Karp's patching algorithm on random perturbations of dense digraphs

Alan Frieze* Peleg Michaeli
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh PA 15213

March 29, 2025

Abstract

We consider the following question. We are given a dense digraph D_0 with minimum in- and out-degree at least αn, where $\alpha > 0$ is a constant. We then add random edges R to D_0 to create a digraph D. Here an edge e is placed independently into R with probability $n^{-\epsilon}$ where $\epsilon > 0$ is a small positive constant. The edges $E(D)$ of D are given independent edge costs $C = C(e), e \in E(D)$, where C has a density $f(x) = a + bx + o(x)$ as $x \to 0$. Here $a > 0, b$ are constants. The prime examples will be the uniform $[0,1]$ distribution $(a = 1, b = 0)$ and the exponential mean 1 distribution $EXP(1)$ $(a = 1, b = -1)$. Let $C(i,j), i,j \in [n]$ be the associated $n \times n$ cost matrix where $C(i,j) = \infty$ if $(i,j) \notin E(D)$. We show that w.h.p. the patching algorithm of Karp finds a tour for the asymmetric traveling salesperson problem whose cost is asymptotically equal to the cost of the associated assignment problem. Karp's algorithm runs in polynomial time.

1 Introduction

Let $\mathcal{D}(\alpha, n)$ be the set of digraphs with vertex set $[n] = \{1, 2, \ldots, n\}$ and with minimum in- and out-degree at least αn. We are given a digraph $D_0 \in \mathcal{D}(\alpha, n)$ and then we add random edges R to D_0 to create a digraph D. Here an edge e is placed independently into R with probability $n^{-\varepsilon}$ where $\varepsilon > 0$ is a small positive constant. The edges $E(D)$ of D are given independent edge costs $C(e), e \in E(D)$, with a density $f(x)$ where

$$f(x) = a + bx + o(x) \text{ as } x \to 0. \tag{1}$$

Here $a > 0, b$ are constants. The prime examples will be the uniform $[0,1]$ distribution $(a = 1, b = 0)$ and the exponential mean 1 distribution $EXP(1)$ $(a = 1, b = -1)$. Let $C(i,j), i,j \in [n]$ be the associated $n \times n$ cost matrix where $C(i,j) = \infty$ if $(i,j) \notin E(D)$. Here we are interested in using the relationship between the Assignment Problem (AP) and the Asymmetric Traveling Salesperson Problem (ATSP) associated with the cost matrix $C(i,j), i,j \in [n]$ to find a tour whose cost $\widehat{v}(ATSP)$ satisfies $v(AP) \leq v(ATSP) \leq \widehat{v}(ATSP) \leq$

*Research supported in part by NSF grant DMS1952285, email: frieze@cmu.edu

$(1+o(1))v(AP)$ w.h.p., where $v(\bullet)$ denotes the optimal cost. We say that the output of the algorithm is *asymptotically optimal* i.e. it produces a tour whose cost is at most $(1+o(1))$ times optimal

The problem AP is that of computing the minimum cost perfect matching in the complete bipartite graph $K_{n,n}$ when edge (i,j) is given a cost $C(i,j)$. Equivalently, when translated to the complete digraph $\vec{K}_n$ it becomes the problem of finding the minimum cost collection of vertex disjoint directed cycles that cover all vertices. (Here it is simplest to think of loops as cycles of length 1.) The problem ATSP is that of finding a single cycle of minimum cost that covers all vertices. As such it is always the case that $v(ATSP) \geq v(AP)$. Karp [27] considered the case where $D = \vec{K}_n$. He showed that if the cost matrix is comprised of independent copies of the uniform $[0,1]$ random variable $U(1)$ then w.h.p. $v(ATSP) = (1+o(1))v(\text{AP})$. He proves this by the analysis of a *patching* algorithm (see below). Karp's result has been refined in [17], [22] and [28].

Karp's Patching Algorithm: First solve the AP to obtain a minimum cost perfect matching M and let $\mathcal{A}_M = \{C_1, C_2, \ldots, C_\ell\}$ be the associated collection of vertex disjoint cycles covering $[n]$. Then *patch* two of the cycles together, as explained in the next paragraph. Repeat until there is one cycle.

A pair $e = (x,y), f = (u,v)$ of edges in different cycles C_1, C_2 are said to be a *patching pair* if the edges $e' = (u,y), f' = (x,v)$ both exist. In which case we can replace C_1, C_2 by a single cycle $(C_1 \cup C_2 \cup \{e', f'\}) \setminus \{e, f\}$. The edges e, f are chosen to minimise the increase in cost of the set of cycles.

Theorem 1. *Suppose that $D_0 \in \mathcal{D}(\alpha, n)$, $\alpha > 0$ where α is constant. Suppose that D is created by adding random edges R to D_0 and that each edge of D is given an independent cost drawn from distribution* (1). *An edge $e \notin E(D_0)$ is placed independently into R with probability $n^{-\varepsilon}$ where $\varepsilon > 0$ is a sufficiently small positive constant. Then w.h.p. $v(ATSP) = (1+o(1))v(AP)$ and Karp's patching algorithm finds a tour (Hamilton cycle) in polynomial time whose cost $\widehat{v}(ATSP) = (1+o(1))v(AP)$ w.h.p.*

For the moment assume that C is distributed as $EXP(1)$ i.e $\mathbb{P}(X \geq x) = e^{-x}$. We will discuss more general distributions in Section 5.

This model for instances of the ATSP arises in the following context: Karp's heuristic is well understood for the case of the complete digraph with random weights. If we want to understand its performance on other digraphs then we must be sure that the class of digraphs we consider is Hamiltonian w.h.p. The class of digraphs $\mathcal{D}(\alpha, n)$ is a good candidate, but we can only guarantee Hamiltonicity if $\alpha \geq 1/2$. If we want to allow arbitrary α then the most natural thing to do is add $o(n^2)$ random edges, as we have done.

It is often the case that adding some randomness to a combinatorial structure can lead to significant positive change. Perhaps the most important example of this and the inspiration for a lot of what has followed, is the seminal result of Spielman and Teng [39] on the performance of the simplex algorithm, see also Vershynin [41] and Dadush and Huiberts [12].

Spielman and Teng [39] inspired the following model of Bohman, Frieze and Martin [8]. They consider adding random edges to an arbitrary member G of $\mathcal{G}(\alpha)$. Here α is a positive constant and $\mathcal{G}(\alpha)$ is the set of graphs with vertex set $[n]$ and minimum degree at least αn. They show that adding $O(n)$ random edges to G is enough to create a Hamilton cycle w.h.p. This is in contrast to the approximately $\frac{1}{2} n \log n$ edges needed if we rely only on the random edges. Research on this model and its variations has been quite substantial, see for example [4], [5], [6], [7], [9], [10], [13], [16], [23], [30], [31], [32], [36], [37], [38], [40].

Notation Let G denote the bipartite graph with vertex partition $A = \{a_1, a_2, \ldots, a_n\}$, $B = \{b_1, b_2, \ldots, b_n\}$ and an edge $\{a_i, b_j\}$ for every directed edge $(i,j) \in E(D)$. A matching M of G induces a collection $\mathcal{A}_M$ of vertex disjoint paths and cycles in D and vice-versa. If the matching is perfect, then there are only cycles.

The proof requires a number of definitions of values, graphs, digraphs and trees. It might be helpful to the reader if we list them along with their definitions. See Appendix A.

2 Proof of Theorem 1

We begin by analyzing the solution to the AP on the graph G. (We slightly relax the definition of AP to that of finding a minimum cost perfect matching on a bipartite graph.) We prove the following:

Lemma 2. *W.h.p. the optimal solution to the AP contains only edges of cost $C(i,j) \leq \gamma_n = n^{-(1-2\varepsilon)}$.*

(We note that this lemma implies that only $O(n^{1+\varepsilon})$ of the added edges play a role in the algorithm.)

Lemma 3. *W.h.p. after solving the AP, the number ν_C of cycles is at most $r_0 \log n$ where $r_0 = n^{1-3\varepsilon}$.*

Bounding the number of cycles has been the most difficult task. Karp proved that the number of cycles is $O(\log n)$ w.h.p. when we are dealing with the complete digraph $\vec{K}_n$ with independent uniform $[0,1]$ costs. Karp's proof is very clean but it relies on the key insight that in this case the optimal assignment comes from a uniform random permutation. This seems unlikely to be true in our case and this requires building a proof from scratch.

Given Lemmas 2, 3, the proof is straightforward. We can begin by temporarily replacing costs $C(e) > \gamma_n$ by infinite costs before we solve the AP. Lemma 2 implies that w.h.p. we get the same optimal assignment as we would without the cost changes. Having solved the AP, the memoryless property of the exponential distribution, implies that the unused edges in $E(D)$ of cost greater than γ_n have a cost which is distributed as $\gamma_n + EXP(1)$.

Let $\mathcal{C} = C_1, C_2, \ldots, C_\ell$ be a cycle cover and let $k_i = |C_i|$ where $k_1 \leq k_2 \leq \cdots \leq k_\ell$, $2 \leq \ell \leq r_0 \log n$. (There is nothing more to do if $\ell = 1$.) Different edges in C_i give rise to disjoint patching pairs. We ignore the saving associated with deleting the edges e, f of the cycles and only look at the extra cost $C(e') + C(f')$ incurred. We will also only consider the random edges R when looking for a patch. The number of possible patching pairs $\pi_{\mathcal{C}}$ satisfies

$$\pi_{\mathcal{C}} \geq \sum_{i<j} k_i k_j = \frac{1}{2}\left(n^2 - \sum_{i=1}^{\ell} k_i^2\right) \geq \frac{1}{2}\left(n^2 - ((n-\ell+1)^2 + \ell - 1)\right) \geq \frac{\ell(n-1)}{2}.$$

Each of these $\pi_{\mathcal{C}}$ pairs uses a disjoint set of edges. We define the sets

$$R_\ell = \left\{ e \in R : C(e) \leq \gamma_n + \frac{1}{(\ell n^{1-5\varepsilon/2})^{1/2}} \right\}, \ 1 \leq \ell \leq r_0.$$

