Karp’s patching algorithm on random perturbations of dense digraphs

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

July 20, 2024

Abstract

We consider the following question. We are given a dense digraph $D_0$ with minimum in- and out-degree at least $\alpha n$, 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 costs $C(e), e \in E(D)$, where $C(e)$ is an independent copy of the exponential mean one random variable $\text{EXP}(1)$ i.e. $\mathbb{P}(\text{EXP}(1) \geq x) = e^{-x}$. 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 that is asymptotically equal to that of the associated assignment problem. Karp’s algorithm runs in polynomial time.

1 Introduction

Let $\mathcal{D}(\alpha)$ be the set of digraphs with vertex set $[n]$ and with minimum in- and out-degree at least $\alpha n$. We are given a digraph $D_0 \in \mathcal{D}(\alpha)$ 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 $n^{-\epsilon}$ where $\epsilon > 0$ is a small positive constant. The edges $E(D)$ of $D$ are given costs $C(e), e \in E(D)$, where $C(e)$ is an independent copy of the exponential mean one random variable $\text{EXP}(1)$ i.e. $\mathbb{P}(C(e) \geq x) = e^{-x}$. Let $C(i,j), i,j \in [n]$ be the associated $n \times n$ cost matrix where $C(i,j) = \infty$ if $(i,j) \notin E(D)$. One is 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 asymptotically solve the latter.

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. 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 $v(\text{ATSP}) \geq v(\text{AP})$ where $v(\bullet)$ denotes the optimal cost. Karp [27] considered

∗Research supported in part by NSF grant DMS1952285
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. \( v(\text{ATSP}) = (1 + o(1))v(\text{AP}) \). 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 \( A_M = \{C_1, C_2, \ldots, C_t\} \) 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 \( D_0 \in \mathcal{D}(\alpha), \alpha > 0 \) where \( \alpha \) 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 \( \text{EXP}(1) \) cost. Here an edge \( e \notin E(D_0) \) is placed independently into \( R \) with probability \( n^{-\varepsilon} \) where \( \varepsilon > 0 \) is a small positive constant. Then w.h.p. \( v(\text{ATSP}) = (1 + o(1))v(\text{AP}) \) and Karp’s patching algorithm finds a tour (Hamilton cycle) of the claimed cost in polynomial time.

The use of \( \text{EXP}(1) \) as opposed to \( U(1) \) is an artifact of our proof. In particular it enables us to claim that a certain tree is uniformly distributed among the spanning trees of a certain graph, see Lemma \([9]\).

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 \( \mathcal{D}(\alpha) \) is a good candidate, but we can only guarantee Hamiltonicity if \( \alpha \geq 1/2 \). If we want to allow arbitrary \( \alpha \) 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 \([11]\) and Dadush and Huiberts \([12]\).

The paper \([39]\) 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 \( \alpha \) is a positive constant and \( \mathcal{G}(\alpha) \) is the set of graphs with vertex set \([n]\) and minimum degree at least \( \alpha n \). 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 \( 1/2 n \log n \) edges needed if we rely only on the random edges. Research on this model and its variations has been quite substantial, see for example \([1], [5], [6], [7], [9], [10], [13], [16], [23], [30], [31], [32], [36], [37], [38], [40]\).

**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 \( A_M \) 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 solving the AP. We prove the following:

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

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

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 \( \bar{K}_n \). Karp’s proof is very clean but rather fragile. It relies on the key insight that if \( D = \bar{K}_n \) then the optimal assignment comes from a uniform random permutation. This seems unlikely to be true in general 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 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 \( \gamma_n \) 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 \( \pi_C \) satisfies

\[
\pi_C \geq \sum_{i<j} k_i k_j = \frac{1}{2} \left( n^2 - \sum_{i=1}^{\ell} k_i^2 \right) \geq \frac{1}{2} \left( n^2 - \left( (n-\ell+1)^2 + \ell - 1 \right) \right) \geq \frac{\ell n}{2}.
\]

Each of these \( \pi_C \) pairs uses a disjoint set of edges. We define the sets

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

Each edge of \( E(\bar{K}_n) \setminus E(D_0) \) appears in \( R_\ell \) with probability at least \( p_\ell = n^{-\epsilon} \left( \frac{1 - o(1)}{\ell n^{1-5\epsilon/2}} \right)^{1/2} \), independent of other edges. (The factor \( n^{-\epsilon} \) 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 \( R_\ell \). If \( \mathcal{E}_\ell \) does not occur then we reduce the number of cycles by at least one. We have

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

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

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

Given the last equality and the fact that w.h.p. \( v(\text{AP}) > (1 - o(1))\zeta(2) > 1 \) we see that Karp’s patching heuristic is asymptotically optimal. The lower bound of \( (1 - o(1))\zeta(2) \) on \( v(\text{AP}) \) 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 \( \varepsilon \gamma / 10 \). (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 at most \( \gamma n \). 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 \( Bin(N, p) \).

