

*Partially supported by NSF grant DMS-1313596.
[†]University of Chicago, United States of America.E-mail: jianding@galton.uchicago.edu
[‡]University of Chicago, United States of America.E-mail: sgoswami@galton.uchicago.edu

Theorem 1.1. *Let $\lambda = 1/\mathrm{e} + \eta$ where $\eta > 0$. Then there exist absolute constants $C_1, C_2, \eta^* > 0$ such that for all $\eta \le \eta^*$,*

$$\lim_{n\to\infty} \mathbb{P}\big(n\mathrm{e}^{-C_1/\sqrt{\eta}} \le L(n,\lambda) \le n\mathrm{e}^{-C_2/\sqrt{\eta}}\big) = 1\,. \tag{1.1}$$

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 $\lambda < \mathrm{e}^{-1}$ and $L(n,\lambda)$ is of order n when $\lambda > \mathrm{e}^{-1}$. 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 λ is around e^{-1} within 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 η with $\eta = \lambda - \mathrm{e}^{-1} \downarrow 0$. Another interesting result proved in [4] states that $L(n,\lambda) \ge n^{1/4}$ in a somewhat similar regime namely $\lambda \ge 1/\mathrm{e} + \beta(\log n)^{-2}$, where $\beta > 0$ is an absolute constant. Notice that substituting $\eta = C(\log n)^{-2}$ in (1.1), we indeed get a fractional power of n. In fact our method should work, subject to some technical modifications, all the way down to $\eta = C(\log n)^{-2}$ for a large absolute constant C. However, we do not attempt any rigorous proof of this in the current paper.

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 have been proved in [5]. It might be relevant to mention here that the method used in [5] could be potentially useful for nailing down the second phase transition detected in [4], namely the transition from $\eta = \alpha(\log n)^{-2}$ to $\eta = \beta(\log n)^{-2}$ where α, β are positive constants.

Another related question is the classical travelling 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 $\mathrm{e}^{-1} < \lambda < \lambda^*$ (here λ^* is the asymptotic value for TSP), interpolating the critical behavior and the extremal case of TSP. A question of the same flavor on steiner tree is given in [13].

One can also look at the maximum size of tree with average weight below a certain threshold, where a phase transition was 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 $\lim_{n\to\infty} \mathbb{P}\big(L(n,\lambda) = O(\log n)\big) = 1$ when $\lambda < 1/\mathrm{e}$ (see also [4, Theorem 1.3]). For $\lambda > 1/\mathrm{e}$, a sprinkling method was employed in [1] to show that with high probability $L(n,\lambda) = \Theta(n)$. The author first proved that with high probability there exist a large number of paths with average weight slightly above $1/\mathrm{e}$ and then used a certain greedy algorithm to connect these paths into a single long path with average weight slightly above $1/\mathrm{e}$. 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 $1/\mathrm{e}$ and with the maximal deviation (defined in (3.2) below) at most $O(\log n)$, thereby yielding that with high probability $L(n, 1/\mathrm{e}) = \Theta((\log n)^3)$. 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.

TSP, where one studies paths (cycles) that visit every single vertex, is in a sense analogous to the question of finding the minimal value λ 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, 9] as well as the sprinkling method employed in [1].

In order to prove the upper bound, our main intuition is that if $L(n,\lambda)$ were greater than $\mathrm{e}^{-C_2/\sqrt{\eta}}n$ then we would have a larger number of short and light paths (a light path refers to a path with small average weight — at most a little above $1/\mathrm{e}$) than we would typically expect. Formally, let $\ell = \frac{c_1}{\eta}$ where c_1 is a small positive constant, and consider the number of paths (denoted by $N_{\eta/c_1,c_2}$) with length ℓ 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 gives $\mathbb{E}N_{\eta/c_1,c_2} = O(1)n\ell\mathrm{e}^{-c_3/\sqrt{\eta}}$ for a positive constant c_3 depending on c_1 and c_2. Now we consider the number of paths (denoted by N_δ) of length $\delta(\lambda)n$ and average weight at most λ. Such paths have two possibilities: (1) The path contains substantially more than $\mathbb{E}N_{\eta/c_1,c_2}$ many downcrossings, which is unlikely by Markov's inequality. (2) The path does not have substantially more than $\mathbb{E}N_{\eta/c_1,c_2}$ many downcrossings. This is also unlikely for the following reasons: (a) A straightforward first moment computation gives that $\mathbb{E}N_\delta = O(n)\mathrm{e}^{c_4\delta n\eta}$ for a constant $c_4 > 0$; (b) The number of downcrossings along a path of this kind, or a random variable that is "very likely" smaller, should dominate a Binomial random variable $\mathrm{Bin}(\delta n/\ell, c_5)$ where $c_5 > 0$ is an absolute constant (since in the random walk bridge, every subpath of size ℓ has a positive chance to have such a downcrossing). If we choose δ suitably large as in Theorem 1.1, 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 $\mathrm{e}^{-c_6\delta n/\ell}$ for a constant $c_6 > 0$ depending in c_1. If we choose c_1 small enough this probability cost kills the growth of $\mathrm{e}^{c_4\delta n\eta}$ in $\mathbb{E}N_\delta$. Therefore, paths of this kind do not exist either. The details 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 λ and is of length $\mathrm{e}^{c_7/\sqrt{\eta}}$ 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 $\mathrm{e}^{-c_8/\sqrt{\eta}}n$ 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 at most a multiple of η, we have to use a more delicate algorithm. The details are carried out in Section 3.

Notation convention. For a graph G, we denote by $V(G)$ and $E(G)$ the set of vertices and edges of G respectively. A path in a graph G is an (finite) ordered tuple of vertices $(v_0, v_1, \cdots, v_m)$, all distinct. For a path $\pi = (v_0, v_1, \cdots, v_m)$, we also use π to denote the graph whose vertices are $v_0, v_1, \cdots, v_m$ and edges are $(v_0, v_1), \cdots, (v_{m-1}, v_m)$. This would be clear from the context. The weight of an edge e in $\mathcal{W}_n$ is denoted by W_e and we define the total weight $W(\pi)$ of a path π as $\sum_{e\in E(\pi)} W_e$. The collection of all paths in $\mathcal{W}_n$ of length $\ell \in [n]$ is denoted as Π_ℓ. We let $\lambda = 1/\mathrm{e} + \eta$ where η is a fixed positive number. A path is called λ-*light* if its average weight is at most λ, and a path is called (λ, C)-*light* if its total weight is at most $\lambda\ell - C\sqrt{\ell}$ where ℓ is length of the path. For nonnegative real or integer valued variables $x_0, x_1, \cdots, x_n$, let S be a statement involving $x_0, x_1, \cdots, x_n$. We say that S holds "for large x_0 (given $x_1, \cdots, x_n$)" or "when x_0 is large (given $x_1, \cdots, x_n$)" if it holds for any fixed values of $x_1, \cdots, x_n$ in their respective domains and $x_0 \geq a_0$ where a_0 is some positive number depending on the fixed values of $x_1, \cdots, x_n$. In case a_0 is an absolute constant, the phrase "(given $x_1, \cdots, x_n$)" will be dropped. We use "for small x_0" or "when x_0 is small" with or without the qualifying phrase "(given $x_1, x_2, \cdots, x_n$)" in similar situations if the statement S holds instead for $0 < x_0 \leq a_0$. Throughout this paper the order notations $O(.), \Theta(.), o(.)$ etc. are assumed to be with respect to $n \to \infty$ while keeping all the other involved parameters (such as ℓ, η etc.) fixed. We will use $C_1, C_2, \ldots$ to denote constants, and each C_i will denote the same number throughout of the rest of the paper.

2 Proof of the upper bound

Let η' be a multiple of η by a constant bigger than 1 whose precise value is to be selected. Set $\ell = \lfloor 1/\eta' \rfloor$ and let $N_{\eta'}$ be the number of "$(\lambda, 1)$-light" paths of length ℓ. We assume $\eta < 1$ so that $\ell \geq 1$. As outlined in the introduction, we shall first control $N_{\eta'}$.

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 $\mathrm{Gamma}(k,\theta)$ is given by

$$f_{\theta,k}(z) = \theta^k z^{k-1} \mathrm{e}^{-\theta z}/(k-1)! \text{ for all } z \geq 0, \theta > 0 \text{ and } k \in \mathbb{N}. \tag{2.1}$$

By (2.1) and the Stirling's formula, we carry out a straightforward computation and get that

$$\begin{aligned} \mathbb{E}N_{\eta'} &= (1+o(1)) \times n^{\ell+1} \times \mathbb{P}\Big(\mathrm{Gamma}(\ell, 1/n) \leq \lambda\ell - \sqrt{\ell}\Big) \\ &= (1+o(1)) \times n^{\ell+1} \times \frac{\mathrm{e}^{-(\lambda\ell-\sqrt{\ell})/n}(\lambda\ell - \sqrt{\ell})^\ell}{\ell! n^\ell} \\ &= (1+o(1)) C_0(\eta) \alpha \mathrm{e}^{\mathrm{e}\eta/\eta'} \sqrt{\eta'} \mathrm{e}^{-\mathrm{e}/\sqrt{\eta'}} n, \end{aligned} \tag{2.2}$$

where $C_0(\eta) \to 1$ as $\eta \to 0$, and α is a positive constant. Furthermore the factors $1 + o(1)$ are strictly less than 1.