Each edge of $E(\vec{K}_n) \setminus E(D_0)$ appears in R_ℓ with probability at least $p_\ell = n^{-\varepsilon}\left(\frac{1-o(1)}{\ell n^{1-5\varepsilon/2}}\right)^{1/2}$, independent of other edges. (The factor $n^{-\varepsilon}$ accounts for being included in the random set R. Then if $C(e) > \gamma_n$ we use the memoryless property to get the second factor.) Let $\mathcal{E}_\ell$ be the event that at some stage in the patching process, $|\mathcal{C}| = \ell$ and that there is no patch using only edges in R_ℓ. If $\mathcal{E}_\ell$ does not occur then we reduce the number of cycles by at least one. We have

$$\mathbb{P}(\exists 2 \leq \ell \leq r_0 : \mathcal{E}_\ell) = \sum_{\ell=2}^{r_0} \mathbb{P}\left(\mathcal{E}_\ell \cap \bigcap_{\lambda=\ell+1}^{r_0} \neg\mathcal{E}_\lambda\right) \leq \sum_{\ell=2}^{r_0} \mathbb{P}\left(\mathcal{E}_\ell \mid \bigcap_{\lambda=\ell+1}^{r_0} \neg\mathcal{E}_\lambda\right) \leq \sum_{\ell=2}^{r_0} \frac{\mathbb{P}(\mathcal{E}_\ell)}{1-\sum_{\lambda=\ell+1}^{r_0}\mathbb{P}(\mathcal{E}_\lambda)}$$
$$\leq \sum_{\ell=2}^{r_0} \frac{(1-p_\ell^2)^{\ell(n-1)/2}}{1-\sum_{\lambda=\ell+1}^{r_0}(1-p_\lambda^2)^{\lambda(n-1)/2}} = \sum_{\ell=2}^{r_0} \frac{\left(1-\frac{1-o(1)}{\ell n^{1-\varepsilon/2}}\right)^{\ell(n-1)/2}}{1-\sum_{\lambda=\ell+1}^{r_0}\left(1-\frac{1-o(1)}{\lambda n^{1-\varepsilon/2}}\right)^{\lambda(n-1)/2}} = o(1).$$

W.h.p. the patches involved in these cases add at most the following to the cost of the assignment:

$$2\sum_{\ell=1}^{r_0 \log n}\left(\gamma_n + \frac{1}{(\ell n^{1-5\varepsilon/2})^{1/2}}\right) \leq 2r_0\gamma_n \log n + 2\left(\frac{2r_0}{n^{1-5\varepsilon/2}}\right)^{1/2} = o(1). \tag{2}$$

Given the last equality and the fact that w.h.p. $v(\text{AP}) > (1-o(1))\zeta(2) > 1$ where $\zeta(2) = \sum_{k=1}^\infty k^{-2}$, we see that Karp's patching heuristic is asymptotically optimal. The lower bound of $(1-o(1))\zeta(2)$ on $v(AP)$ comes from [3].

3 Proof of Lemma 2

We show that w.h.p. for any pair of vertices $a \in A, b \in B$ and any perfect matching M between A and B that there is an M-alternating path from a to b that only uses at most $10/\varepsilon$ non-M edges, each of cost at most $\varepsilon\gamma_n/10$. (A path is M-alternating if its edges alternate between being in M and not being in M.) So the difference in cost between added and deleted edges is at most γ_n. We need to prove a slightly more general version where $r \geq r_0$ replaces n, see Lemma 6.

The idea of the proof is based on the fact that w.h.p. the sub-digraph induced by edges of low cost is a good expander. There is therefore a low cost path between every pair of vertices. Such a path can be used to replace an expensive edge.

Chernoff Bounds: We use the following inequalities associated with the Binomial random variable $Bin(N,p)$.

$$\mathbb{P}(Bin(N,p) \leq (1-\theta)Np) \leq e^{-\theta^2 Np/2}.$$
$$\mathbb{P}(Bin(N,p) \geq (1+\theta)Np) \leq e^{-\theta^2 Np/3} \qquad \text{for } 0 \leq \theta \leq 1.$$
$$\mathbb{P}(Bin(N,p) \geq \gamma Np) \leq \left(\frac{e}{\gamma}\right)^{\gamma Np} \qquad \text{for } \gamma \geq 1.$$

Proofs of these inequalities are readily accessible, see for example [21]. We have the same bounds for the Hypergeometric distribution with mean Np. This follows from Theorem 4 of Hoeffding [24].

Assume now that $a_1, a_2, \ldots, a_n$ is a random permutation of A and similarly for B. We let $A_r = \{a_1, a_2, \ldots, a_r\}$ and $B_r = \{b_1, b_2, \ldots, b_r\}$. We let $G_r = (A_r \cup B_r, E_r)$ denote the subgraph of G induced by $A_r \cup B_r$.

Lemma 4. *If $r \geq r_0$ then with probability $1-o(n^{-1})$, (i) G_r has minimum degree at least $\alpha_0 r$ where $\alpha_0 = (1-o(1))\alpha$ and (ii) G_r is connected and (iii) G_r contains a perfect matching.*

Proof. G_r is the union of two bipartite graphs, Γ_0, Γ_1 where Γ_0 comes from D_0 and Γ_1 comes from R. The degree of a vertex in Γ_0 has a hypergeometric distribution with mean $r\alpha$ and so the minimum degree condition follows from the Chernoff bounds above. If m, p satisfy $p = m/2n^2 = n^{-\varepsilon}/2$ then the Chernoff bounds imply that adding edges to D_0 with probability p will add fewer than m random edges w.h.p. On the other hand Frieze [19] showed that w.h.p. $K_{r,r,p}$ has a Hamilton cycle. For p as large as given, this can easily be shown to be $1 - o(n^{-1})$ if $r \geq r_0$. This is because the probability there is no Hamilton cycle in $K_{r,r,p}$ is of the same order as the probability that there is an isolated vertex. And this is at most $2r(1-p)^r \leq 2ne^{-r_0 n^{-\varepsilon}} = o(n^{-1})$. This verfies connectivity and the existence of a perfect matching. □

We let AP_r denote the problem of finding a minimum weight perfect matching between A_r and B_r. Let $M_r = \{(a_i, \phi(a_i)) : i = 1, 2, \ldots, r\}$ denote the optimal solution to AP_r.

For a set $S \subseteq A_r$ we let

$$N_0(S) = \left\{ b_j \in B_r : \exists a_i \in S \text{ such that } (a_i, b_j) \in R \text{ and } C(i,j) \leq \beta_r = \frac{\varepsilon \gamma_r}{10} \right\} \text{ where } \gamma_r = r^{-(1-2\varepsilon)}.$$

Lemma 5. *If $r \geq r_0$ then the following holds q.s.*[1]*,*

$$|N_0(S)| \geq \frac{\varepsilon r^\varepsilon |S|}{40} \text{ for all } S \subseteq A_r, 1 \leq |S| \leq r^{1-\varepsilon}. \tag{3}$$

Proof. For a fixed $S \subseteq A_r$, $s = |S| \geq 1$ we have that $|N_0(S)|$ is distributed as $Bin(r, q_s)$ in distribution, where $1 - q_s = (1 - n^{-\varepsilon} + n^{-\varepsilon} e^{-\beta_r})^s \leq (1 - \frac{1}{2} n^{-\varepsilon} \beta_r)^s$. It follows that $q_s \geq n^{-\varepsilon} \beta_r s/3$ for $s \leq r^{1-\varepsilon}$ and so $rq_s \geq \frac{\varepsilon r^{\varepsilon/2} s}{30}$. Let $\nu_s = \frac{\varepsilon r^{\varepsilon/2} s}{40}$. Then, using the Chernoff bounds, we have

$$\mathbb{P}(\neg(3)) \leq \sum_{s=1}^{r^{1-\varepsilon}} \binom{r}{s} \mathbb{P}(Bin(r, q_s) \leq \nu_s) \leq \sum_{s=1}^{r^{1-2\varepsilon}} \left(\frac{re}{s}\right)^s e^{-\Omega(\varepsilon r^{\varepsilon/2} s)} = \sum_{s=1}^{r^{1-2\varepsilon}} \left(\frac{re}{s} \cdot e^{-\Omega(\varepsilon r^{\varepsilon/2})}\right)^s = e^{-\Omega(\varepsilon r^{\varepsilon/2})}.$$

□

We let $\vec{G}_r$ denote a digraph with vertex set $A_r \cup B_r$. The edges of $\vec{G}_r$ consists of $\vec{M}_r = \{(b_{\phi(a_i)}, a_i) : i \in [r]\}$ directed from B_r to A_r and edges $\vec{D}_r = \{(a_i, b_j) : (i,j) \in D, j \neq \phi(i)\}$ directed from A_r to B_r. The edge $(b_{\phi(a_i)}, a_i)$ of $\vec{M}_r$ is given cost $-C(i, \phi(i))$ for $i \in [r]$. The edge (a_i, b_j) of $\vec{D}_r$ is given cost $C(i,j)$. Paths in $\vec{G}_r$ are *alternating* in the sense that ignoring orientation, they alternate between being in M_r and not being in M_r.

Lemma 6. *The following holds q.s.: assume that $r \geq r_0$. Then for all $a \in A_r, b \in B_r$, $\vec{G}_r$ contains an alternating path from a to b of total cost less than γ_r.*

Proof. Fix $a = a_1, b = b_1$ and let an alternating path $P = (a_1 = x_1, y_1, \ldots, y_{k-1}, x_k, y_k = b_1)$ be *acceptable* if $C(x_i, y_i) \leq \beta_r$, $i = 1, 2, \ldots, k$. Now consider the sequence of sets $S_0 = \{a_1\}, S_1, S_2, \ldots \subseteq A_r, T_1, T_2, \ldots \subseteq B_r$ defined as follows:

$T_i = N_0\left(\bigcup_{j<i} S_j\right)$ and $S_i = \phi^{-1}(T_i)$. It follows from (3) that w.h.p.

$$|S_i| = |T_i| \geq \left(\frac{\varepsilon r^\varepsilon}{40}\right)^i \text{ as long as } \left(\frac{\varepsilon r^\varepsilon}{40}\right)^i \leq r^{1-\varepsilon}.$$

[1]A sequence of events $\mathcal{A}_r$ holds *quite surely* (q.s.) if $\mathbb{P}(\neg \mathcal{A}_r) = o(r^{-K})$ as $r \to \infty$ for any constant $K > 0$

So define i_0 to be the smallest integer i such that $\left(\frac{\varepsilon r^{\varepsilon}}{40}\right)^i \geq r^{1-\varepsilon}$. Note that $i_0 < 2/\varepsilon$. Thus w.h.p. $|S_{i_0}| \geq r^{1-\varepsilon}$. Replace S_{i_0} by a subset of S_{i_0} of size $r^{1-\varepsilon}$ and then after this, we have that w.h.p. $|S_{i_0+1}| \geq \frac{\varepsilon r}{40}$.