We also need a bound on the second moment of $N_{\eta'}$ to control its concentration around $\mathbb{E}N_{\eta'}$. For $\gamma \in \Pi_\ell$, define F_γ to be the event that γ is $(\lambda, 1)$-light. Then clearly we have $N_{\eta'} = \sum_{\gamma\in\Pi_\ell} \mathbf{1}_{F_\gamma}$. In order to compute $\mathbb{E}(N_{\eta'})^2$, we need to estimate $\mathbb{P}(F_\gamma \cap F_{\gamma'})$ for $\gamma, \gamma' \in \Pi_\ell$. In the case $E(\gamma) \cap E(\gamma') = \emptyset$, we have F_γ and $F_{\gamma'}$ independent of each other and thus $\mathbb{P}(F_{\gamma'}|F_\gamma) = \mathbb{P}(F_{\gamma'})$. In the case $|E(\gamma) \cap E(\gamma')| = j > 0$, we have

$$\mathbb{P}(F_{\gamma'}|F_\gamma) \quad \leq \quad \mathbb{P}\Big(\mathrm{Gamma}(\ell - j, 1/n) \leq \lambda\ell\Big) \leq \tfrac{1}{(\ell-j)!} \tfrac{(\lambda\ell)^{\ell-j}}{n^{\ell-j}}. \tag{2.3}$$

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 γ' such that $|E(\gamma) \cap E(\gamma')| = j$ is at most $O(n^{\ell-j})$. Altogether, we obtain that

$$\begin{aligned}
\mathbb{E}N_{\eta'}^2 &= \textstyle\sum_{\gamma,\gamma'\in\Pi_\ell}\mathbb{P}(F_\gamma\cap F_{\gamma'}) = \sum_{\gamma\in\Pi_\ell}\mathbb{P}(F_\gamma)\sum_{\gamma'\in\Pi_\ell}\mathbb{P}(F_{\gamma'}|F_\gamma)\\
&\le \textstyle\sum_{\gamma\in\Pi_\ell}\mathbb{P}(F_\gamma)\Big(\displaystyle\sum_{\gamma':E(\gamma'\cap\gamma)=\emptyset}\mathbb{P}(F_{\gamma'}) + \sum_{1\le j\le\ell}\sum_{\gamma':|E(\gamma'\cap\gamma)|=j}\frac{1}{(\ell-j)!}\frac{(\lambda\ell)^{\ell-j}}{n^{\ell-j}}\Big)\\
&= \textstyle\sum_{\gamma\in\Pi_\ell}\mathbb{P}(F_\gamma)\Big(\displaystyle\sum_{\gamma':E(\gamma'\cap\gamma)=\emptyset}\mathbb{P}(F_{\gamma'}) + \sum_{1\le j\le\ell}\frac{O(n^{\ell-j})}{(\ell-j)!}\frac{(\lambda\ell)^{\ell-j}}{n^{\ell-j}}\Big)\\
&\le \textstyle\sum_{\gamma\in\Pi_\ell}\mathbb{P}(F_\gamma)\Big(\mathbb{E}N_{\eta'}+O(1)\Big) = \mathbb{E}N_{\eta'}\Big(\mathbb{E}N_{\eta'}+O(1)\Big).
\end{aligned}\tag{2.4}$$

Since $\mathbb{E}N_{\eta'}=\Omega(1)$ as implied by (2.2), (2.4) yields that

$$\mathbb{E}N_{\eta'}^2 = (\mathbb{E}N_{\eta'})^2(1+o(1)).\tag{2.5}$$

As a consequence of Markov's inequality (applied to $|N_{\eta'}-\mathbb{E}N_{\eta'}|^2$), we get that

$$\mathbb{P}\big(N_{\eta'}\ge 2\mathbb{E}N_{\eta'}\big)=o(1).\tag{2.6}$$

Next, we set out to show that any long λ-light path should have a large number of subpaths which are $(\lambda,1)$-light. Let π be a path of length δn for some $\delta>0$. Denote its successive edge weights by $X_1,X_2,\dots X_{\delta n}$ and let $S_k=\sum_{i=1}^k X_i$ for $1\le k\le\delta n$. Probabilities of events involving edge weights of π, unless specfically mentioned, will be assumed to be conditioned on "$\{A(\pi)\le\lambda\}$" throughout the remainder of this section. Now divide π into edge-disjoint subpaths of length ℓ (with the last subpath of length possibly less than ℓ in the case ℓ does not divide δn) and denote the k-th subpath by b_k^π for $1\le k\le\delta n/\ell$. Call any such subpath a downcrossing if it is $(\lambda,1)$-light. Let $D_{k,\eta',n}^\pi$ be the event that b_k^π is a downcrossing. The following well-known result about exponential random variables (see, e.g., [3, Theorem 6.6]) will be very useful.

Lemma 2.1. *Let $W_1,W_2,\dots,W_N$ be i.i.d. exponential random variables with mean $1/\theta$, and let $S_k=\sum_{i=1}^k W_i$ for $1\le k\le N$. Then the random vector $(\frac{W_1}{S_N},\dots,\frac{W_{N-1}}{S_N})$ follows $\mathrm{Dirichlet}(\mathbf{1}_N)$ distribution, S_N follows $\mathrm{Gamma}(N;\theta)$ distribution, and they are independent of each other. Here $\mathbf{1}_N$ is the N-dimensional vector whose all entries are 1.*

We will also require the following simple lemma which we prove for sake of completeness.

Lemma 2.2. *Let $Z_1,Z_2,\dots,Z_N$ be i.i.d. exponential random variables with mean 1 and let $S_N=\sum_{i=1}^N S_N$. Then*

$$\mathbb{P}(S_N\ge N+\alpha)\le \mathrm{e}^{-\alpha^2/4N}\ \textit{for all}\ 0<\alpha\le(2-\sqrt{2})N\,,\tag{2.7}$$

$$\mathbb{P}(S_N\le N-\alpha)\le \mathrm{e}^{-\alpha^2/2N}\,,\ \textit{for all}\ \alpha>0\,.\tag{2.8}$$

Proof. By Markov's inequality, we get that for any $\alpha>0$ and $0<\theta<1$,

$$\mathbb{P}(S_N\ge N+\alpha)=\mathbb{P}(e^{\theta S_N}\ge e^{\theta(N+\alpha)})\le \tfrac{\mathrm{e}^{-\theta N-\theta\alpha}}{(1-\theta)^N}.$$

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

As hinted in the introduction, let us begin with the intent to prove that the number of downcrossings along the first half of π (or any fraction of it) dominates a Binomial random variable $\mathrm{Bin}(\delta n/2\ell,p)$ for some positive, absolute constant p. So essentially we need to prove that a subpath b_k^π can be a downcrossing with probability p

regardless of the first $(k-1)\ell$ edges of π that precede it. Now conditional distribution of $X_{(k-1)\ell+1}, X_{(k-1)\ell+2}, \cdots, X_{\delta n}$ given $X_1, X_2, \cdots, X_{(k-1)\ell}$ and $A(\pi) \le \lambda$ is essentially the the distribution of $X_{(k-1)\ell+1}, X_{(k-1)\ell+2}, \cdots, X_{\delta n}$ conditioned on $\sum_{i=(k-1)\ell+1}^{\delta n} X_i \le \lambda\delta n - S_{(k-1)\ell}$. On the other hand we get from Lemma 2.1 that conditional mean and variance of $W(b_k^\pi)$ given $S_{\delta n} - S_{(k-1)\ell} = \mu(\delta n - (k-1)\ell)$ are $\mu\ell$ and $\mu^2\ell(1+o(1))$ respectively for all $\mu > 0$ and $k \le \delta n/2$. Hence it is plausible to expect that probability of the event $\{W(b_k^\pi) \le \Lambda_k(\ell - C\sqrt{\ell})\}$ conditional on any set of values for $X_1, X_2, \cdots, X_{(k-1)\ell}$ is bounded away from 0 for large ℓ and n, where $\Lambda_k = \Lambda_k^\pi = (S_{\delta n} - S_{(k-1)\ell})/(\delta n - (k-1)\ell)$ and C is some positive number. Let us denote the event $\{W(b_k^\pi) \le \Lambda_k(\ell - C\sqrt{\ell})\}$ by $A_{k,\eta',\delta,n}^{C,\pi}$. Thus it seems more immediate to prove the stochastic domination for number of occurrences of $A_{k,\eta',n}^{C,\pi}$'s which, for the time being, can be treated as a "proxy" for the number of downcrossings. The formal statement is given in the next lemma where we use 6 as the value of C since this allows us to avoid unnecessary named variables and also suits our specific needs for the computations carried out at the end of this section.

Lemma 2.3. *Let $N_{\eta',n}^\pi$ be the number of occurrences of events $A_{k,\eta',n}^\pi = A_{k,\eta',n}^{6,\pi}$ for $1 \le k \le \delta n/2\ell$. Then for any $0 < \eta' < \eta_0$ where η_0 is a positive, absolute constant and any $0 < \delta_0 < 1$ there exists a positive integer $n_d = n_d(\delta_0, \eta')$ and an absolute constant $c > 0$ such that for all $\delta \ge \delta_0$ and $n \ge n_d$ the conditional distribution of $N_{\eta',n}^\pi$ given $\{A(\pi) \le \lambda\}$ stochastically dominates the binomial distribution $\mathrm{Bin}(\delta n/2\ell, c)$.*

Proof. Notice that it suffices to prove that there exist positive absolute constants ℓ_0, c such that uniformly for $\mu > 0$, $\ell \ge \ell_0$ and large L (given ℓ)

$$\mathbb{P}(S_\ell \le \tfrac{S_L}{L}(\ell - 6\sqrt{\ell}) | S_L = \mu L) \ge c\,.$$

To this end, we see that for $L > \ell$

$$\begin{aligned}\mathbb{P}(S_\ell \le \tfrac{S_L}{L}(\ell - 6\sqrt{\ell}) | S_L = \mu L) &= \mathbb{P}(\tfrac{S_\ell}{S_L} \le (\ell - 6\sqrt{\ell})/L | S_L = \mu L) \\ &= \mathbb{P}(\tfrac{S_\ell}{S_L} \le (\ell - 6\sqrt{\ell})/L)\,, \end{aligned} \tag{2.9}$$

where the last equality follows from Lemma 2.1. Since distribution of $\frac{S_\ell}{S_L}$ 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

$$\mathbb{P}\big(S_L/L \le 1 - 1/(2\sqrt{\ell})\big) \le \mathrm{e}^{-L/8\ell}\,.$$

So for $\ell - 6\sqrt{\ell} > 0$, we get

$$\mathbb{P}(S_\ell \le \tfrac{S_L}{L}(\ell - 6\sqrt{\ell})) \ge \mathbb{P}(S_\ell \le \ell - 6.5\sqrt{\ell}) - \mathrm{e}^{-L/8\ell}. \tag{2.10}$$

By central limit theorem there exist absolute numbers $\ell_0, c' > 0$ such that $\mathbb{P}(S_\ell \le \ell - 6.5\sqrt{\ell}) \ge c'$ for $\ell \ge \ell_0$. Hence from (2.10) it follows that for any $\ell \ge \ell_0$ there exists $L_0 = L_0(\ell)$ such that the right hand side of (2.9) is at least $c = 0.99c'$ for $L \ge L_0$. □

Now what remains to show is that the number of downcrossings $\tilde{N}_{\eta',n}^\pi$ along π is bigger than $N_{\eta',n}^\pi$ with high probability. Notice that the occurrence of $A_{k,\eta',n}^\pi \setminus D_{k,\eta',n}^\pi$ implies that Λ_k must be "significantly" above λ. But that can only be caused by a substantial drop in S_k for some $1 \le k \le \delta n/2$, an event that occurs with small probability.