For a set $T \subseteq B_r$ we let

$$\widehat{N}_0(T) = \{a_i \in A_r : \exists b_j \in T \text{ such that } (a_i, b_j) \in E(D) \text{ and } C(i,j) \leq \beta_r\}.$$

We then define $\widehat{T}_0 = \{b_1\}, \widehat{T}_1, \widehat{T}_2, \ldots, \widehat{T}_{i_0+1} \subseteq B, \widehat{S}_1, \widehat{S}_2, \ldots, \widehat{S}_{i_0+1} \subseteq A_r$ by $\widehat{S}_i = \widehat{N}_0\left(\bigcup_{j<i} \widehat{T}_j\right)$ and $\widehat{T}_i = \phi(\widehat{S}_i)$ and argue as above that $|\widehat{T}_{i_0+1}| \geq \frac{\varepsilon r}{40}$ with probability $1 - e^{-\Omega(\varepsilon r^{\varepsilon}/2)}$.

For $S \subseteq A_r, T \subseteq B_r$ let

$$E_R(S,T) = \{a_i \in S, b_j \in T : (i,j) \in R, C(i,j) \leq \beta_r\}.$$

Then,

$$\mathbb{P}\left(\exists S \subseteq A_r, T \subseteq B_r : |S|, |T| \geq \frac{\varepsilon r}{40}, E_R(S,T) = \emptyset\right) \leq 2^{2r} \exp\left\{-\frac{\varepsilon^2 r^2}{1600 r^{1-2\varepsilon}}\right\} = e^{-\Omega(r^{1+2\varepsilon})}.$$

It follows that with the claimed probability there will be an edge in $E_R(S_{i_0+1}, \widehat{T}_{i_0+1})$ and we have found an alternating path of length at most $2i_0 + 3$ using edges of cost at most β_r and this completes the proof of Lemma 6. □

Corollary 7. *No edge of M_r has cost more than γ_r, q.s.*

Proof. Suppose $e = \{a_i, b_j\} \in M_r$ and $C(e) > \gamma_r$. It follows from Lemma 6 that there is an alternating path $P = (a_1, \ldots, b_j)$ of cost at most γ_r. But then deleting e and the M_r-edges of P and adding the non-M_r edges of P to M_r creates a matching from A_r to B_r of lower cost than M_r, contradiction. □

4 Proof of Lemma 3

Here is an inital outline of the proof strategy.

1. We analyse the sequential shortest path algorithm for solving the assignment problem. By this, we mean that given M_r, we obtain M_{r+1} by solving a shortest path problem. A shortest path here corresponds to a least cost augmenting path that increases the matching cost by the minimum.

2. We make the edge lengths non-negative so that we can use Dijkstra's algorithm.

3. We estimate the number of short cycles created in this process. We bound the number of short cycles only, as there cannot be many vertex disjoint long cycles.

4.1 Linear programming formulation of AP

For the remainder of the proof, assume that $r \geq r_0$. We consider the linear program $\mathcal{LP}_r$ that underlies the assignment problem and its dual $\mathcal{D}_r$. We obtain M_{r+1} from M_r via a shortest augmenting path P_r and we examine the expected number of *short cycles* created by this path. A simple accounting then proves Lemma 3.

Let $\mathcal{LP}_r$ be the linear program

$$\text{Minimise } \sum_{i,j\in[r]} C(i,j)x_{i,j} \text{ subject to } \sum_{j\in[r]} x_{i,j}=1,\ i\in[r],\ \sum_{i\in[r]} x_{i,j}=1,\ j\in[r],\ x_{i,j}\geq 0.$$

The linear program $\mathcal{D}_r$ dual to $\mathcal{LP}_r$ is given by:

$$\text{Maximise } \sum_{i=1}^{r} u_i + \sum_{j=1}^{r} v_j \text{ subject to } u_i+v_j \leq C(i,j), i,j\in[r].$$

4.1.1 Trees and bases

An optimal basis of $\mathcal{LP}_r$ can be represented by a spanning tree T_r of G_r that contains the perfect matching M_r, see for example Ahuja, Magnanti and Orlin [1], Chapter 11. We have that for every optimal basis T_r,

$$C(i,j) = u_i + v_j \text{ for } (a_i,b_j)\in E(T_r) \tag{4}$$

and

$$C(i,j) \geq u_i + v_j \text{ for } (a_i,b_j)\notin E(T_r). \tag{5}$$

Note that replacing u_i by $\widehat{u}_i = u_i-\lambda, i=1,2,\ldots,r$ and v_i by $\widehat{v}_i = v_i+\lambda, i=1,2,\ldots,r$ for any λ has no afffect on these constraints. We say that $\mathbf{u},\mathbf{v}$ and $\widehat{\mathbf{u}},\widehat{\mathbf{v}}$ are equivalent. It follows that we can always fix the value of one component of $\mathbf{u}$,$\mathbf{v}$.

For a fixed tree T and $\mathbf{u}$,$\mathbf{v}$ let $\mathbf{C}(T,\mathbf{u},\mathbf{v})$ denote the set of cost matrices C such that the edges of T satisfy (4). The following lemma implies that the space of cost matrices (essentially) partitions into sets defined by $T,\mathbf{u},\mathbf{v}$. As such, we can prove Lemma 3 by showing that there are few cycles for almost all $\mathbf{u}$,$\mathbf{v}$ and spanning trees satisfying (4), (5).

Lemma 8. *(a) Given $\mathbf{u}$,$\mathbf{v}$, if T_1,T_2 are distinct spanning trees of G_r then $\mathbf{C}(T_1,\mathbf{u},\mathbf{v})\cap\mathbf{C}(T_2,\mathbf{u},\mathbf{v})$ has measure zero, given $\mathbf{u}$,$\mathbf{v}$.*
(b) If $u_1=u_1'=0$ and $(\mathbf{u},\mathbf{v})\neq(\mathbf{u}',\mathbf{v}')$ then for any spanning tree T of G_r, we have that $\mathbf{C}(T,\mathbf{u},\mathbf{v})\cap\mathbf{C}(T,\mathbf{u}',\mathbf{v}')=\emptyset$.

Proof. (a) Suppose that $C\in\mathbf{C}(T_1,\mathbf{u},\mathbf{v})\cap\mathbf{C}(T_2,\mathbf{u},\mathbf{v})$. We root T_1,T_2 at a_1 and let $u_1=0$. The equations (4) imply that for $i\in[r]$, u_i is the alternating sum and difference of costs on the path $P_{i,k}$ from a_1 to a_i in T_k. So, unless $P_{i,1}=P_{i,2}$ for all i, there will be an additional non-trivial linear combination of the $C(i,j)$ that equals zero. This has probability zero.
(b) There is a 1-1 correspondence between the costs of the tree edges and $\mathbf{u},\mathbf{v}$. □

The next goal is to show that w.h.p. we can choose optimal dual variables of absolute value at most $2\gamma_r = 2r^{-(1-2\varepsilon)}$. Let $\mathcal{E}$ be the event that we can choose $|u_i|,|v_j|\leq 2\gamma_r$ for all i,j.

Lemma 9. *$\mathcal{E}$ holds q.s.*

Proof. Let $\mathcal{E}_0$ be the event described in Lemma 6. Fix a_i,b_j and let $P=(a_{i_1},b_{j_1},\ldots,a_{i_k},b_{j_k})$ be the alternating path from a_i to b_j promised by $\mathcal{E}_0$. Then, using (4) and (5), we have

$$\gamma_r \geq C(P) = \sum_{l=1}^{k} C(i_l,j_l) - \sum_{l=1}^{k-1} C(i_{l+1},j_l) \geq \sum_{l=1}^{k}(u_{i_l}+v_{j_l}) - \sum_{l=1}^{k-1}(u_{i_{l+1}},v_{i_l}) = u_i+v_j. \tag{6}$$

Fix $u_s = 0$ for some s. For each $i \in [r]$ there is some $j \in [r]$ such that $u_i + v_j = C(i,j)$. This is because of the fact that a_i meets at least one edge of T and we assume that (4) holds. We also know that from (6) occurs that $u_{i'} + v_j \leq \gamma_r$ for all $i' \neq i$. It follows that $u_i - u_{i'} \geq C(i,j) - \gamma_r \geq -\gamma_r$ for all $i' \neq i$. Since i is arbitrary, we deduce that $|u_i - u_{i'}| \leq \gamma_r$ for all $i, i' \in [r]$. Since $u_s = 0$, this implies that $|u_i| \leq \gamma_r$ for $i \in [r]$. We deduce by a similar argument that $|v_j - v_{j'}| \leq \gamma_r$ for all $j, j' \in [r]$. Now because for the optimal matching edges $\left\{a_i, b_{\phi(i)}\right\}, i \in [r]$ we have $u_i + v_{\phi(i)} = C(i, \phi(i)) \leq \gamma_r$, we see that $|v_j| \leq 2\gamma_r$ for $j \in [r]$. □

Corollary 7 bounds the cost of the edges in M_r. The next lemma proves the same bound for the costs of the other edges in T_r.

Lemma 10. *$C(e) \leq \gamma_r$ q.s. for all $e \in T_r$.*

Proof. We condition on the values $\mathbf{u}$,$\mathbf{v}$. Suppose there is an edge $e = (a_i, b_j) \in T_r \setminus M_r$ such that $C(i,j) > \gamma_r$. Suppose we replace $C(e)$ by γ_r and resolve AP. Corollary 7 implies that the matching will be unchanged. The optimal tree might change to T_r'. Now the values $\mathbf{u}$,$\mathbf{v}$ are obtained from u_1 by taking the paths in T_r, T_r' and alternately adding and subtracting the edges in paths. Because the costs (and maybe the tree) has changed there will be a non-trivial sum of positive and negative costs that sum to zero or one that sums to γ_r. Both events have probability 0. □

The next two lemmas help us to understand the structure of the tree T_r. Fix M_r and let $G_r^* = G_r^*(\mathbf{u}, \mathbf{v})$ be the subgraph of G_r induced by the edges (a_i, b_j) for which $u_i + v_j \geq 0$. We first show that T_r is a uniform random spanning tree of G_r^*, containing M_r.

Let $\mathcal{T}_r(\mathbf{u}, \mathbf{v})$ denote the set of spanning trees of $G_r^*(\mathbf{u}, \mathbf{v})$ that contain the edges of M_r. This is non-empty because $T_r \in \mathcal{T}_r(\mathbf{u}, \mathbf{v})$. Let $f(\mathbf{u}, \mathbf{v})$ denote the joint density of $\mathbf{u}, \mathbf{v}$ then

Lemma 11. *If $T \in \mathcal{T}_r(\mathbf{u}, \mathbf{v})$ then*

$$\mathbb{P}(T_r = T \mid \mathbf{u}, \mathbf{v}) f(\mathbf{u}, \mathbf{v}) = \prod_{(a_i, b_j) \in G_r^*(\mathbf{u}, \mathbf{v})} e^{-(u_i + v_j)}, \tag{7}$$

which is independent of T.