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

$$\mathbb{P}(E_{\eta',n}^\pi | A(\pi) \le \lambda) \le 2n\mathrm{e}^{-\delta n \eta'/16} \text{ for all } \delta \ge \delta_0 \text{ and } n \ge n_s\,. \tag{2.11}$$

Proof. For $1 \le k \le \delta n/2\ell$, let $\ell_k = (k-1)\ell$, $n_s = \lceil 2\ell/\delta_0 \rceil$ and $E^\pi_{k,\eta',n} = \{\Lambda_k \ge \lambda + \sqrt{\eta'}\}$. Assume $n \ge n_s$ so that $\delta n/2\ell \ge 1$. On $E^\pi_{k,\eta',n}$, we have

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

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

$$\mathbb{P}(E^\pi_{k,\eta',n} | A(\pi) \le \lambda) \le \mathbb{P}(S_{\ell_k} \le \tfrac{S_{\delta n}}{\delta n}\big(\ell_k - \sqrt{\eta'}(\delta n - \ell_k)\big)) \tag{2.12}$$

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 $X_1, X_2, \dots X_L$ are i.i.d. exponential variables with mean 1. It is routine to check that

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

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

$$\begin{aligned}\mathbb{P}\Big(S_{\ell_k} \le \frac{S_{\delta n}}{\delta n}\big(\ell_k - \sqrt{\eta'}(\delta n - \ell_k)\big)\Big) &\le \mathbb{P}\Big(S_{\ell_k} \le \ell_k - \sqrt{\eta'}\delta n/4\Big) + \mathbb{P}\Big(\frac{S_{\delta n}}{\delta n} \ge 1 + \sqrt{\eta'}/2\Big) \\ &\le \mathrm{e}^{-\delta n \eta'/16} + \mathrm{e}^{-\delta n \eta'/16},\end{aligned}$$

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

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

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

Proof of Theorem 1.1: upper bound. Assume that $\eta' < 1/4 \wedge \eta_0$ where η_0 is same as given in the statement of Lemma 2.3. Fix a $\delta_0 = \delta_0(\eta')$ in $(0,1)$ and let $n_0 = n_0(\delta_0, \eta') = n_d(\delta_0, \eta') \vee n_s(\delta_0, \eta')$, where n_d, n_s are as stated in Lemmas 2.3 and 2.4 respectively. In the remaining part of this section we will assume that $n \ge n_0$ and $\delta \ge \delta_0$, so that Lemmas 2.3 and 2.4 become applicable. Now let π be a path with length δn. From Lemma 2.4 we get that with large probability $\Lambda_k \le \lambda + \sqrt{\eta'}$ for all k between 1 and $\delta n/2\ell$. But it takes a routine computation to show that $A^\pi_{n,k,\eta'} \setminus \{\Lambda_k \le \lambda + \sqrt{\eta'}\} \subseteq D^\pi_{n,k,\eta'}$ when η' is small. Thus $\tilde{N}^\pi_{\eta',n} \ge N^\pi_{\eta',n}$ except on $E^\pi_{\eta',n}$. Consequently Lemma 2.3 allows us to use binomial distribution to bound quantities like $\mathbb{P}(\tilde{N}^\pi_{\eta',n} \le x)$ with a “small error term” caused by the rare event $E^\pi_{\eta',n}$. Formally,

$$\begin{aligned}\mathbb{P}\Big(\tilde{N}^\pi_{\eta',n} \le 2\mathbb{E}N_{\eta'} | A(\pi) \le \lambda\Big) &\le \mathbb{P}\Big(N^\pi_{\eta',n} \le 2\mathbb{E}N_{\eta'} | A(\pi) \le \lambda\Big) + \mathbb{P}\Big(E_{\eta',n} | A(\pi) \le \lambda\Big) \\ &\le \mathbb{P}\Big(N^\pi_{\eta',n} \le 2\mathbb{E}N_{\eta'} | A(\pi) \le \lambda\Big) + 2n\mathrm{e}^{-\delta n \eta'/16}\,,\end{aligned}$$

where the last inequality follows from Lemma 2.4. Therefore, by Lemma 2.3, we get that

$$\mathbb{P}\Big(\tilde{N}^\pi_{\eta',n} \le 2\mathbb{E}N_{\eta'} | A(\pi) \le \lambda\Big) \le \mathbb{P}\Big(\mathrm{Bin}(\delta n/2\ell, c) \le 2\mathbb{E}N_{\eta'}\Big) + 2n\mathrm{e}^{-\delta n \eta'/16}\,. \tag{2.13}$$

Next let us define a new event as

$$\Xi_{\eta,\delta_0,n} = \textstyle\bigcup_{k \ge \delta_0 n} \bigcup_{\pi \in \Pi_k} \big\{\tilde{N}^\pi_{\eta',n} \ge 2\mathbb{E}N_{\eta'}, A(\pi) \le \lambda\big\}.$$

So $\Xi_{\eta,\delta_0,n}$ is the event that there exists a λ-light path π with $len(\pi) \ge \delta_0 n$ and which contains at least $2\mathbb{E}N_{\eta'}$ many downcrossings. Thus occurrence of $\Xi_{\eta,\delta_0,n}$ implies that

$N_{\eta'} \geq 2\mathbb{E}N_{\eta'}$ which has small probability owing to (2.6). On the other hand if $\Xi_{\eta,\delta_0,n}$ does not occur, $L(n,\lambda) \geq \delta_0 n$ implies the existence of a λ-light path of length at least $\delta_0 n$ that has no more than $2\mathbb{E}N_{\eta'}$ many downcrossings. Formally,

$$\begin{aligned}
&\mathbb{P}\big(L(\lambda,n) \geq \delta_0 n\big) = \mathbb{P}\big(\Xi_{\eta,\delta_0,n}\big) + \mathbb{P}\big(\{L(\lambda,n) \geq \delta_0 n\} \setminus \Xi_{\eta,\delta_0,n}\big) \\
&\leq \mathbb{P}\big(N_{\eta'} \geq 2\mathbb{E}N_{\eta'}\big) + \mathbb{P}\Big(\textstyle\bigcup_{k \geq \delta_0 n} \bigcup_{\pi \in \Pi_k} \big\{\tilde{N}^{\pi}_{\eta',n} \leq 2\mathbb{E}N_{\eta'}, A(\pi) \leq \lambda\big\}\Big) \\
&\leq o(1) + \textstyle\sum_{k \geq \delta_0 n} \sum_{\pi \in \Pi_k} \mathbb{P}\Big(\tilde{N}^{\pi}_{\eta',n} \leq 2\mathbb{E}N_{\eta'} | A(\pi) \leq \lambda\Big)\mathbb{P}\big(A(\pi) \leq \lambda\big)\,.
\end{aligned} \tag{2.14}$$

Now choose $\delta_0 = \delta_0(\eta')$ such that

$$\delta_0 n \eta' c/4 = 2\mathbb{E}N_{\eta'}\,. \tag{2.15}$$

Since $1/\ell \geq \eta'$, we then get from Binomial concentration that for $\delta \geq \delta_0$,

$$\mathbb{P}\Big(\mathrm{Bin}(\delta n/2\ell, c) \leq 2\mathbb{E}N_{\eta'}\Big) \leq \mathrm{e}^{-\delta n \eta' c^2/16}\,.$$

Plugging this into (2.13) we have

$$\mathbb{P}\Big(\tilde{N}^{\pi}_{\eta',n} \leq 2\mathbb{E}N_{\eta'} | A(\pi) \leq \lambda\Big) \leq 2n\mathrm{e}^{-len(\pi)\eta'/16} + \mathrm{e}^{-len(\pi)\eta' c^2/16}\,,$$

whenever $len(\pi) \geq \delta_0 n$. A straightforward computation using (2.1) yields

$$\textstyle\sum_{\pi \in \Pi_k} \mathbb{P}\big(A(\pi) \leq \lambda\big) \leq \dfrac{n}{\sqrt{2\pi k}}\mathrm{e}^{\mathrm{e}k\eta}\,.$$

The last two displays and (2.14) together imply that

$$\mathbb{P}\big(L(\lambda,n) \geq \delta_0 n\big) \leq o(1) + \textstyle\sum_{k \geq \delta_0 n} (2n\mathrm{e}^{-k\eta'/16} + \mathrm{e}^{-k\eta' c^2/16})\mathrm{e}^{\mathrm{e}k\eta}\dfrac{n}{\sqrt{2\pi k}}\,. \tag{2.16}$$

Setting $\eta' = 32\mathrm{e}\eta/c^2$ we get from (2.16),

$$\mathbb{P}\big(L(n,\lambda) \geq \delta_0 n\big) = o(1)\,.$$

It remains to be checked whether δ_0 obtained from (2.15) has the correct functional form as in (1.1). To this end recall from (2.2) that

$$2\mathbb{E}N_{\eta'} \leq 3\alpha \mathrm{e}^{\mathrm{e}\eta/\eta'}\sqrt{\eta'}\mathrm{e}^{-\mathrm{e}/\sqrt{\eta'}}n\,,$$

where η is small enough so that $C_0(\eta)$ in (2.2) is less than $3/2$. Hence $\delta_0 \leq \mathrm{e}^{-C_2/\sqrt{\eta}}$ for some absolute constant C_2 when η is small.

□

3 Proof of the lower bound

3.1 Existence of a large number of vertex-disjoint light paths

As we mentioned in the introduction, the proof of lower bound is divided into two steps. In the first step we split the vertices into two parts and show that there exist a large number of short (i.e. of $O(1)$ length) vertex-disjoint λ-light paths containing vertices from only one part. In the second step we use vertices in the other part as “links” to connect a subcollection of the short paths obtained from step 1 into a long (i.e. of $\Theta(n)$ length) and light path. The current and next subsections are devoted to these two steps in respective order.

In light of the preceding discussion, let us first select a complete subgraph $\mathcal{W}_n^*$ of $\mathcal{W}_n$ containing $n_* = n_{*;\eta,\zeta_1} = (1-\zeta_1\eta)n$ vertices where $\eta, \zeta_1 \in (0,1)$. To be specific we can order the vertices of $\mathcal{W}_n$ in some arbitrary way and define $\mathcal{W}_n^*$ as the subgraph induced by "first" n_* vertices. It will be shown that there are substantially many short and light paths that can be formed with the vertices in $V(\mathcal{W}_n^*)$. We will in fact require slightly more from a path than just being λ-light. For $\pi \in \Pi_\ell$ and some $\zeta_2 > 0$, define

$$G_\pi = G_{\pi;\eta,\zeta_2} = \left\{\lambda\ell - 1 \le W(\pi) \le \lambda\ell, M(\pi) \le (\zeta_2/\sqrt{\eta}).(W(\pi)/\lambda\ell)\right\}, \tag{3.1}$$

where $M(\pi)$ is the maximum deviation of π away from the linear interpolation between the starting and ending edges, formally given by

$$M(\pi) = \sup_{1\le k\le \ell} \left| \textstyle\sum_{i=1}^k W_{e_i} - \frac{k}{\ell} W(\pi) \right| . \tag{3.2}$$

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:

$$\mathbb{P}(M(\pi) \le (\zeta_2/\sqrt{\eta}).(W(\pi)/\lambda\ell) \mid W(\pi) = w) \equiv \text{constant for all } w > 0. \tag{3.3}$$

Call a path $\pi \in \Pi_\ell$ *good* if G_π occurs. Since we are only interested in good paths whose vertices come from $V(\mathcal{W}_n^*)$, we need some related notations. For $\ell \in \mathbb{N}$, denote by $\Pi_\ell^* = \Pi_{\ell;\eta,\zeta_1}^*$ the set of all paths of length ℓ in $\mathcal{W}_n^*$ and by $N_\ell^* = N_{\ell;\eta,\zeta_1,\zeta_2}^*$ the total number of good paths in Π_ℓ^*, i.e., $N_\ell^* = \sum_{\pi\in\Pi_\ell^*} \mathbf{1}_{G_\pi}$. In order to carry out second moment analysis of N_ℓ^* we need to control the correlation between $\mathbf{1}_{G_\pi}$ and $\mathbf{1}_{G_{\pi'}}$ where π, $\pi' \in \Pi_\ell^*$. It is plausible that such correlation depends on the number of common edges between π and π' and in fact bounding the correlation in terms of the number of common edges was sufficient for proving (2.5) in Section 2. But in this case we need an additional measurement instead of just $|E(\pi) \cap E(\pi')|$. This is discussed in detail in [4, 9] and some of their results will be used. Let π be a path in Π_ℓ^* and $S \subseteq E(\pi)$. A segment of π is called an S-component or a component of S if it is a maximal segment of π whose all edges belong to S. Notice that S-components can be defined solely in terms of S. For two paths π and π', define a functional $\theta(\pi,\pi')$ to be the number of S-components where $S = E(\pi)\cap E(\pi')$. As π and π' are self-avoiding, $\theta(\pi,\pi')$ is basically the number of maximal segments shared between π and π'. We refer the readers to Figure 1 for an illustration.

The following result ([4, Lemma 2.9]) relates cardinality of $V(S)$, the union of all endpoints of edges in $S = E(\pi)\cap E(\pi')$, to $\theta(\pi,\pi')$ and $|S|$.

$$|V(S)| = |S| + \theta(\pi,\pi')\,.$$

The pair $\big(\theta(\pi,\pi'), |E(\pi)\cap E(\pi')|\big)$ turns out to be sufficient for bounding the correlation between $\mathbf{1}_{G_\pi}$ and $\mathbf{1}_{G_{\pi'}}$ from above. Consequently it makes sense to partition Π_ℓ^* based on the value of this pair. More formally for $\pi \in \Pi_\ell^*$ and integers $i \le j$, define the set $A_{i,j}$ as

$$A_{i,j} \equiv A_{i,j}(\pi) = \{\pi' \in \Pi_\ell^* : \theta(\pi,\pi') = i, |E(\pi)\cap E(\pi')| = j\}. \tag{3.4}$$

We need a number of lemmas from [4].

Lemma 3.1. *([4, Lemma 2.10]) For any $1 \le \ell \le n_*$ and any $\pi \in \Pi_\ell^*$, we have that for any positive integers $i \le j$*

$$|A_{i,j}(\pi)| \le \binom{\ell+1}{2i}\binom{n_*-i-j}{\ell+1-i-j} 2^i (\ell+1-j)! \le \ell^{3i} n_*^{\ell+1-i-j}.$$

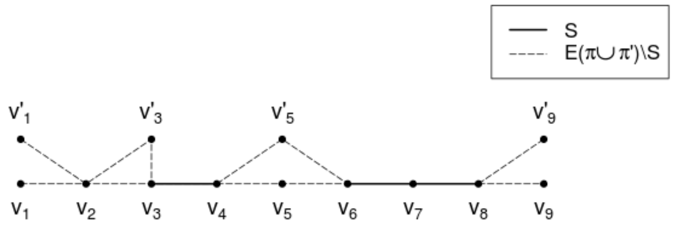


Figure 1 – **Components of the set of edges common to two paths.** In this figure the sequences of vertices $v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9$ and $v_1', v_2, v_3', v_3, v_4, v_5', v_6, v_7, v_8, v_9'$ define the paths π and π' respectively. The dark edges belong to $S = E(\pi) \cap E(\pi')$. Here $\theta(\pi, \pi') = 2$ with the segments (v_3, v_4) and (v_6, v_7, v_8) being the two S-components.

Lemma 3.2. *([4, Lemma 2.3]) Let Z_i be i.i.d. exponential variables with mean $\theta > 0$ for $1 \le i \le \ell$. For $1/4 \le \rho \le 4$, consider the variable*

$$M_\ell = \sup_{1\le k\le \ell} \mid \textstyle\sum_{i=1}^{k} Z_i - \rho k \mid . \tag{3.5}$$

Then there exist absolute constants c^*, $C^* > 0$ such that for all $r \ge 1$ and $\ell \ge r^2$,

$$\mathrm{e}^{-C^*\ell/r^2} \le \mathbb{P}(M_\ell \le r \mid \textstyle\sum_{i=1}^{\ell} Z_i = \rho\ell) \le \mathrm{e}^{-c^*\ell/r^2} .$$

Lemma 3.3. *[4, Lemma 3.2] Let Z_i be i.i.d. exponential variables with mean $\theta > 0$ for $i \in \mathbb{N}$. Consider $1 \le r \le \sqrt{\ell}$ and the integer intervals $[a_1, b_1], [a_2, b_2], \cdots, [a_m, b_m]$ such that $1 \le a_1 \le b_1 \le a_2 \le \cdots \le a_m \le b_m \le \ell$ and $q = \sum_{i=1}^{m}(b_i - a_i + 1) \le \ell - 1$. Let $1/4 \le \rho \le 1$ and M_ℓ be defined as in the previous lemma. Also write $A = \cup_{i=1}^{m}[a_i, b_i] \cap \mathbb{N}$ and $p_\ell = \mathbb{P}(M_\ell \le r \mid \sum_{i=1}^{\ell} Z_i = \rho\ell)$. Then for all z_j such that*

$$\textstyle\sum_{j=a_i}^{b_i} z_j - \rho(b_i - a_i + 1) \le 2r ,$$

we have

$$\mathbb{P}(M_\ell \le r \mid \textstyle\sum_{i=1}^{\ell} Z_i = \rho\ell, Z_j = z_j \text{ for all } j \in A) \le C_3 r\sqrt{q \wedge (\ell - q)}\, p_\ell 10^{100mr} \mathrm{e}^{C^* q/r^2} , \tag{3.6}$$

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) Although the bounds on p_ℓ in Lemma 3.2 do not contain any ρ (as it was restricted to a bounded interval), p_ℓ actually depends on r only through the ratio r/ρ. This follows from an application of Lemma 2.1 with little manipulation. (3) Lemma 3.3 is same as Lemma 3.2. in [4] except that in the latter q is restricted to be at most $\ell - 10r$. But we can easily extend this to all $q \le \ell - 1$. To see this assume $\ell - 1 \ge q \ge \ell - 10r$. Then the right hand side in (3.6) becomes at least $C_3 p_\ell \mathrm{e}^{C^*\ell/r^2}\mathrm{e}^{-10C^*/r}$. Now from Lemma 3.2 we get $p_\ell \mathrm{e}^{C^*\ell/r^2} \ge 1$. So the right hand side in (3.6) is bigger than $C_3\mathrm{e}^{-10C^*}$ whenever $\ell - 1 \ge q \ge \ell - 10r$. 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 $\mathcal{G}_n$ where each vertex corresponds to a good path in $\mathcal{W}_n^*$ and an edge is present whenever the corresponding paths intersect at one vertex at least. Thus the presence of a large number of vertex disjoint good paths in $\mathcal{W}_n^*$ is equivalent to the existence of a large independent subset (i.e., a subset that has no edge among them) in the graph $\mathcal{G}_n$. 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 $|V|^2/(2|E| + |V|)$. Notice $2|E|$ is the total degree of vertices in G.*

In light of Lemma 3.5, we wish to show that with high probability the total degree of vertices in $\mathcal{G}_n$ is not big relative to $|V(\mathcal{G}_n)|$. For this purpose, it is desirable to show that the typical number of good paths that intersect with a fixed good path $\pi \in \Pi_\ell^*$ is not big. Thus, we need to estimate $\sum_{\pi' \in \Pi_{\ell,\pi}^*} \mathbb{P}(G_{\pi'}|G_\pi)$ where $\Pi_{\ell,\pi}^*$ is the collection of all paths π' in $\mathcal{W}_n^*$ sharing at least one vertex with π. Drawing upon the discussions preceding (3.4), we will first estimate $\mathbb{P}(G_{\pi'}|G_\pi)$ for a specific value of the pair $(\theta(\pi,\pi'), |E(\pi) \cap E(\pi')|)$. Our next lemma is very similar to Lemma 3.3 in [4].