Proof. Fixing $\mathbf{u}$,$\mathbf{v}$ and T_r fixes the lengths of the edges in T_r. If $(a_i, b_j) \notin E(T_r)$ then $\mathbb{P}(C(i,j) \geq u_i + v_j) = e^{-(u_i+v_j)^+}$ where $x^+ = \max\{x, 0\}$. If $(a_i, b_j) \in E(T_r)$ then the density of $C(i,j)$ is $e^{-(u_i+v_j)}$. So,

$$\mathbb{P}(T_r = T \mid \mathbf{u}, \mathbf{v}) f(\mathbf{u}, \mathbf{v}) = \prod_{(a_i, b_j) \notin E(T)} e^{-(u_i+v_j)^+} \prod_{(a_i, b_j) \in E(T)} e^{-(u_i+v_j)} = \prod_{(a_i, b_j) \in G_r^*(\mathbf{u}, \mathbf{v})} e^{-(u_i+v_j)}. \tag{8}$$

□

Thus

$$T_r \text{ is a uniform random member of } \mathcal{T}_r(\mathbf{u}, \mathbf{v}). \tag{9}$$

This property is key for the proof. It also explains why we need to modify costs when they are not $EXP(1)$. Equation (8) does not hold for more general distributions of cost.

We need the following simple graph theoretic lemma:

Lemma 12. *Let G be a graph with n vertices and with minimum degree d. Let T be a spanning tree of G maximum degree Δ. Let v be a vertex of T. Then T contains at least $\lfloor d/\Delta^3 \rfloor$ edge disjoint paths $P = (v_1, v_2, v_3, v_4)$ such that (i) $\{v, v_4\} \in E(G)$ and (ii) P constitutes the last 3 edges in the path from v to v_4. We refer to these paths as* useful *paths.*

Proof. First consider the case where $G = K_n$. We prove this by induction with $n = \Delta^3$ as the base case. If $n > \Delta^3$ we choose a leaf w at furthest distance from v. Let the path from v to w end with $x_0, x_1, \ldots, x_k = w$. Deleting the edges of the tree rooted at x_{k-3} after removing the edge $\{x_{k-4}, x_{k-3}\}$ yields a tree with at least $n - \Delta^3$ vertices and at least $n/\Delta^3 - 1$ paths of length 3. We obtain the general case by contracting the edges of T that contain vertices not adjacent to v. □

The next lemma will show that G_r^* has a large minimum degree. We use this to show that there are many non-basic variables with small reduced costs. It is not obvious at the moment, but this makes it unlikely that the shortest augmenting path with respect to M_r creates a short cycle.

We need to know that w.h.p. each vertex a_i is connected in G_r to many b_j for which $u_i + v_j \geq 0$. We fix a tree T and condition on $T_r = T$. For $i = 1, 2, \ldots, r$ let $L_{i,+} = \{j : (i,j) \in E(G)\}$ and let $L_{j,-} = \{i : (i,j) \in E(G)\}$. Then for $i = 1, 2, \ldots, r$ let $\mathcal{A}_{i,+}$ be the event that $|\{j \in L_{i,+} : u_i + v_j \geq 0\}| \leq r^{1-7\varepsilon}$ and let $\mathcal{A}_{j,-} = \mathcal{A}_{j,-}(\eta)$ be the event that $|\{i \in L_{j,-} : u_i + v_j \geq 0\}| \leq r^{1-7\varepsilon}$.

Lemma 13. *Fix a spanning tree T of G_r^* that contains M_r. Let $\mathcal{E}$ be as in Lemma 9. Then*

$$\mathbb{P}\left((\mathcal{A}_{i,+}(\eta) \vee \mathcal{A}_{i,-}(\eta)) \wedge \mathcal{E} \mid T_r = T\right) = O(n^{-anyconstant}).$$

Proof. We assume that $C(i,j) \leq \gamma_r$ for $(a_i, b_j) \in T$. The justification for this is Lemma 10. The number of edges in G_r of cost at most γ_r incident with a fixed vertex is dominated by $Bin(r, \gamma_r)$ and so with probability $1 - e^{-\Omega(r^{2\varepsilon})}$ the maximum degree in G_r can be bounded $2r^{2\varepsilon}$. This degree bound applies to the trees we consider.

We fix s and put $u_s = 0$. The remaining values $u_i, i \neq s, v_j$ are then determined by the costs of the edges of the tree T. Let $\mathcal{B} = \mathcal{B}(T)$ be the event that $C(i,j) > u_i + v_j$ for all $(a_i, b_j) \notin E(T)$. Note that if (4) holds and $\mathcal{B}$ occurs then $T_r = T$, and vice-versa.

We now condition on the set E_T of edges (and the associated costs) of $\{(a_i, b_j) \notin E(T)\}$ such that $C(i,j) \geq 4\gamma_r$. Let $F_T = \{(a_i, b_j) \notin E(T)\} \setminus E_T$. Note that $|F_T|$ is dominated by $Bin(r^2, 1 - e^{-2\gamma_r})$ and so $|F_T| \leq 5r^2\gamma_r$ with probability $1 - e^{-\Omega(r^{2\varepsilon})}$.

Let $Y = \{C(i,j) : (a_i, b_j) \in E(T)\}$ and let $\delta_1(Y)$ be the indicator for $\mathcal{A}_{s,+} \wedge \mathcal{E}$. We write,

$$\mathbb{P}(\mathcal{A}_{s,+} \wedge \mathcal{E} \mid \mathcal{B}) = \frac{\int \delta_1(Y) \mathbb{P}(\mathcal{B} \mid Y) dC}{\int \mathbb{P}(\mathcal{B} \mid Y) dC} \tag{10}$$

Then we note that since $(a_i, b_j) \notin F_T \cup E(T)$ satisfies the condition (5),

$$\mathbb{P}(\mathcal{B} \mid Y) = \prod_{(a_i, b_j) \in F_T} \exp\left\{-(u_i(Y) + v_j(Y))^+\right\} = e^{-W}, \tag{11}$$

where $W = W(Y) = \sum_{(a_i,b_j)\in F_T}(u_i(Y)+v_j(Y))^+ \leq 20r^2\gamma_r^2 = 20r^{4\varepsilon}$. Then we have

$$\begin{aligned}\int_Y \delta_1(Y)\mathbb{P}(\mathcal{B}\mid Y)\,dC &= \int_Y e^{-W}\delta_1(Y)\,dC \\ &\leq \left(\int_Y e^{-2W}\,dC\right)^{1/2}\times\left(\int_Y \delta_1(Y)^2\,dC\right)^{1/2} \\ &\leq e^{-\mathbb{E}(W)}\left(\int_Y e^{-2(W-\mathbb{E}(W))}dC\right)^{1/2}\times\mathbb{P}(\mathcal{A}_{s,+}\mid\mathcal{E})^{1/2} \\ &\leq e^{-\mathbb{E}(W)}e^{20r^{4\varepsilon}}\mathbb{P}(\mathcal{A}_{s,+}\mid\mathcal{E})^{1/2}. \qquad (12)\end{aligned}$$

$$\int \mathbb{P}(\mathcal{B}\mid Y)dC = \mathbb{E}(e^{-W}) \geq e^{-\mathbb{E}(W)}. \qquad (13)$$

It then follows from (10),(12) and (13) that

$$\mathbb{P}(\mathcal{A}_{s,+}\wedge\mathcal{E}\mid\mathcal{B}) \leq e^{20r^{4\varepsilon}}\mathbb{P}(\mathcal{A}_{s,+}\mid\mathcal{E})^{1/2}. \qquad (14)$$

Equation (14) removes the conditioning on $\mathcal{B}$ at the expense of inflating our probability estimate by something tolerable. For the remainder of the lemma we assume that the $C(i,j)$ for $(a_i,b_j)\in T$ satisfy $C\leq\gamma_r$ and that $\mathcal{E}$ holds. Denote this conditioning by $\mathcal{F}$. Let b_j be a neighbor of a_s in G_r^* and let $P_j = (i_1 = s, j_1, i_2, j_2, \ldots, i_k, j_k = j)$ define the path from a_s to b_j in T. Then it follows from (4) that $v_{j_l} = v_{j_{l-1}} - C(i_l, j_{l-1}) + C(i_l, j_l))$. Thus v_j is the final value S_k of a random walk $S_t = X_0 + X_1 + \cdots + X_t, t = 0,1,\ldots,k$, where $X_0 \geq 0$ and each $X_t, t\geq 1$ is the difference between two independent copies of $EXP(1)$ that are conditionally bounded above by γ_r. Assume for the moment that $k\geq 3$ and let $x = u_{i_{k-3}} \in [-2\gamma_r, 2\gamma_r]$. Given x we see that there is some positive probability $p_0 = p_0(x) = \mathbb{P}(x + X_{k-2} + X_{k-1} + C(i_k, j_k) > 0 \mid \mathcal{F})$. Let $\eta = \min\{x\geq -2\gamma_r : p_0(x)\} > 0$.

The number of edges in G_+ of cost at most γ_r incident with a fixed vertex is dominated by $Bin(n,\gamma_r)$ and so w.h.p. the maximum degree of the trees we consider can be bounded by $2r^{2\varepsilon}$. Applying Lemma 12 we see that given $\mathcal{E}$ we can find a set K of at least $\alpha r^{1-6\varepsilon}/8$ values $k\in L_{s,+}$ such that $v_k, k\in K$ forms the set of endpoints of disjoint useful paths. Then we have that given $\mathcal{F}$, independently of each other we have $v_k > 0, k\in K$ with probability at least $\eta > 0$. Applying (14) and the Chernoff bounds we have

$$\mathbb{P}(\mathcal{A}_{s,+}\wedge\mathcal{E}\mid\mathcal{B}) \leq e^{20r^{4\varepsilon}}\mathbb{P}(Bin(\alpha r^{1-6\varepsilon}/8,\eta)\leq r^{1-7\varepsilon}) \leq e^{20r^{4\varepsilon}-\Omega(r^{1-6\varepsilon})}.$$

Taking the union bound over choices of $s\leq r$ and $+,-$ proves the lemma. □

So, from now on we assume that $\mathcal{E}$ occurs and that $\mathcal{A}_{s,\pm}$ does not occur for any $s = 1,2,\ldots,r$.

4.2 Construction of the augmenting path

As previously mentioned, we will go from M_r to M_{r+1} by solving a shortest path problem. We let $\vec{G}_r$ be the orientation of G_{r+1} with edges oriented from A_{r+1} to B_{r+1} **except** for the edges of M_r which are oriented from B_r to A_r. The *forward* edges $(a_i,b_j)\notin M_r$ are given their costs $C(i,j)$. The *backward* edges in $(a_i,b_j)\in M_r$ are given costs $-C(i,j)$. This reflects the idea that traversing a forward edge means adding it and traversing a backward edge means deleting it from the matching. We obtain M_{r+1} from M_r by finding a minimum cost (augmenting) path $P_r = (x_1 = a_{r+1}, y_1, x_2, \ldots, x_\sigma, y_\sigma = b_{r+1})$ from a_{r+1} to b_{r+1} in $\vec{G}_r$. As defined so far, the backward edges have a negative cost. In order to use Dijkstra's algorithm, we must modify the costs so that they become non-negative.

We let

$$u_{r+1} = \min\{C(r+1,j) - v_j(T_r) : j \in [r]\} \text{ and}$$
$$v_{r+1} = \min\{C(r+1,r+1) - u_{r+1}, \min\{C(i,r+1) - u_i(T_r) : i \in [r]\}\}. \quad (15)$$

We use costs $\widehat{C}(i,j) = C(i,j) - u_i - v_j$ in our search for a shortest augmenting path. Our choice of u_{r+1}, v_{r+1} and (4), (5) implies that $\widehat{C}(i,j) \geq 0$ and that matching edges have cost zero. This idea for making edge costs non-negative is well known, see for example Kleinberg and Tardos [29]. The $\widehat{C}$ cost of a path P from a_{r+1} to $b_{r+1} \in B$ differs from its C cost by $-(u_{r+1} + v_{r+1})$, independent of P.

We now introduce some conditioning $\mathcal{C}$. We fix $M_r = \{(a_i, b_{\phi(i)}), i = 1, 2, \ldots, r\}$ and assume that $\mathbf{u}, \mathbf{v} \in \mathcal{U} = \{u_i, v_i : |u_i|, |v_i| \leq 2\gamma_r\}$ and that for all i, neither $\mathcal{A}_{i,+}$ nor $\mathcal{A}_{i,-}$ of Lemma 13 hold. The constraints (4), (5) on the $C(i,j)$ become that

$$\begin{aligned} C(i,\phi(i)) &= u_i + v_{\phi(i)} \text{ for } i = 1, 2, \ldots, r \\ C(i,j) &\geq u_i + v_j, \text{ otherwise.} \end{aligned} \quad (16)$$

Note that with this conditioning, the tree T_r of basic variables is not completely determined. The tree T_r will not be exposed all at once, but we will expose it as necessary. We also define some extra conditioning $\mathcal{C}+$ that will only be needed in Section 4.3.2, when we deal with non-basic edges. Not only will we fix M_r, but we will also fix T_r and $\mathbf{u}, \mathbf{v} \in \mathcal{U}$.

4.3 Dijkstra's algorithm

We let $\Gamma_r^* = \Gamma_r^*(\mathbf{u}, \mathbf{v})$ denote the (multi)graph obtained from G_r^* by contracting the edges of M_r and let $\widehat{T}_r$ be the tree obtained from T_r by contracting these edges. We have to consider multigraphs because we may find that $(a_i, \phi(a_j))$ and $(a_j, \phi(a_i))$ are both edges of $G_r^*(\mathbf{u}, \mathbf{v})$. Of course, T_r can only contain at most one of such a pair. It follows from (9) that given $\mathbf{u}$,$\mathbf{v}$, $\widehat{T}_r$ is a uniform random spanning tree of Γ_r^*.

We use Dijkstra's algorithm to find the shortest augmenting path from a_{r+1} to b_{r+1} in the digraph $\vec{G}_r$. Because each $b_j \in B_r$ has a unique out-neighbor $a_{\phi^{-1}(j)}$ and $\widehat{C}(b_j, a_{\phi^{-1}(j)}) = 0$, we can think of the Dijkstra algorithm as operating on a digraph $\vec{\Gamma}_r$ with vertex set A_{r+1}. The edges of $\vec{\Gamma}_r$ are derived from paths $(a_i, \phi(a_j), a_j)$ in $\vec{G}_r$. (We are just contracting the edges of M_r.) The cost of this edge will be $\widehat{C}(i,j)$ which is the cost of the path $(a_i, \phi(a_j), a_j)$ in $\vec{G}_r$. Given an alternating path $P = (a_{i_1}, b_{j_1}, a_{i_2}, \ldots a_{i_k})$ where $\phi(a_{i_t}) = b_{j_t}$ for $t \geq 2$ there is a corresponding $\psi(P) = (a_{i_1}, a_{i,2}, \ldots, a_{i_k})$ of the same length in $\vec{G}_r$.

The Dijkstra algorithm applied to $\vec{G}_r$ produces a sequence of values $0 = d_1 \leq d_2 \leq \cdots \leq d_{r+1}$. The d_i are the costs of shortest paths. Suppose that after k rounds we have a set of vertices S_k for which we have found a shortest path of length d_i to $a_i \in S_k$ and that d_l for $a_l \notin S_k$ is our current estimate for the cost of a shortest path from a_{r+1} to a_l. The algorithm chooses $a_{l^*} \notin S_k$ to add to S_k to create S_{k+1}. Here l^* minimises $d_i + \widehat{C}(i,l)$ over $a_i \in S_k, a_l \notin S_k$. It then updates the $d_l, a_l \notin S_{k+1}$ appropriately. In this way, the Dijkstra algorithm builds up a tree DT_k that is made up of the known shortest paths after k rounds. Here $DT_1 = a_{r+1}$.

Let $\theta_{i,\ell} = d_k - d_i + u_i - u_\ell + C(\ell, \phi(\ell))$. Note that if $i \leq k < \ell$ then $0 \leq d_i + \widehat{C}(i,\ell) - d_k = C(i,\ell) - \theta_{i,\ell}$. Having fixed $\mathbf{u}, \mathbf{v}$ and T_r the only restriction on $C(i,\ell)$ for (i,ℓ) non-basic is that $C(i,\ell) \geq \theta_{i,\ell}$. This holds regardless of the other non-basic costs $C(p,q)$, $(p,q) \neq (i,\ell)$. The memoryless property of the exponential distribution then implies that under the conditioning $\mathcal{C}+$ (defined at the end of Section 4.2), the non-basic/non-tree values

$C(i,\ell)$ are independently distributed as follows:

$$\begin{array}{l}\text{If } \theta_{i,\ell} \geq 0 \text{ then } d_i + \widehat{C}(i,\ell) - d_k \text{ is distributed as } EXP(1).\\ \text{Otherwise, } d_i + \widehat{C}(i,\ell) - d_k \text{ is distributed as } -\theta_{i,\ell} + EXP(1) \leq u_\ell - u_i + EXP(1).\end{array} \tag{17}$$

4.3.1 Final argument

Referring to the augmenting path $P_r = (x_1 = a_{r+1}, y_1, x_2, \ldots, x_\sigma, y_\sigma = b_{r+1})$, suppose that $1 \leq \tau < \sigma$ and that $\widehat{M_{r,\tau}}$ is the matching obtained from M_r by adding the edges $(x_k, y_k), k = 1, 2, \ldots, \tau$ and deleting the edges $(x_{k+1}, y_k), k = 1, 2, \ldots, \tau - 1$. Suppose now that $x_\tau = a_i$ and $y_\tau = b_j$. Observe that vertex i is the head of a path, Q say, in the set of paths and cycles $\mathcal{A}_{\widehat{M_{r,\tau}}}$. ($Q$ is directed towards i.) We say that vertex x_τ *creates* a short cycle if j lies on Q and the sub-path of Q from j to i has length at most $\ell_1 = n^{4\varepsilon}$. In this case we also say that the edge (i, j) creates a short cycle. Extending the notation, we say that x_σ *creates* a short cycle if $r + 1$ ($y_\sigma = b_{r+1}$) is the tail of Q and the length of Q is at most ℓ_1. For $r \geq r_0$ we only count the creation of a small cycle by an edge (x, y) if this is the first such edge involving x. (In this way we avoid an overcount of the number of short cycles.) Call this a *virgin short cycle*. Let χ_r denote the number of virgin short cycles created in iteration r. We then have that

$$\mathbb{E}(\nu_C) \leq \frac{r_0}{2} + \frac{n}{\ell_1} + \sum_{r=r_0}^{n} \mathbb{E}(\chi_r). \tag{18}$$

Here n/ℓ_1 bounds the number of large cycles induced by M_n, i.e. those of length greater than ℓ_1. The $r_0/2$ term bounds the contributions from the matching M_{r_0}. The sum bounds the expected number of small cycles induced by M_n. To see this, suppose that C is a non-virgin short cycle and that it was created by adding the edge (x, y). There must have been some earlier virgin short cycle created by adding an edge (x, z) and this will be counted in the sum.

We claim that

$$\Sigma_C := \sum_{r=r_0}^{n} \mathbb{E}(\chi_r) \leq \ell_1 n^{1-11\varepsilon}. \tag{19}$$

Assume (19) for the moment. Then we have,

$$\mathbb{E}(\nu_C) \leq \frac{r_0}{2} + \frac{n}{\ell_1} + \ell_1 n^{1-11\varepsilon} \leq r_0. \tag{20}$$

Lemma 3 now follows from the Markov inequality, so it only remains to prove (19).

4.3.2 Proof of (19)

We fix $r \geq r_0$.

Edges incident with a_{r+1} or b_{r+1} The costs of edges incident with one of a_{r+1}, b_{r+1} are unconditioned at the start of the search for P_r. They have not been part of the optimization so far. Let ξ_i be the minimum C-cost of an alternating path from $b_i, i \leq r$ to b_{r+1} through G_r. It follows from Lemma 6 that w.h.p. $\xi_j \leq \gamma_r$ for all $j \leq r$. To create the shortest augmenting path from a_{r+1} to b_{r+1} we must find the minimum μ^* of the $C(a_{r+1,j}) + \xi_j$. There are at least $\alpha_0 r$ indices j for which the edge (a_{r+1}, b_j) exists in G_r, see Lemma 4.

It follows that w.h.p. $\mu^* \leq \min_j C(a_{r+1}, b_j) + \gamma_r \leq 2\gamma_r$. There are at most ℓ_1 indices j that would lead to the creation of a short cycle and for these the probability that $C(a_{r+1,j}) + \xi_j \leq 2\gamma_r$ is at most $2\gamma_r$. Thus in expectation, edges incident with a_{r+1} in this context, only contribute $O(\ell_1\gamma_r)$ to the number of short cycles over all. The same argument can be applied for edges incident with b_{r+1}.