Lemma 3.6. *Let $\pi \in \Pi_\ell^*$ and $\pi' \in A_{i,j}$ with $1 \le i \le j \le \ell$. Then there exist absolute constants $\eta_1, C_4 > 0$ such that for $0 < \eta < \eta_1$, $\zeta_2 > 1 \vee \sqrt{2C^*/\mathrm{e}}$ and $\ell \ge \zeta_2^2/\eta$ we have*

$$\mathbb{P}(G_{\pi'}|G_\pi) \le C_4(1 + o(1))\mathbb{P}(G_\pi) n^j \sqrt{\ell/\eta}\, \mathrm{e}^{-j\eta} \mathrm{e}^{1000\zeta_2 i/\sqrt{\eta}}\,. \tag{3.7}$$

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 $0 < \eta_1 \le 1$ such that $1 + \mathrm{e}\eta \ge \mathrm{e}^{(1+\mathrm{e}/2)\eta}$ for all $0 < \eta < \eta_1$. Note that $\mathbb{P}(G_{\pi'} \mid G_\pi) = \mathfrak{p}_1 \cdot \mathfrak{p}_2$, where

$$\begin{aligned} \mathfrak{p}_1 &= \mathbb{P}(\lambda\ell - 1 \le W(\pi') \le \lambda\ell \mid G_\pi)\,, \\ \mathfrak{p}_2 &= \mathbb{P}\big(M(\pi') \le (\zeta_2/\sqrt{\eta}).(W(\pi')/\lambda\ell) \mid G_\pi, \lambda\ell - 1 \le W(\pi') \le \lambda\ell\big)\,. \end{aligned}$$

Since the maximum deviation of a good path from its linear interpolation between starting and ending edges is at most $\zeta_2/\sqrt{\eta}$, the weight of an S-component, say s, is at least $W(\pi)|s|/\ell - 2\zeta_2/\sqrt{\eta}$ when π is good. Here $|s|$ denotes the number of edges in s. Adding over all the $\theta(\pi,\pi')$ components of S we get that $\sum_{e\in S} W_e \ge W(\pi)|S|/\ell - 2\theta(\pi,\pi')\zeta_2/\sqrt{\eta}$ on G_π. As $\pi' \in A_{i,j}$ and weight of a good path is at least $\lambda\ell - 1$, the previous inequality implies that on G_π,

$$\textstyle\sum_{e\in S} W_e \ge \lambda j - 1 - 2i\zeta_2/\sqrt{\eta}\,.$$

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

$$\begin{aligned} \mathfrak{p}_1 &\le \mathbb{P}(\textstyle\sum_{e\in S'} W_e \le \lambda|S'| + 1 + 2i\zeta_2/\sqrt{\eta} \mid G_\pi) && (3.8)\\ &= \mathbb{P}\big(\mathrm{Gamma}(\ell - j, 1/n) \le \lambda(\ell - j) + 1 + 2i\zeta_2/\sqrt{\eta}\big) \\ &\le C_4' n^{-(\ell-j)}(\ell - j)^{-1/2}(1 + \mathrm{e}\eta)^{\ell-j} \mathrm{e}^{\frac{2i\mathrm{e}\zeta_2}{\sqrt{\eta}(1+\mathrm{e}\eta)}}\,, && (3.9) \end{aligned}$$

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

$$\mathfrak{p}_2 \le C_3 \mathbb{P}\big(M(\pi) \le \zeta_2/\sqrt{\eta} \mid W(\pi) = \lambda l\big)\sqrt{j \wedge (\ell - j)/\eta}\, 10^{100 i\zeta_2/\sqrt{\eta}} \mathrm{e}^{C^* j\eta/\zeta_2^2}\,, \tag{3.10}$$

when $j \le \ell - 1$ and $\ell \ge \zeta_2^2/\eta$ (see the conditions in Lemma 3.3). Using (3.3) again, we get that

$$\begin{aligned}\mathbb{P}\big(M(\pi) \le \zeta_2/\sqrt{\eta} \mid W(\pi) = \lambda\ell\big) &= \mathbb{P}\big(M(\pi) \le (\zeta_2/\sqrt{\eta}).(W(\pi)/\lambda\ell) \,|\lambda\ell - 1 \le W(\pi) \le \lambda\ell\big)\\ &= \mathbb{P}(G_\pi)/\mathbb{P}(\lambda\ell - 1 \le W(\pi) \le \lambda\ell)\\ &= \mathbb{P}(G_\pi)/\mathbb{P}\big(\lambda\ell - 1 \le \mathrm{Gamma}(\ell, 1/n) \le \lambda\ell\big)\\ &\le C_4''(1+o(1))\mathbb{P}(G_\pi)\ell!\big(n/\lambda\ell\big)^\ell ,\end{aligned}$$

where $C_4'' > 0$ is an absolute constant and the last inequality follows from (2.1). Plugging the preceding inequality into (3.10) and using the fact $\ell! \le \mathrm{e}\sqrt{\ell}(\ell/\mathrm{e})^\ell$ (Stirling's approximation)

$$\mathfrak{p}_2 \le \mathrm{e}C_3C_4''(1+o(1))\mathbb{P}(G_\pi)n^\ell\sqrt{\ell(\ell-j)/\eta}(1+\mathrm{e}\eta)^{-\ell}10^{100i\zeta_2/\sqrt{\eta}}\mathrm{e}^{C^*j\eta/\zeta_2^2} .$$

Combined with (3.9), it yields that

$$\mathbb{P}(G_{\pi'}|G_\pi) \le \mathrm{e}C_3C_4'C_4''\zeta_2(1+o(1))\mathbb{P}(G_\pi)\sqrt{\ell/\eta}n^j(1+e\eta)^{-j}10^{100i\zeta_2/\sqrt{\eta}}\mathrm{e}^{C^*j\eta/\zeta_2^2}\mathrm{e}^{\frac{2ie\zeta_2}{\sqrt{\eta}(1+e\eta)}} .$$

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

$$\mathbb{P}(G_{\pi'}|G_\pi) \quad\le\quad \mathrm{e}C_3C_4'C_4''(1+o(1))\mathbb{P}(G_\pi)n^j\sqrt{\ell/\eta}\mathrm{e}^{-j\eta}\mathrm{e}^{1000\zeta_2 i/\sqrt{\eta}}$$

provided $j \le \ell - 1$. 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

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

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

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

for an absolute constant $C_4''' > 0$. Hence, with the choice of $C_4 = 1/C_4''' \vee \mathrm{e}C_3C_4'C_4''$ the right hand side of (3.7) is at least 1, and thus (3.7) holds in this case. □

Armed with Lemma 3.6, we can now obtain an upper bound on $\sum_{\pi' \in \Pi^*_{\ell,\pi}} \mathbb{P}(G_{\pi'}|G_\pi)$. Similarly we can bound $\sum_{\pi' \in \Pi^*_\ell} \mathbb{P}(G_{\pi'}|G_\pi)$ which is useful for the computation of $\mathbb{E}((N^*_\ell)^2)$ in view of the following simple observation:

$$\mathbb{E}((N^*_\ell)^2) = \textstyle\sum_{\pi\in\Pi^*_\ell}\mathbb{P}(G_\pi)\sum_{\pi'\in\Pi^*_\ell}\mathbb{P}(G_{\pi'}|G_\pi) = \mathbb{E}(N^*_\ell)\sum_{\pi'\in\Pi^*_\ell}\mathbb{P}(G_{\pi'}|G_\pi) , \tag{3.11}$$

where the last equality follows from the fact that $\sum_{\pi'\in\Pi^*_\ell}\mathbb{P}(G_{\pi'}|G_\pi)$ is independent of π.

Lemma 3.7. *Let $0 < \zeta_1 < 1/4$ and let ζ_2, ℓ, η satisfy the same conditions as stated in Lemma 3.6. Then there exists an absolute constant $C_5 > 0$ such that,*

$$\textstyle\sum_{\pi'\in\Pi^*_{\ell,\pi}}\mathbb{P}(G_{\pi'}|G_\pi) \le C_5(1+o(1))\mathrm{e}^{1000\zeta_2/\sqrt{\eta}}\sqrt{\ell^7/\eta}\frac{\mathbb{E}N^*_\ell}{n} , \tag{3.12}$$

$$\textstyle\sum_{\pi'\in\Pi^*_\ell}\mathbb{P}(G_{\pi'}|G_\pi) \le (1+o(1))\mathbb{E}N^*_\ell . \tag{3.13}$$

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

$$\begin{aligned}\textstyle\sum_{\pi'\in A_{i,j}}\mathbb{P}(G_{\pi'}|G_\pi) &\le (1+o(1))n_*^{\ell+1}\mathbb{P}(G_\pi)\tfrac{\xi(\eta,\ell,i,j,\zeta_1)}{n^i}\\ &\le (1+o(1))\mathbb{E}N^*_\ell\tfrac{\xi(\eta,\ell,i,j,\zeta_1)}{n^i} ,\end{aligned} \tag{3.14}$$

where $\xi(\eta, \ell, i, j, \zeta_1)$ is a number depending only on $(\eta, \ell, i, j, \zeta_1)$ (so in particular, $\xi(\eta, \ell, i, j, \zeta_1)$ does not depend on n). It is also clear that

$$\textstyle\sum_{\pi' \in A_{0,0}} \mathbb{P}(G_{\pi'} | G_\pi) \le \sum_{\pi' \in A_{0,0}} \mathbb{P}(G_{\pi'}) \le \mathbb{E} N_\ell^* \,.$$

Combined with (3.14), it yields (3.13). It remains to prove (3.12). 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 π' with $\theta(\pi, \pi') = 1$ or $|V(\pi') \cap V(\pi)| = 1$. Thus, we revisit (3.14) for the case of $i = 1$. By Lemmas 3.6 and 3.1 again, we get that

$$\begin{aligned}
\textstyle\sum_{1 \le j \le \ell} \sum_{A_{1,j}} \mathbb{P}(G_{\pi'}|G_\pi) &\le 2\frac{C_4}{n}(1+o(1))\mathrm{e}^{\frac{1000\zeta_2}{\sqrt{\eta}}} \sqrt{\tfrac{\ell^7}{\eta}} n_*^{\ell+1} \mathbb{P}(G_\pi) \textstyle\sum_{1 \le j \le \ell} \mathrm{e}^{-j\eta}(1-\zeta_1\eta)^{-j} \\
&\le 2\frac{C_4}{n}(1+o(1))\mathrm{e}^{\frac{1000\zeta_2}{\sqrt{\eta}}} \sqrt{\tfrac{\ell^7}{\eta}} \mathbb{E}N_\ell^* (1 - \mathrm{e}^{-\frac{\eta}{2}})^{-1} \\
&\le 8C_4(1+o(1))\mathrm{e}^{\frac{1000\zeta_2}{\sqrt{\eta}}} \sqrt{\tfrac{\ell^7}{\eta^3}} \tfrac{\mathbb{E}N_\ell^*}{n} \,,
\end{aligned} \tag{3.15}$$

where the last two inequalities follow from the facts that $\zeta_1 < 1/4$ and $\mathrm{e}^{-\eta/2} \le 1 - \eta/4$ whenever $0 < \eta < 1$. We still need to consider paths that share vertices with π but no edges. For $1 \le i \le \ell$, define B_i to be the collection of paths which shares i vertices with π but no edges, i.e.,

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

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 π' in at most $\binom{\ell+1}{i} i!$ 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

$$|B_i| \le \tbinom{\ell+1}{i}^2 i! n_*^{\ell+1-i} \,.$$

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

$$\textstyle\sum_{\pi' \in B_i} \mathbb{P}(G_{\pi'}|G_\pi) \le (1+o(1))\binom{\ell+1}{i}^2 i! (1-\zeta_1\eta)^{-i} \frac{\mathbb{E}N_\ell^*}{n^i} \le (8+o(1))\ell^2 \frac{\mathbb{E}N_\ell^*}{n} \,. \tag{3.16}$$