Basic Edges Consider the point where we have carried out k iterations of the Dijkstra algorithm and we are about to add a $(k+1)$st vertex to the tree of known shortest paths. A path $(a, \phi(a'), a')$ in the tree T_r gives rise to a *basic* edge (a, a'). Basic edges have $\widehat{C}$ value zero and so if there are basic edges oriented from DT_k to $A_{r+1} \setminus DT_k$ then one of them will be added to the shortest path tree and we will have $d_{k+1} = d_k$. We need to argue that they are unlikely to create short cycles. At this point we will only have exposed basic edges that are part of DT_k.

Fix $a_i \in V(DT_k)$. We want to show that given the history of the algorithm, the probability of creating a short cycle via an edge incident with a_i is sufficiently small. At the time a_i is added to DT_r there will be a set L_1 of size at most ℓ_1 for which adding the edge corresponding to $(a_i, b_j, a_{\phi^{-1}(j)}), a_j \in L_1$ creates a short cycle. This set is not increased by the future execution of the algorithm. At this point we have only exposed edges of $\vec{G}_r$ pointing into a_i.

Let $e = (a_i, x)$, $x \in A_r$. We claim that

$$\mathbb{P}((a_i, x) \in \widehat{T}_r) = O\left(\frac{1}{r^{1-7\varepsilon}}\right) \tag{21}$$

from which we can deduce that

$$\mathbb{P}(\text{an added basic edge is bad}) = O\left(\frac{\ell_1}{r^{1-7\varepsilon}}\right), \tag{22}$$

where *bad* means that the edge creates a short cycle.

To prove (21) we use two well known facts: (i) if $e = \{a, b\}$ is an edge of a connected (multi)graph G and T denotes a uniform random spanning tree then $\mathbb{P}(e \in T) = R_{eff}(a, b)$ where R_{eff} denotes effective resistance, see for example [34]; (ii) $R_{eff}(a, b) = \frac{\tau(a,b)+\tau(b,a)}{2|E(G)|}$ where $\tau(x, y)$ is the expected time for a random walk starting at x to reach y, see for example [15]. We note that in the context of (21), we may have exposed some edges of T_r. Fortunately, edge inclusion in a random spanning tree is negatively correlated i.e. $\mathbb{P}(e \in T_r \mid f_1, \ldots, f_s \in T_r) \leq \mathbb{P}(e \in T_r)$, see for example [34].

Given (i) and (ii) and Lemma 11 it only remains to show that with $G = \Gamma_r^* = \Gamma_r^*(\mathbf{u}, \mathbf{v})$ that $\tau(a, x) = O(r^{1+7\varepsilon})$, for $a, x \in A_r$. For this we only have to show that the mixing time for a random walk on Γ_r is sufficiently small. After this we can use the fact that the expected time to visit a vertex a from stationarity is $1/\pi_a \leq r^{1+7\varepsilon}$ where π denotes the stationary distribution, see for example [33]. We estimate the mixing time of a walk by its *conductance*.

Let $\deg(v) \geq r^{1-7\varepsilon}$ denote degree in Γ_r^*. For $S \subseteq A_r$, let $\Phi_S = e(S, \bar{S})/\deg(S)$ where $e(S, \bar{S})$ is the number of edges of $\Gamma_r^*(\mathbf{u}, \mathbf{v})$ with one end in S and $\deg(S) = \sum_{v \in S} \deg(v)$. Let $\Phi = \min\{\Phi_S : \deg(S) \leq \deg(A_r)/2\}$. Note that if $\deg(S) \leq \deg(A_r)/2$ then $\deg(\bar{S}) \geq \deg(A_r)/2 \geq r^{2-7\varepsilon}/2$ which implies that $|\bar{S}| \geq r^{1-7\varepsilon}/2$ and so $|S| \leq r(1 - r^{-7\varepsilon})$.

Assume first that $|S| \leq \eta r^{1-7\varepsilon}/2$. Then

$$\Phi_S \geq \frac{\sum_{v \in S}(\deg(v) - |S|)^+}{\deg(S)} \geq \frac{r^{1-7\varepsilon}|S|/2}{r|S|} = \frac{r^{-7\varepsilon}}{2}.$$

If $\eta r^{1-7\varepsilon}/2 \leq |S| \leq r - \eta r^{1-7\varepsilon}/2$ then we use the random edges R. We sum over the $2^{O(r)}$ choices for S and the $r^{O(r)}$ choices for $\widehat{T}_r$. Then we see via the Chernoff bounds that with probability $1 - e^{-\Omega(r^{2-15\varepsilon})}$ there are at least $r^{2-15\varepsilon}$ edges in R from S to $\bar{S}$. The failure probability $e^{-\Omega(r^{2-15\varepsilon})}$ is small enough to handle the $r^{O(r)}$ choices of S, T_r. So,

$$\Phi_S \geq \frac{r^{2-15\varepsilon}}{r^2} = r^{-15\varepsilon}. \tag{23}$$

It then follows that after r steps of the random walk the total variation distance between the walk and the steady state is at most $r^2(1-\Phi_S^2/2)^r = e^{-\Omega(r^{1-30\varepsilon})}$, see for example [33]. This completes our verification of (21) and hence (22).

We will also need a bound on the number of basic edges in any path in the tree DT_r constructed by Dijkstra's algorithm. Aldous [2], Chung, Horn and Lu [11] discuss the diameter of random spanning trees. Section 6 of [2] provides an upper bound for the diameter that we use for the following.

Lemma 14. *The diameter of $\widehat{T}_r$ is $O(r^{1/2+30\varepsilon})$ with probability $1-o(r^{-2})$.*

Proof. Let A be the adjacency matrix of Γ_r^* and let D be the diagonal matrix of degrees $\deg(v), v \in A_r$ and let $L = I - D^{-1/2}AD^{-1/2}$ be the Laplacian. Let $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{r-1}$ be the eigenvalues of L and let $\sigma = 1 - \lambda_1$. We have $\lambda_1 \geq \Phi^2/2$ (see for example Jerrum and Sinclair [26]). So we have

$$\sigma \leq 1 - \frac{1}{2r^{30\varepsilon}}. \tag{24}$$

Now let $\rho_0 = r^{1/2}$ and δ denote the minimum degree in Γ_r^* and

$$s = \left\lceil \frac{3}{\log(1/\sigma)} \cdot \frac{r^2}{(\rho_0+1)\delta} \right\rceil = O(r^{1/2+30\varepsilon}).$$

It is shown in [11] that

$$\mathbb{P}(diam(T) \geq 2(\rho_0 + js)) \leq \frac{r}{2^{j-2}}. \tag{25}$$

Putting $j = 5\log r$ into (25) yields the lemma.

(Unfortunately, there are no equation references for (25). It appears in Section 6 of [2] and Section 5 of [11]. In [11], $\sigma = \max\{1-\lambda_1, \lambda_{n-1} - 1\}$. It is used to bound the mixing time of a lazy random walk on Γ_r^* and in our context we can drop the λ_{n-1} term.) □

Non-Basic Edges Each $a_i \in DT_k$ corresponds to an alternating path P_i. As such there are at most ℓ_1 choices of ℓ such that (i,ℓ) would create a bad edge. This is true throughout an execution of the Dijkstra algorithm. Also, while we initially only know that the $C(i,\ell), \ell \neq \phi(i)$ are $EXP(1)$ subject to (16), as Dijkstra's algorithm progresses, we learn lower bounds on $C(i,\ell)$ through (17). For this part of the argument we condition as for $\mathcal{C}+$. Note that assuming $\mathbf{u}, \mathbf{v} \in \mathcal{U}$ only affects the costs of basic edges. The reduced costs $C(i,\ell)$ will thus be (conditionally) independent.

We have to show that w.h.p. there are many non-basic pairs (i,ℓ) "competing" to be the next edge added to DT_k. This makes the choice of a bad edge unlikely. Examining (17) we see that for there to be any chance that an edge (i,ℓ) has low cost, it must be that $u_\ell - u_i$ must be at least some small negative value. The following shows that in most cases there will be sufficiently many $a_\ell \notin DT_k$ for which this is true.

Suppose that vertices are added to DT_r in the sequence $\mathbf{i} = i_1, i_2, \ldots, i_r$ and that $u_{j_1} < u_{j_2} < \cdots < u_{j_r}$ and that $\mathbf{j} = j_1, j_2, \ldots, j_r$. For $r_0 < j \leq r$ let

$$F(\mathbf{i}, \mathbf{u}, j) = |\Phi(\mathbf{i}, \mathbf{u}, j)| \text{ where } \Phi(\mathbf{i}, \mathbf{u}, j) = \left\{t > j : \ u_{i_t} \leq u_{i_j} + \gamma_r \varepsilon_r\right\} \text{ where } \varepsilon_r = r^{-100\varepsilon}.$$

Let $X_r(\mathbf{i}, \mathbf{u}) = \{j \leq r : \ F(\mathbf{i}, \mathbf{u}, j) \leq r\varepsilon_r^2\}$. If $j \notin X_r(\mathbf{i}, \mathbf{u})$ let $\widehat{\Phi}(\mathbf{i}, \mathbf{u}, j)$ be the first $r\varepsilon_r^2$ members of $\Phi(\mathbf{i}, \mathbf{u}, j)$ in the sequence $\mathbf{i}$. Note that given $\mathbf{u}$, $X_r(\mathbf{i}, \mathbf{u})$ is determined by $\mathbf{i}$,$\mathbf{j}$.

Lemma 15. $|X_r(\mathbf{i}, \mathbf{u})| \leq 4r\varepsilon_r$.

Proof. Assume without loss that $i_t = t$ and replace the notation $\Phi(\mathbf{i}, \mathbf{u}, j)$ by $\Phi(\mathbf{u}, j)$. We show that we can assume that $u_1 \leq u_2 \leq \cdots \leq u_r$. Assume that $u_k = \max\{u_1, \ldots, u_r\}$ and that $k < r$. Consider amending $\mathbf{u}$ by interchanging u_k and u_r. Fix $j < r$. We enumerate the possibilities and show that $F(\mathbf{u}, j)$ does not increase.

If $j \geq k$ then we have that $k \notin \Phi(\mathbf{u}, j)$ and $\Phi(\mathbf{u}, j)$ may lose element r, since u_r has increased. Assume then that $j < k$.

Before	$k \notin \Phi(\mathbf{u}, j), r \notin \Phi(\mathbf{u}, j)$	After	No change.
Before	$k \notin \Phi(\mathbf{u}, j), r \in \Phi(\mathbf{u}, j)$	After	$k \in \Phi(\mathbf{u}, j), r \notin \Phi(\mathbf{u}, j)$.
Before	$k \in \Phi(\mathbf{u}, j), r \notin \Phi(\mathbf{u}, j)$		Not possible.
Before	$k \in \Phi(\mathbf{u}, j), r \in \Phi(\mathbf{u}, j)$	After	No change.

So in all cases $F(\mathbf{u}, j)$ does not increase. u_r is now the maximum of the u_i. After this we can assume that $u_r = \max\{u_1, \ldots, u_r\}$. We now apply the argument above but restricted to $u_1, \ldots, u_{r-1}$ or use induction on r.

Next let k_1 be the smallest index k in $X_r(\mathbf{i}, \mathbf{u})$ and let $J_1 = [u_{k_1}, u_{k_1} + \gamma_r \varepsilon_r]$. The interval J_1 contains at most $r\varepsilon_r^2$ of the values u_i. Then let k_2 be the smallest index k in $X_r(\mathbf{i}, \mathbf{u})$ with $u_k > u_{k_1} + \gamma_r \varepsilon_r$ and let $J_2 = [u_{k_2}, u_{k_2} + \gamma_r \varepsilon_r]$ and so on. Using the fact that $\mathbf{u} \in \mathcal{U}$ we see that in this way we cover $X_r(\mathbf{i}, \mathbf{u})$ with at most $4\varepsilon_r^{-1}$ intervals each containing at most $r\varepsilon_r^2$ of the values u_j for which $j \in X_r(\mathbf{i}, \mathbf{u})$. $\square$

Now let

$$K_r = \left\{k : |X_r(\mathbf{i}, \mathbf{u}) \cap [k - r\varepsilon_r^2/2, k]| \geq r\varepsilon_r^2/4\right\}.$$

And for each $k \notin K_r$, let Ψ_k be the first $r\varepsilon_r^2/4$ members of $[k - r\varepsilon_r^2/2, k] \setminus X_r(\mathbf{i}, \mathbf{u})$ in the sequence $\mathbf{i}$. Note that given $\mathbf{u}$, Ψ_k is determined by $\mathbf{i}$,$\mathbf{j}$.

Lemma 16. $|K_r| \leq 2|X_r(\mathbf{i}, \mathbf{u})| \leq 8r\varepsilon_r$.

Proof. Let $z_{j,k}$ be the indicator for (j, k) satisfying $k - r\varepsilon_r^2/2 \leq j \leq k$ and $j \in X_r(\mathbf{i}, \mathbf{u})$. Then if $z = \sum_{j,k} z_{j,k}$ we have

$$z \geq \sum_{k \in K_r} r\varepsilon_r^2/4 = |K_r| r\varepsilon_r^2/4.$$

$$z \leq \sum_{j \in X_r(\mathbf{i}, \mathbf{u})} r\varepsilon_r^2/2 \leq r\varepsilon_r^2 |X_r(\mathbf{i}, \mathbf{u})|/2.$$

and the lemma follows from Lemma 15. $\square$

It follows from the definition of K_r that if $k \notin K_r$ then there are at least $r\varepsilon_r^2/4 \times r\varepsilon_r^2$ pairs $(j \in \Psi_k, \ell \in \widehat{\Phi}(\mathbf{i}, j))$ such that $j \leq k < \ell$ and $u_\ell \leq u_j + \gamma_r \varepsilon_r$. Note that $\theta_{j,\ell} \geq -\varepsilon_r \gamma_r$ for each such pair. We next estimate for $k \notin K_r$ and $r_0 \leq k \leq r$ and $j \leq k < \ell \leq r$ the probability that (j, ℓ) minimises $d_i + \widehat{C}(i, \ell)$. The Chernoff bounds imply that w.h.p. $r^2 \varepsilon_r^4 n^{-\varepsilon}/5 \gg r \log r$ of these pairs appear as edges in the random edge set R. (We can afford to multiply by $r!^2$ so that this claim holds for all choices of $\mathbf{i}$,$\mathbf{j}$ and hence of Ψ_k, $\widehat{\Phi}(\mathbf{i}, j), j \notin X_r(\mathbf{i}, \mathbf{u})$.) Given this, it follows from the final inequality in (17) that

$$\mathbb{P}(\text{an added non-basic edge is bad} \mid \mathcal{C}+) \leq \ell_1 \left(\varepsilon_r \gamma_r + \frac{5n^\varepsilon}{r^2 \varepsilon_r^4} \right) \leq 2\ell_1 \varepsilon_r \gamma_r. \tag{26}$$

Explanation: There are at most ℓ_1 possibilities for a bad edge $e = (a_j, a_m)$ being added. The term $\varepsilon_r \gamma_r$ bounds the probability that the cost of edge e is less than $\varepsilon_r \gamma_r$. Failing this, e will have to compete with at least $r^2 \varepsilon_r^4 n^{-\varepsilon}/5$ other pairs for the minimum.

We will now put a bound on the length L of a sequence $(t_k, x_k), k = 1, 2, \ldots, L$ where $t_k, k \notin K_r$ is an iteration index where a non-basic edge (y_k, x_k) is added to DT_r. The expected number of such sequences can be bounded by

$$\sum_{\substack{t_1 < t_2 < \cdots < t_L \\ x_1, x_2, \ldots, x_L}} (2\varepsilon_r \gamma_r)^L \leq \binom{r}{L}^2 (2\varepsilon_r \gamma_r)^L \leq \left(\frac{2r^2 e^2 \varepsilon_r \gamma_r}{L^2} \right)^L = o(n^{-2}), \tag{27}$$

if $L^2 \geq 3e^2 \varepsilon_r \gamma_r r^2$ or $L \geq 3er^{1/2 - 51\varepsilon}$.

Explanation: We condition on the tails y_k of the edges added at the given times. Then there are at most r possibilities for the head x_k and then $2\varepsilon_r \gamma_r$ bounds the probability that (y_k, x_k) is added, see (26).

Combining Lemma 14 and (27) we obtain a bound of $r^{1-20\varepsilon}$ on the diameter of DT_r. (Each path in DT_r consists of a sequence of non-basic edges separated by paths of $\widehat{T}_r$ and so we multiply the two bounds.)

4.3.3 Finishing the proof of (19)

Let $\zeta_{r,k}$ be the 0,1 indicator for e_k being a virgin bad edge i.e. one that creates a virgin short cycle. Note that $\sum_{r=r_0}^n \sum_{k=1}^r \zeta_{r,k} \leq n$. We remind the reader that the following inequalities are claimed to be true for sufficiently small $\varepsilon > 0$,

We write

$$\Sigma_C \leq \sum_{r=r_0}^n \sum_{k=1}^r \mathbb{P}(e_k \text{ is bad} \mid \mathcal{C}) \zeta_{r,k} + o(1),$$

given that $\mathbb{P}(\mathcal{C}) = 1 - o(1)$.

We then have that with C equal to the hidden constant in (22),

$$\sum_{r=r_0}^n \sum_{\substack{k=1 \\ k \notin K_r}}^r \mathbb{P}(e_k \text{ is bad} \mid \mathcal{C}) \zeta_{r,k} \leq C\ell_1 \sum_{r=r_0}^n \frac{r^{1-20\varepsilon}}{r^{1-7\varepsilon}} + 2 \sum_{r=r_0}^n \ell_1 \sum_{k=k_0}^r \gamma_r \varepsilon_r \zeta_{r,k}. \tag{28}$$

Explanation: For each $a_i \in DT_k$, the set of possible bad edges does not increase for each $k' > k$. This is because each $a_i \in DT_k$ is associated with an alternating path that does not change with k'. The first term bounds the expected number of bad basic edges, using (22) and our bound on the diameter of DT_r. The second sum deals with non-basic edges and uses (26).

Now

$$\ell_1 \sum_{r=r_0}^{n} \frac{r^{1-20\varepsilon}}{r^{1-7\varepsilon}} \leq \ell_1 n^{1-12\varepsilon} \log n \tag{29}$$

and

$$\sum_{r=r_0}^{n} \sum_{k=1}^{r} \zeta_{r,k} \ell_1 \gamma_r \varepsilon_r \leq \ell_1 \gamma_{r_0} \varepsilon_{r_0} \sum_{r=r_0}^{n} \sum_{k=1}^{r} \zeta_{r,k} \leq \ell_1 \gamma_{r_0} \varepsilon_{r_0} n = o(\ell_1). \tag{30}$$

Finally, it follows from the fact that only edges of cost at most γ_r are added that for any $k \leq r$, $\mathbb{P}(e_k \text{ is bad} \mid \mathcal{C}) \leq \ell_1 \gamma_r$. (There are always at most ℓ_1 choices of edge that could be bad and the probability they have cost at most γ_r is $1 - e^{-\gamma_r} \leq \gamma_r$.) So, from Lemma 16,

$$\sum_{r=r_0}^{n} \sum_{k \in K_r} \mathbb{P}(e_k \text{ is bad} \mid \mathcal{C}) \zeta_{r,k} \leq \ell_1 \sum_{r=r_0}^{n} |K_r| \gamma_r \leq 8\ell_1 \sum_{r=r_0}^{n} r \gamma_r \varepsilon_r \leq \ell_1 n^{1-90\varepsilon}. \tag{31}$$

After adding the $O(\ell_1 r \gamma_r)$ contribution from the edges incident with a_{r+1}, b_{r+1}, this completes the justification for (19) and the proof of Lemma 3.

5 More general distributions

Replacing $C(i,j)$ by $aC(i,j)$ yields a distribution as in (1) where $a = 1$. We assume then that $a = 1$.

Suppose first that the cost density function can be re-expressed as $f(x) = e^{-bx + O(x^2)}$ as $x \to 0$, where $b \neq 0$. We let $F(x) = \mathbb{P}(C \geq x) = b^{-1} e^{-bx + O(x^2)}$ as $x \to 0$. In this case we run Karp's algorithm with the given costs. Let $E^+ = \{(i,j) : u_i + v_j \geq 0\}$. Equation (8) becomes

$$\mathbb{P}(T_r = T \mid \mathbf{u}, \mathbf{v}) = \prod_{(a_i, b_j) \in E^+ \setminus E(T)} F(u_i + v_j) \prod_{(a_i, b_j) \in E(T)} f(u_i + v_j).$$

So, for trees T_1, T_2 we have where ,

$$\begin{aligned} \frac{\mathbb{P}(T_r = T_1 \mid \mathbf{u}, \mathbf{v})}{\mathbb{P}(T_r = T_2 \mid \mathbf{u}, \mathbf{v})} &= \prod_{(a_i, b_j) \in E(T_1) \setminus E(T_2)} \frac{f(u_i + v_j)}{F(u_i + v_j)} \prod_{(a_i, b_j) \in E(T_2) \setminus E(T_1)} \frac{F(u_i + v_j)}{f(u_i + v_j)} \\ &= \prod_{(a_i, b_j) \in E(T_1) \setminus E(T_2)} b e^{O(\gamma_r^2)} \prod_{(a_i, b_j) \in E(T_2) \setminus E(T_1)} b^{-1} e^{O(\gamma_r^2)} \\ &= e^{O(r\gamma_r^2)} = o(1). \end{aligned}$$

So, we replace uniformity by near uniformity and this is enough for the proof.