Combined with (3.15), it completes the proof of (3.12). □

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

$$H_{\pi,\pi'} = H_{\pi,\pi';\eta,\zeta_2} = \begin{cases} G_\pi \cap G_{\pi'} & \text{if } V(\pi) \cap V(\pi') \neq \emptyset \,, \\ \emptyset, & \text{otherwise} \,. \end{cases}$$

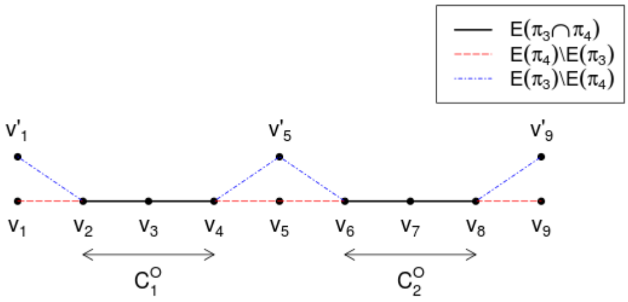
Writing $N'_\ell = N'_{\ell;\eta,\zeta_1,\zeta_2} = \sum_{\pi,\pi' \in \Pi_\ell^*} \mathbf{1}_{H_{\pi,\pi'}}$, we see that $N'_\ell = 2|E(\mathcal{G}_n)| + |V(\mathcal{G}_n)|$. Also notice that $N_\ell^* = |V(\mathcal{G}_n)|$. As an immediate consequence of Lemma 3.7, we can compute an upper bound of $\mathbb{E}N'_\ell$ as follows:

$$\mathbb{E}N'_\ell = \textstyle\sum_{\pi \in \Pi_\ell^*} \mathbb{P}(G_\pi) \sum_{\pi' \in \Pi^*_{\ell,\pi}} \mathbb{P}(G_{\pi'}|G_\pi) \le C_5(1+o(1))\mathrm{e}^{1000\zeta_2/\sqrt{\eta}} \sqrt{\ell^7/\eta^3} \frac{(\mathbb{E}N_\ell^*)^2}{n} . \tag{3.17}$$

If N_ℓ^* and N'_ℓ are concentrated around their respective means in the sense that $N_\ell^* = \mathbb{E}N_\ell^*(1+o(1))$ and $N'_\ell = \mathbb{E}N'_\ell(1+o(1))$ with high probability, then we can use Lemma 3.5 and (3.17) to derive a lower bound on the size of a maximum independent subset of $\mathcal{G}_n$. 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 former has already been addressed by (3.13) (see (3.11)). For the latter 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'_ℓ. Our next lemma will be useful for this purpose.

Lemma 3.8. *Let $\pi_1, \pi_2, \pi_3, \pi_4$ be paths in Π^*_ℓ 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 $0 \le j \le \ell$ and $1 \le j' \le 2\ell - j$. Also assume that $V(\pi_3 \cap \pi_4) \ne \emptyset$. Then,*

$$|V(\pi_3) \cap V(\pi_4)| + |V(\pi_3 \cup \pi_4) \cap V(\pi_1 \cup \pi_2)| \ge j + j' + 2\,. \tag{3.18}$$



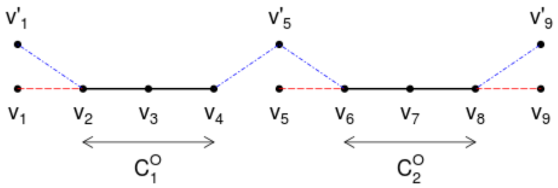


Figure 2 – **Removing edges from union of two paths.** In these figures the sequences of vertices $v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9$ and $v'_1, v_2, v_3, v_4, v'_5, v_6, v_7, v_8, v'_9$ define the paths π_4 and π_3 respectively. C^O_1 and C^O_2 are the two connected components of $\pi_3 \cap \pi_4$. In the figure at the top, the vertices v_4, v_5, v_6, v'_5 define a cycle. After removing the edge (v_4, v_5) from the only segment in $E(\pi_4) \setminus E(\pi_3)$ between C^O_1 and C^O_2, we get an acyclic graph displayed at the bottom.

Proof. Suppose the graph $\pi_3 \cap \pi_4$ has exactly $k+1$ (connected) components namely $C^O_1, \cdots, C^O_{k+1}$. Notice that k is nonnegative as $\pi_3 \cap \pi_4 \ne \emptyset$. Since $|E(\pi_3 \cap \pi_4)| = j$ and $\pi_3 \cap \pi_4$ is acyclic with $k+1$ components, we have that $|V(\pi_3 \cap \pi_4)| = j + k + 1$. Now suppose it were shown that $\pi_3 \cup \pi_4$ can be made acyclic by removing at most k edges while keeping the vertex set same and call this new graph as H. One would then have,

$$|V\big(H \cap (\pi_1 \cup \pi_2)\big)| \ge |E\big(H \cap (\pi_1 \cup \pi_2)\big)| + 1 \ge |E\big((\pi_3 \cup \pi_4) \cap (\pi_1 \cup \pi_2)\big)| - k + 1 = j' - k + 1\,.$$

Adding this to $|V(\pi_3 \cap \pi_4)| = j + k + 1$ would immediately give (3.18). In the remaining part of this proof we will show that one can remove k edges from $\pi_3 \cup \pi_4$ so that the resulting graph becomes acyclic.

Let $\mathcal{C}$ be a cycle in $\pi_3 \cup \pi_4$. Since π_3 and π_4 are acyclic, $\mathcal{C}$ consists of an alternating sequence of segments in $E(\pi_4) \setminus E(\pi_3)$ and $E(\pi_3) \setminus E(\pi_4)$ interspersed with segments in any one of the C_i^O's (possibly trivial i.e. consisting of a single vertex). This implies that for some $1 \le i, i' \le k+1$, $\mathcal{C}$ contains a (nontrivial i.e. of positive length) segment in $E(\pi_4) \setminus E(\pi_3)$ joining C_i^O and $C_{i'}^O$. In fact $i \neq i'$ since π_4 is acyclic. Hence the only case we need to consider is when $k \ge 1$. As π_4 is a path, $C_1^O, C_2^O, \cdots, C_{k+1}^O$'s are vertex-disjoint segments (possibly trivial) aligned along π_4 in some order with k intervening (nontrivial) segments in $E(\pi_4) \setminus E(\pi_3)$. Pick one edge from each of these k segments. It follows from the discussions so far that $\mathcal{C}$ must contain one of these edges. Consequently removing these k edges from $\pi_3 \cup \pi_4$ would make the resulting graph acyclic. We refer the readers to Figure 2 for an illustration. □

We will now use (3.18) and Lemma 3.7 to show that N_ℓ^* and N_ℓ' concentrate around their expected values.

Lemma 3.9. *Assume the same conditions on ζ_1, ζ_2, ℓ and η as in Lemma 3.7. Then there exists $g_{\ell,\eta} = g_{\ell,\eta;\zeta_1,\zeta_2} : \mathbb{N} \mapsto [0,\infty)$ depending on ℓ, η (and ζ_1, ζ_2) with $g_{\ell,\eta}(n) \to 0$ as $n \to \infty$ such that the following hold:*
(1) $\mathbb{P}\big(|N_\ell^* - \mathbb{E}N_\ell^*| \le g_{\ell,\eta}(n)\mathbb{E}N_\ell^*\big) \to 1$ as $n \to \infty$;
(2) $\mathbb{P}\big(|N_\ell' - \mathbb{E}N_\ell'| \le g_{\ell,\eta}(n)\mathbb{E}N_\ell'\big) \to 1$ as $n \to \infty$.

Proof. The proof of (1) is rather straightforward. By (3.11) and (3.13) we see that

$$\mathbb{E}((N_\ell^*)^2) \le (\mathbb{E}N_\ell^*)^2(1 + o(1))\,.$$

An application of Markov's inequality then yields Part (1). In order to prove Part (2), we first argue that $\mathbb{E}N_\ell' = \Theta(n)$. Similar to the computation of (2.2), we can show that $\mathbb{E}N_\ell^*$ is $O(n)$. But then (3.17) tell us that same is also true for $\mathbb{E}N_\ell'$. For the lower bound, notice that given any path π_1 in Π_ℓ^*, there are $\Theta(n^\ell)$ many paths in Π_ℓ^* that intersect π_1 in exactly one vertex. Furthermore for any such pair (π_1, π_2) we have

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

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

$$\mathbb{E}N_\ell' = \Theta(n^{\ell+1})\textstyle\sum_{\pi_2 \in \Pi^*_{\ell,\pi_1}} \mathbb{P}(G_{\pi_1} \cap G_{\pi_2}) \ge \Theta(n^{\ell+1})\Theta(n^\ell)\Theta(n^{-2\ell}) = \Theta(n).$$

Next we estimate $\mathbb{E}((N_\ell')^2)$. For this purpose, we first consider two fixed $\pi_1, \pi_2 \in \Pi_\ell^*$ such that $V(\pi_1) \cap V(\pi_2) \neq \emptyset$. For $0 \le j \le \ell$ and $1 \le j' \le 2\ell - j$, let $\Pi_{\pi_1,\pi_2}^{\ell,j,j'}$ 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 $(\pi_3, \pi_4) \in \Pi_{\pi_1,\pi_2}^{\ell,j,j'}$, 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

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

Now let $\Pi_{\pi_1,\pi_2}^{\ell,j,j'}(n_1,n_2) \subseteq \Pi_{\pi_1,\pi_2}^{\ell,j,j'}$ contain all the pairs (π_3,π_4) such that $|V(\pi_3) \cap V(\pi_4)| = n_1 \ge 1$ and $|V(\pi_3 \cup \pi_4) \cap V(\pi_1 \cup \pi_2)| = n_2$. Then $|V(\pi_3 \cup \pi_4) \setminus V(\pi_1 \cup \pi_2)| = 2\ell + 2 - n_1 - n_2$ and consequently $|\Pi_{\pi_1,\pi_2}^{\ell,j,j'}(n_1,n_2)| = O(n^{2\ell+2-n_1-n_2})$. By Lemma 3.8, we know that for $n_1 + n_2 \ge j + j' + 2$ for $(\pi_3,\pi_4) \in \Pi_{\pi_1,\pi_2}^{\ell,j,j'}(n_1,n_2)$. Therefore,

$$\textstyle\sum_{(\pi_3,\pi_4) \in \Pi_{\pi_1,\pi_2}^{\ell,j,j'}} \mathbb{P}(H_{\pi_3,\pi_4} | H_{\pi_1,\pi_2}) = \sum_{1 \le n_1, n_2 \le 2\ell+2} \sum_{(\pi_3,\pi_4) \in \Pi_{\pi_1,\pi_2}^{\ell,j,j'}(n_1,n_2)} \mathbb{P}(H_{\pi_3,\pi_4} | H_{\pi_1,\pi_2}) = O(1)\,.$$