When $b = 0$, such as when C is uniform $[0,1]$ then we proceed as follows: In the analysis above we have assumed that $f(x) = e^{-x}$ i.e. that the costs are distributed as exponential mean 1, $EXP(1)$. We extend the analysis to costs C with density function $f(x) = 1 + O(x^2)$ as $x \to 0$ as follows: Given $C(e) = x$, we define $\widehat{C}(e) = y = y(x)$ where $\widehat{C}(e)$ is exponential mean 1 and $\mathbb{P}(\widehat{C}(e) \leq y) = \mathbb{P}(C(e) \leq x)$ i.e. $1 - e^{-y} = \mathbb{P}(C(e) \leq x) = x + O(x^2)$, for small x. We then find that $y + O(y^2) = x + O(x^2)$ and so $x = y + O(y^2)$.

We run Karp's algorithm with C replaced by $y(C)$. We have shown that w.h.p. the tour found by Karp's heuristic has cost $\sum_{i=1}^{n} \widehat{U}_i$ where $\widehat{U}_i \leq \zeta_n$ for all i and so the corresponding C costs U_i satisfy $U_i = \widehat{U}_i + O(\widehat{U}_i^2)$. Consequently, the increase in cost of using C over $\widehat{C}$ is $O(n\zeta_n^2) = o(1)$. Of course it would be more satisfying to apply the algorithm directly to C and we conjecture that the proof can be modified to verify this.

6 Final Remarks

We have extended the proof of the validity of Karp's patching algorithm to random perturbations of dense digraphs with minimum in- and out-degree at least αn. It would be of interest to extend the analysis to the case where $\alpha > 1/2$ and there is no random set R. At the moment, the only gap in a proof comes from our inability to prove that there are enough acceptable non-basic edges to prove(26)

References

[1] R. Ahuja, T. Magnanti and J. Orlin, Network Flows: Theory, Algorithms and Applications, Prentice Hall, 1991.

[2] D. Aldous, The Random Walk Construction of Uniform Spanning Trees and Uniform Labelled Trees, *SIAM Journal on Discrete Mathematics* 3 (1990) 450-465.

[3] D. Aldous, The $\zeta(2)$ limit in the random assignment problem, *Random Structures and Algorithms* 4 (2001) 381-418.

[4] S. Antoniuk, A. Dudek and A. Ruciński, Powers of Hamiltonian cycles in randomly augmented Dirac graphs – the complete collection, *Journal of Graph Theory* 104 (2023) 811-835.

[5] I. Araujo, J. Balogh, R. Krueger, S. Piga and A. Treglown, On oriented cycles in randomly perturbed digraphs, *Combinatorics, Probability and Computing* 33 (2024) 157-178.

[6] J. Balogh, A. Treglown and A. Wagner, Tilings in randomly perturbed dense graphs, *Combinatorics, Probability and Computing* 28 (2019) 159-176.

[7] W. Bedenknecht, J. Han and Y. Kohayakawa, Powers of tight Hamilton cycles in randomly perturbed hypergraphs, *Random Structures and Algorithms* 55 (2019) 795-807.

[8] T. Bohman, A.M. Frieze and R. Martin, How many random edges make a dense graph Hamiltonian?, *Random Structures and Algorithms* 22 (2003) 33-42.

[9] T. Bohman, A.M. Frieze, M. Krivelevich and R. Martin, Adding random edges to dense graphs, *Random Structures and Algorithms* 24 (2004) 105-117.

[10] J. Böttcher, R. Montgomery, O. Parczyk and Y. Person, Embedding spanning bounded degree graphs in randomly perturbed graphs, *Electronic Notes in Discrete Mathematics* 61 (2017) 155-161.

[11] F. Chung, P. Horn and L. Lu, Diameter of random spanning trees of a given graph, *Journal of Graph Theory* 69 (2012) 223–240.

[12] D. Dadush and S. Huiberts, A friendly smoothed analysis of the simplex method, Proceedings of STOC 2018.

[13] S. Das and A. Treglown, Ramsey properties of randomly perturbed graphs: cliques and cycles, to appear in *Combinatorics, Probability and Computing* 29 (2020) 830-867.

[14] E. Dijkstra, A note on two problems in connexion with graphs, *Numerische Mathematik* 1 (1959) 269-271.

[15] P. Doyle and J. Snell, Random walks and electric networks.

[16] A. Dudek, C. Reiher, A. Ruciński and M. Schacht, Powers of Hamiltonian cycles in randomly augmented graphs, *Random Structures and Algorithms* 56 (2020) 122-141.

[17] M.E. Dyer and A.M. Frieze, On patching algorithms for random asymmetric travelling salesman problems, *Mathematical Programming* 46 (1990) 361-378.

[18] P. Erdős and A. Rényi, On random matrices, *Publ. Math. Inst. Hungar. Acad. Sci.* 8 (1964) 455-461.

[19] A.M. Frieze, Limit distribution for the existence of hamiltonian cycles in random bipartite graphs, *European Journal of Combinatorics* 6 (1985) 327-334.

[20] A.M. Frieze, The effect of adding randomly weighted edges, *SIAM Journal on Discrete Mathematics* 35 (2021) 1182-1200.

[21] A.M. Frieze and M. Karoński, Introduction to Random Graphs, Cambridge University Press, 2015.

[22] A.M. Frieze and G. Sorkin, The probabilistic relationship between the assignment and asymmetric traveling salesman problems, *SIAM Journal on Computing* 36 (2007) 1435-1452.

[23] J. Han and Y. Zhao, Hamiltonicity in randomly perturbed hypergraphs, *Journal of Combinatorial Theory, Series* B (2020) 14-31.

[24] W. Hoeffding, Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association* 58 (1963) 13-30.

[25] S. Janson, One, two and three times $\log n/n$ for paths in a complete graph with random weights, *Combinatorics, Probability and Computing* 8 (1999) 347-361.

[26] M. Jerrum and A. Sinclair, Approximate Counting, Uniform Generation and Rapidly Mixing Markov Chains, *Information and Computation* 82 (1989) 93-133.

[27] R.M. Karp, A patching algorithm for the non-symmetric traveling salesman problem, *SIAM Journal on Computing* 8 (1979) 561–573.

[28] R.M. Karp and J.M. Steele, Probabilistic analysis of heuristics, in The traveling salesman problem: a guided tour of combinatorial optimization, E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys Eds. (1985) 181-206.

[29] J. Kleinberg and E. Tardos, Algorithm Design, Addison-Wesley, 2005.

[30] M. Krivelevich, M. Kwan and B. Sudakov, Cycles and matchings in randomly perturbed digraphs and hypergraphs, *Combinatorics, Probability and Computing* 25 (2016) 909-927.

[31] M. Krivelevich, M. Kwan and B. Sudakov, Bounded-degree spanning trees in randomly perturbed graphs, *SIAM Journal on Discrete Mathematics* 31 (2017) 155-171.

[32] M. Krivelevich, B. Sudakov and P. Tetali, On smoothed analysis in dense graphs and formulas, *Random Structures and Algorithms* 29 (2006) 180-193.

[33] D. Levin, Y. Peres and E. Wilmer, Markov chains and mixing times, American Mathematical Society, 2017.

[34] R. Lyons and Y. Peres, Probability on trees and networks, Cambridge University Press, 2017.

[35] C. McDiarmid, On the method of bounded differences, *in Surveys in Combinatorics, ed. J. Siemons, London Mathematical Society Lecture Notes Series* 141, Cambridge University Press, 1989.

[36] A. McDowell and R. Mycroft, Hamilton ℓ-cycles in randomly-perturbed hypergraphs, *The electronic journal of combinatorics* 25 (2018).

[37] O. Parczyk, 2-universality in randomly perturbed graphs, *European Journal of Combinatorics* 87 (2020) 103-118.

[38] E. Powierski, Ramsey properties of randomly perturbed dense graphs.

[39] D. Spielman and S. Teng, Smoothed Analysis of the Simplex Algorithm, *Journal of the ACM* 51 (2004) 385-463.

[40] B. Sudakov and J. Vondrak, How many random edges make a dense hypergraph non-2-colorable?, *Random Structures and Algorithms* 32 (2008) 290-306.

[41] R. Vershynin, Beyond Hirsch Conjecture: Walks on Random Polytopes and Smoothed Complexity of the Simplex Method, *SIAM Journal on Computing* 39 (2009) 646-678.

A Definitions

$$
\begin{aligned}
\varepsilon &= \text{ a sufficiently small positive constant.}\\
\ell_1 &= n^{4\varepsilon}.\\
r_0 &= n^{1-3\varepsilon}.\\
\gamma_r &= r^{2\varepsilon-1}.\\
\beta_r &= \frac{\varepsilon\gamma_r}{10}.\\
\varepsilon_r &= r^{-30\varepsilon}.
\end{aligned}
$$

G_r : This is the bipartite subgraph of G induced by A_r, B_r.

M_r :This is the minimum cost perfect matching between A_r and B_r.

Γ_r : This is the subgraph of G_r induced by contracting M_r.

$G_r^*(\mathbf{u},\mathbf{v})$: This is the subgraph of G_r induced by the edges (a_i, b_j) for which $u_i + v_j \geq 0$.

$\Gamma_r^*(\mathbf{u},\mathbf{v})$: This is the graph obtained from $G_r^*(\mathbf{u},\mathbf{v})$ by contracting M_r.

$\vec{G}_r$: This is the digraph obtained by orienting the edges of G_{r+1} from A_{r+1} to B_{r+1}, except for the edges of M_r, which are oriented from B_r to A_r.

$\vec{\Gamma}_r$: This is the digraph obtained from $\vec{G}_r$ by contracting M_r.

T_r : This is the spanning tree of G_r corresponding to an optimal basis.

$\widehat{T}_r$: This is the tree obtained from T_r by contracting M_r.

DT_k : This is the tree comprising the first k vertices selected by Dijkstra's algorithm below.