This implies that

$$\textstyle\sum_{(\pi_1,\pi_2),(\pi_3,\pi_4)} \mathbb{P}(H_{\pi_1,\pi_2} \cap H_{\pi_3,\pi_4}) = O(1)\mathbb{E}N_\ell'\,.$$

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,

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

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 $\mathbb{E}N'_\ell = \Theta(n)$, it gives that $\mathbb{E}((N'_\ell)^2) = (1+o(1))(\mathbb{E}N'_\ell)^2$. 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

$$f(\ell,\eta) = f_{\zeta_2}(\ell,\eta) = \mathrm{e}^{-1000\zeta_2/\sqrt{\eta}}\sqrt{\eta^3/\ell^7}\,. \tag{3.19}$$

Lemma 3.10. *Assume the same conditions on ζ_1, ζ_2, ℓ and η as in Lemma 3.7. Let $S_{n,\eta,\ell} = S_{n,\eta,\ell;\zeta_1,\zeta_2}$ be a set with maximum cardinality among all subsets of Π^*_ℓ containing only pairwise disjoint good paths. Then there exists an absolute constant $C_6 > 0$ such that,*

$$\mathbb{P}(|S_{n,\eta,\ell}| \geq C_6 f(\ell,\eta)n) \to 1 \text{ as } n \to \infty\,. \tag{3.20}$$

Proof. Let $h(\ell,\eta) = C_5\mathrm{e}^{1000\zeta_2/\sqrt{\eta}}\sqrt{\ell^7/\eta^3}$. By Lemma 3.9 and (3.17), we assume without loss of generality that

$$|N^*_\ell - \mathbb{E}N^*_\ell| \leq g_{\ell,\eta}(n)\mathbb{E}N^*_\ell \text{ and } N'_\ell \leq (1+o(1))h(\ell,\eta)\tfrac{(\mathbb{E}N^*_\ell)^2}{n}(1+g_{\ell,\eta}(n)),$$

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

$$N^{*2}_\ell / N'_\ell \geq n(1+o(1))/h(\ell,\eta).$$

Therefore, with high probability $|S_{n,\eta,\ell}| \geq n/2h(\ell,\eta)$ which leads to (3.20) for $C_6 = 1/2C_5$. □

3.2 Connecting short light paths into a long one

We set $\zeta_1 = 1/5$ and $\zeta_2 = 1 + \sqrt{2C^*/\mathrm{e}}$ in this subsection. Note that this choice satisfies the conditions in Lemma 3.10. Denote by $\mathcal{E}_{n,\eta,\ell}$ the event $\{|S_{n,\eta,\ell}| \geq C_6 f(\ell,\eta)n)\}$.
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 γ. In order to describe our algorithm for the construction of γ, we need a few more notations. Denote the vertex sets $V(\mathcal{W}^*_n)$ and $V(\mathcal{W}_n) \setminus V(\mathcal{W}^*_n)$ by V_1 and V_2 respectively. Let $\delta > 0$ be a number and $\nu > 0$ be an integer satisfying

$$1 \leq \delta n/\ell \leq |S_{n,\eta,\ell}| \text{ and } \delta n\nu/\ell \leq |V_2|. \tag{3.21}$$

Now label the paths in $S_{n,\eta,\ell}$ as $\pi_1, \pi_2, \ldots$ in some arbitrary way. Our aim is to build up the path γ in step-by-step fashion starting from π_1. In each step we will connect γ to some π_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 π_j's which we call *end segments*. These end segments will allow us to "choose" endpoints of π_j's while connecting them (as such, it is possible that we only keep half of the vertices of π_j in γ). 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 that endpoint by $v_{e,S}$. The following algorithm, referred to as $\mathrm{BRIDGE}(\nu,\ell,\delta)$, will construct a long path γ. See Figure 3 for an illustration.

Initialization. $\gamma = \pi_1$, T is the set of all vertices which are in end segments of π_j's for $j \geq 2$, $M = V_2$, $P = \emptyset$ and designate an end segment of γ as the open end γ_O. Also let v be the endpoint of γ **not** in γ_O.

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

Step 1. Repeat ν times: find the lightest edge e between γ_O and M, remove $v_{e,M}$ from M and include it in P. These ν edges will be called predecessor edges (so at the end of this step, $|P| = \nu$).

Step 2. Find the lightest edge between P and T. Call it e'. Then $v_{e',T}$ comes from an end segment of some path in $S_{n,\eta,\ell}$, say π.

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

Step 4. Remove the vertices on the end segments of π from T and reset P at $\emptyset$.

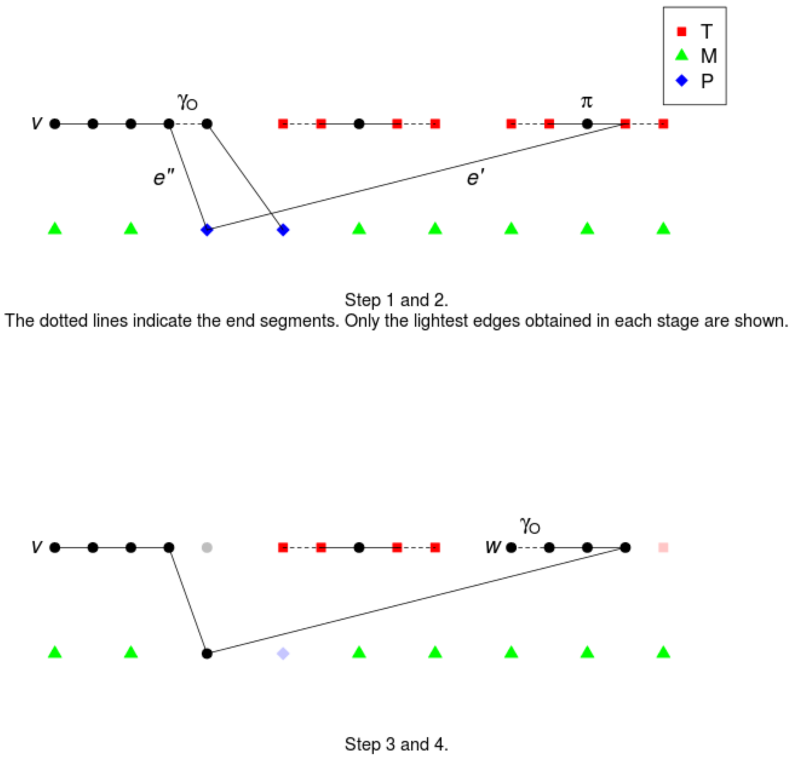


Figure 3 – **Illustrating an iteration of BRIDGE for $\nu = 2$ and $\ell = 4$.** The edges e' and e'' define the path b. So in this iteration the paths γ and π are shortened slightly before being joined via b.

Notice that the conditions in (3.21) ensure that we never run out of vertices in T or M during first $\lfloor \delta n/\ell \rfloor - 1$ iterations of steps 1 to 4. Thus what we described above is a valid algorithm for such choices of δ and ν. Denote the length and average weight of the path γ generated by $\mathrm{BRIDGE}(\nu, \ell, \delta)$ as $L_{\mathrm{BRIDGE}}(\nu, \ell, \delta)$ and $A_{\mathrm{BRIDGE}}(\nu, \ell, \delta)$ respectively when δ, ν, ℓ satisfy these inequalities. For sake of completeness we may define these quantities to be 0 and ∞ respectively and regard the output path γ as "empty" if any one of the inequalities in (3.21) fails to hold. We are now just one lemma short of proving the lower bound in (1.1).

Lemma 3.11. *For any $0 < \eta < \eta_2$ where $\eta_2 > 0$ is an absolute constant there exist positive integers $\nu = \nu(\eta)$, $\ell = \ell(\eta) \geq \zeta_2^2/\eta$ and a positive number $\delta = \delta(\eta)$ such that*

$$\mathbb{P}\big(L_{BRIDGE}(\nu, \ell, \delta) \geq \mathrm{e}^{-C_7/\sqrt{\eta}} n \text{ and } A_{BRIDGE}(\nu, \ell, \delta) \leq 1/\mathrm{e} + 12\eta \mid \mathcal{E}_{n,\eta,\ell}\big) \to 1$$

as n tends to infinity. Here $C_7 > 0$ is an absolute constant.

Proof. We will omit the phrase "conditioned on $\mathcal{E}_{n,\eta,\ell}$" while talking about probabilities in this proof (barring formal expressions) although that is to be implicitly assumed. We use $\mathrm{Exp}(1/\theta)$ to denote the distribution of an exponential random variable with mean $\theta > 0$. Define $\mathcal{B}_{n,\eta,\nu,\ell,\delta}$ to be the event that the total weight of bridges does not exceed $3\ell\eta \times \lfloor \delta n/\ell \rfloor$. Notice that if any one of the inequalities in (3.21) does not hold, γ is "empty" and hence $\mathcal{B}_{n,\eta,\nu,\ell,\delta}$ is a sure event. Suppose δ, ν and ℓ are such that (3.21) is satisfied. We will first bound the average weight $A(\gamma)$ of γ assuming that $\mathcal{B}_{n,\eta,\nu,\ell,\delta}$ occurs. Let ℓ_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 λ and maximum deviation from its linear interpolation is at most $\zeta_2/\sqrt{\eta}$ (see (3.1) as well as the proof for Lemma 3.6). Thus the total weight of edges in γ from the good paths is bounded by $\lambda L + \lfloor \delta n/\ell \rfloor . (2\zeta_2/\sqrt{\eta})$ where $L = \sum_i \ell_i$. Adding this to the total weight of bridges we get with probability tending to 1 as $n \to \infty$

$$W(\gamma) \leq L(1/\mathrm{e} + \eta) + \lfloor \delta n/\ell \rfloor \cdot (2\zeta_2/\sqrt{\eta}) + \lfloor \delta n/\ell \rfloor 3\ell\eta.$$

Since the algorithm selects at least $\ell/2$ edges from each of the $\lfloor \delta n/\ell \rfloor$ good paths it connects, we have $\ell_i \geq \ell/2$ for each i and thus $L \geq \lfloor \delta n/\ell \rfloor \times \ell/2$. Therefore,

$$\begin{aligned} A(\gamma) &\leq 1/\mathrm{e} + \eta + \lfloor \delta n/\ell \rfloor . (2\zeta_2/L\sqrt{\eta}) + \lfloor \delta n/\ell \rfloor 3\ell\eta/L \\ &\leq 1/\mathrm{e} + \eta + 4\zeta_2/\ell\sqrt{\eta} + 6\eta\,. \end{aligned}$$

If $\ell \geq \zeta_2/\eta^{3/2}$, then from the last display we can conclude $A(\gamma) \leq 1/\mathrm{e} + 12\eta$. We can assume this restriction on ℓ for now. Indeed, later we will specify the value of ℓ and it will satisfy the condition $\ell \geq \zeta_2/\eta^{3/2}$.

So it remains to find positive numbers δ, ν, ℓ as functions of η and an absolute constant $\eta_2 > 0$ such that the following three hold for all $0 < \eta < \eta_2$: (a) $\mathbb{P}(\mathcal{B}_{n,\eta,\nu,\ell,\delta} \mid \mathcal{E}_{n,\eta,\ell}) \to 1$ as $n \to \infty$, (b) $\ell \geq \zeta_2/\eta^{3/2} \vee \zeta_2^2/\eta$ (see the statement of the lemma as well as the last paragraph) and (c) γ has the desired length. In the next paragraph we will find a triplet (δ, ν, ℓ) and an absolute constant $\eta_2'' > 0$ such that (a) holds for $0 < \eta < \eta_2''$. In the final paragraph we will show that our choice of (δ, ν, ℓ) also satisfies (b) and (c) whenever $0 < \eta < \eta_2$ where $\eta_2 < \eta_2''$ is an absolute constant.

Let us begin with the crucial observation that, at the start of each iteration the edges between M and γ_O are still unexplored. The same is true for the edges between P and T at the end of Step 1 in any iteration. Consequently their weights are i.i.d. $\mathrm{Exp}(1/n)$ 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 can be bounded above by $2\ell\eta$ and σ^2 respectively and we emphasize that the latter does not depend on n. By Markov's inequality it would then follow that $\lim_{n\to\infty} \mathbb{P}(\mathcal{B}_{n,\eta,\nu,\ell,\delta} \mid \mathcal{E}_{n,\eta,\ell}) = 1$. To that end let us consider the bridge obtained from the m-th iteration where $1 \leq m \leq \lfloor \delta n/\ell \rfloor - 1$. Note that here we implicitly assume (3.21), but this would be shortly shown to be implied by some other constraints involving δ, ν and ℓ. Let e' be the lightest edge between P, T in Step 2 and e be the predecessor edge adjacent to e' (for this iteration). So the bridge weight is simply $W_{e'} + W_e$. By discussions on independence at the beginning of the proof, it follows that $W_{e'}$ and W_e are independent of each other and also of the weights of bridges already chosen. Since these weights are minima of some collections of i.i.d. exponentials, they will be of small magnitude provided that we are minimizing over a large collection of exponentials, i.e., $|T|$, $|M|$ and ν are big. It follows from the description of the algorithm that at each iteration we lose $2\lfloor \ell/4 \rfloor$ many vertices from T and ν many vertices from M. By simple arithmetic we then get,

$$|T| \geq C_6 \lfloor \ell/4 \rfloor f(\ell, \eta) n \text{ and } |M| \geq \zeta_1 \eta n/2\,, \tag{3.22}$$

for all $1 \leq m \leq \lfloor \delta n/\ell \rfloor - 1$ provided

$$\delta \leq C_6 f(\ell, \eta)\ell/2 \text{ and } \nu\delta/\ell \leq \zeta_1 \eta/2\,. \tag{3.23}$$

Notice that these inequalities automatically imply $\delta n/\ell \leq C_6 f(\ell, \eta) n$ and $\delta n\nu/\ell \leq |V_2|$. Thus if δ, ν, ℓ satisfy (3.25), (3.21) would also be satisfied for all large n (given δ, ℓ). Assume for now that (3.23) holds. Since $W_{e'}$ is minimum of $\nu \times |T|$ many independent $\mathrm{Exp}(1/n)$ random variables, it is distributed as $\mathrm{Exp}(\nu|T|/n)$. As for W_e, it is bounded by the maximum weight of the ν 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 $E_1 + E_2 + \ldots E_\nu$, where E_{i+1} is exponential with rate $(|M| - i) \times 1/n \times \lfloor \ell/4 \rfloor$. By (3.22), we can then bound the expected weight of the bridge from above by

$$\tfrac{1}{C_6 \lfloor \ell/4 \rfloor f(\ell,\eta) n} \times \tfrac{1}{\nu} \times n \;+\; \tfrac{\nu}{\left(\frac{\zeta_1 \eta}{2} - \frac{\nu}{n}\right)\lfloor \ell/4 \rfloor} \leq \tfrac{5}{C_6 \nu \ell f(\ell,\eta)} + \tfrac{11\nu}{\zeta_1 \eta \ell}, \tag{3.24}$$

where the last inequality holds for $\ell \geq 20$ and large n (given η, ν). By the same line of arguments, we get that the its variance is bounded by a number that depends only on η, ℓ and ν (so in particular independent of n). To make the right hand side of (3.24) bounded above by $2\ell\eta$, we may require each of the summands in (3.24) to be bounded by $\ell\eta$. After a little simplification this amounts to

$$\nu \geq 5/C_6 \ell^2 \eta f(\ell, \eta)\,, \text{ and } \zeta_1 (\ell\eta)^2 \geq 11\nu\,. \tag{3.25}$$

So we need to pick a positive $\delta = \delta(\eta)$, positive integers $\nu = \nu(\eta), \ell = \ell(\eta)$ and an absolute constant $\eta_2'' > 0$ such that (3.23) and (3.25) hold for $0 < \eta < \eta_2''$. We will deal with (3.25) first which is in fact equivalent to

$$\zeta_1 (\ell\eta)^2/11 \geq \nu \geq 5/C_6 \ell^2 f(\ell, \eta)\,. \tag{3.26}$$

Let us try to find an integer ℓ satsfying $\zeta_1(\ell\eta)^2/11 \geq \left(10/C_6\ell^2 f(\ell,\eta)\right) \vee 2$ since this will ensure the existence of a positive integer ν such that ν, ℓ satisfy (3.26). Using $f(\ell, \eta) = \mathrm{e}^{-1000\zeta_2/\sqrt{\eta}} \sqrt{\eta^3/\ell^7}$, we get that this amounts to

$$\ell \geq \tfrac{C_7' \mathrm{e}^{2000\zeta_2/\sqrt{\eta}}}{\eta^9} \vee \tfrac{C_7''}{\eta}\,,$$

for some positive, absolute constants C_7' and C_7''. Hence there exists an absolute constant $\eta_2''' > 0$ such that the integers $\ell = \lceil \mathrm{e}^{2001\zeta_2/\sqrt{\eta}} \rceil$ and $\nu = \lfloor \zeta_1(\ell\eta)^2/11 \rfloor$ satisfy (3.26) whenever $0 < \eta < \eta_2'''$. Now we need to find δ that would satisfy (3.23) which can be rewritten as,

$$\delta \leq (C_6 f(\ell, \eta)\ell/2) \wedge (\zeta_1 \eta\ell/2\nu)\,. \tag{3.27}$$

Again substituting $f(\ell,\eta) = \mathrm{e}^{-1000\zeta_2/\sqrt{\eta}}\sqrt{\eta^3/\ell^7}$, we can simplify (3.27) to

$$\delta \leq (C_6 \mathrm{e}^{-1000\zeta_2/\sqrt{\eta}}\tfrac{\eta^{3/2}}{2\ell^{5/2}}) \wedge (\zeta_1\eta\ell/2\nu)\,. \tag{3.28}$$

Since $\nu = \lfloor \zeta_1(\ell\eta)^2/11\rfloor$, (3.28) would be satisfied if

$$\delta \leq (C_6 \mathrm{e}^{-1000\zeta_2/\sqrt{\eta}}\tfrac{\eta^{3/2}}{2\ell^{5/2}}) \wedge (11/2\ell\eta)\,.$$

The last display together with our particular choice of ℓ i.e. $\lceil \mathrm{e}^{2001\zeta_2/\sqrt{\eta}}\rceil$ imply that there exists a positive, absolute constant $\eta_2'' < \eta_2'''$ such that $\delta = \mathrm{e}^{-7000\zeta_2/\sqrt{\eta}}$ satisfies (3.27) for $0 < \eta < \eta_2''$. Thus our choice of the triplet (δ,ν,ℓ) satisfies (3.23) and (3.25) for $0 < \eta < \eta_2''$ and consequently the event $\mathcal{B}_{n,\eta,\nu,\ell,\delta}$ occurs with high probability for this choice.

As to the constraint on ℓ, it is also clear that there exists a positive, absolute constant $\eta_2' < \eta_2''$ such that $\ell = \lceil \mathrm{e}^{2001\zeta_2/\sqrt{\eta}}\rceil$ is larger than $\zeta_2/\eta^{3/2} \vee \zeta_2^2/\eta$ for all $0 < \eta < \eta_2'$. Finally it is left to ensure whether γ has the length required by the lemma. Since our particular choice of the triplet (δ,ν,ℓ) satisfies (3.21) for large n (given η), we have that $L_{\mathrm{BRIDGE}}(\nu,\ell,\delta) \geq \lfloor \delta n/\ell \rfloor \times \ell/2$. It then follows that there exists a positive, absolute constant $\eta_2 < \eta_2'$ such that $L_{\mathrm{BRIDGE}}(\nu,\ell,\delta) \geq \mathrm{e}^{-7001\zeta_2/\sqrt{\eta}}n$ for these particular choices of ν,ℓ and δ whenever $0 < \eta < \eta_2$ and n is large (given η). 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. MR-1611094

[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. MR-2121938

[3] A. Dasgupta. Probability for Statistics and Machine Learning: Fundamentals and Advanced Topics. *Springer Texts in Statistics*, 2011. MR-2807365

[4] J. Ding. Scaling window for mean-field percolation of averages. *Ann. Probab.*, 41(6):4407–4427, 2013. MR-3161479

[5] J. Ding, N. Sun, and D. B. Wilson. Preprint, available at http://arxiv.org/abs/1504.00918.

[6] P. Erdős. On the graph theorem of Turán. *Mat. Lapok*, 21:249–251 (1971), 1970. MR-0307975

[7] A. M. Frieze. On the value of a random minimum spanning tree problem. *Discrete Appl. Math.*, 10(1):47–56, 1985. MR-0770868

[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. MR-3094483

[10] M. Mézard and G. Parisi. Mean-field equations for the mathcing 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. MR-2600434

[13] B. Bollobás, D. Gamarnik, O. Riordan and B. Sudakov. On the value of a random minimum weight Steiner tree. *Combinatorica*, 24(2):187–207, 2004. MR-2071332

Acknowledgments. We are grateful to David Aldous for very useful discussions, and we thank an anonymous referee for a careful review of an earlier manuscript and suggesting a simpler proof of Lemma 3.8.