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

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 \( B \). For \( r \geq r_0 \) 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.** The degree of a vertex is hypergeometric 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 dominated by the probability that there is an isolated vertex. And this is at most \( 2r(1 - p)^r \leq 2ne^{-\varepsilon \gamma} = o(n^{-1}) \). This verifies connectivity and the existence of a perfect matching. \( \square \)

**Lemma 5.** For a set \( S \subseteq A_r \) we let

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

If \( r \geq r_0 \) then with probability \( 1 - e^{-\Omega(\varepsilon r^2/2)} \),

\[
|N_0(S)| \geq \frac{\varepsilon r^\varepsilon |S|}{40} \text{ for all } S \subseteq A_r, 1 \leq |S| \leq r^{1 - \varepsilon}. \tag{2}
\]

**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

\[
1 - q_s = (1 - n^{-\varepsilon} + n^{-\varepsilon} e^{-\beta_r})^s \leq (1 - \frac{1}{2} n^{-\varepsilon} \beta_r)^s. \quad \text{It follows that } q_s \geq n^{-\varepsilon} \beta_r s/3 \text{ for } s \leq r^{1 - \varepsilon} \text{ and so } rq_s \geq \frac{\varepsilon r^\varepsilon s}{30}.
\]
Let \( \nu_s = \frac{\varepsilon r^{2}}{40} \). Then, using the Chernoff bounds, we have

\[
\mathbb{P} \left( \left| \frac{v_s}{\mathbb{E}[v_s]} \right| \leq 1-\varepsilon \right) \leq \sum_{s=1}^{r-1-2\varepsilon} \left( \frac{rs}{s} \right)^{s} e^{-\Omega(\varepsilon r^{2}/2s)} = \sum_{s=1}^{r-1-2\varepsilon} \left( \frac{rs}{s} \cdot e^{-\Omega(\varepsilon r^{2}/2)} \right)^{s} = e^{-\Omega(\varepsilon r^{2}/2)}.
\]

We let \( \text{AP}_r \) denote the problem of finding a minimum weight matching between \( A_r \) and \( B_r \). Let \( M_r \) denote the optimal solution to \( \text{AP}_r \).

**Lemma 6.** If \( r \geq r_0 \) then with probability \( 1 - e^{-\Omega(\varepsilon r^{2}/2)} \), \( M_r \) contains only edges of cost \( C(i,j) \leq \gamma_r \).

**Proof.** Suppose that \( M_r \) contains an edge \( e \) of cost greater than \( \gamma_r \). Assume w.l.o.g. that \( e = (a_1, b_1) \). Let an alternating path \( P = (a_1 = x_1, y_1, \ldots, y_{k-1}, x_k, y_k = b_1) \) be acceptable if (i) \( x_1, \ldots, x_k \in A_r, y_1, \ldots, y_k \in B_r \), (ii) \( (x_{i+1}, y_i) \in M_r, i = 1, 2, \ldots, k-1 \) and (iii) \( C(x_i, y_i) \leq \beta_r, i = 1, 2, \ldots, k \). The existence of such a path with \( k \leq 5\varepsilon^{-1} \) implies the existence of another perfect matching with cost \( C(M_r) + k\beta_r - C(e) < C(M_r) \), which contradicts the optimality of \( M_r \). We show below that w.h.p. there is such a path.

Now consider the sequence of sets \( S_0 = \{a_1\}, S_1, S_2, \ldots \subseteq A, T_1, T_2, \ldots \subseteq B \) defined as follows:

\[
T_i = N_0 \left( \bigcup_{j<i} S_j \right) \quad \text{and} \quad S_i = \phi^{-1}(T_i), \quad \text{where} \quad M_r = \{(a_i, \phi(a_i)) : i = 1, 2, \ldots, r\}.
\]

It follows from (2) that w.h.p.

\[
|S_i| = |T_i| \leq r^{1-\varepsilon} \text{ implies that } |S_i| \geq \left( \frac{\varepsilon r^{2}}{40} \right)^{i}.
\]

So define \( i_0 \) to be the smallest integer \( i \) such that \( \left( \frac{\varepsilon r^{2}}{40} \right)^{i} \geq r^{1-\varepsilon} \). Note that \( i_0 \leq 2/\varepsilon \). 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. \( |S_{i_0+1}| \geq \frac{\varepsilon r^{2}}{40} \).

For a set \( T \subseteq B_r \) we let

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

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

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,

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

It follows that w.h.p. there will be an edge in \( E_R(S_{i_0+1}, \tilde{T}_{i_0+1}) \) and we have found an alternating path of length at most \( 2i_0 + 3 \) using edges of cost at most \( \beta_r \) and this completes the proof of Lemma 6 and hence Lemma 2.
4 Proof of Lemma 3

We analyse the solution of $A_{P_r}$ via 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 an augmenting path that increases the matching cost by the minimum. We use a standard trick to make the edge costs in this problem non-negative. Given this, we prove Lemma 3 by showing that Dijkstra’s algorithm creates few cycles w.h.p.

4.1 Linear programming formulation of AP

We consider the linear program $\mathcal{LP}_r$ that underlies the assignment problem and its dual $D_r$. 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.

We consider the linear program $\mathcal{LP}_r$ for finding $M_r$. To be precise we 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 $D_r$ 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 $T_r$, 

$$C(i,j) = u_i + v_j \text{ for } (a_i, b_j) \in E(T_r) \quad (3)$$

and 

$$C(i,j) \geq u_i + v_j \text{ for } (a_i, b_j) \notin E(T_r). \quad (4)$$

Note that if $\lambda$ is arbitrary then replacing $u_i$ by $\hat{u}_i = u_i - \lambda, i = 1, 2, \ldots, r$ and $v_i$ by $\hat{v}_i = v_i + \lambda, i = 1, 2, \ldots, r$ has no affect on these constraints. We say that $(u, v)$ and $(\hat{u}, \hat{v})$ are equivalent. It follows that we can always fix the value of one component of $(u, v)$.

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

Lemma 7. (a) Fix $(u,v)$. If $T_1, T_2$ are distinct spanning trees of $G_r$ then $C(T_1, u, v) \cap C(T_2, u, v)$ has measure zero, given $(u,v)$.

(b) If $u_1 = u_1' = 0$ and $(u,v) \neq (u',v')$ then for any spanning tree $T$ of $G_r$, we have that $C(T, u, v) \cap C(T, u', v') = \emptyset$. 

Proof. (a) Suppose that \( C \in \mathcal{C}(T_1, u, v) \cap \mathcal{C}(T_2, u, v) \). We root \( T_1, T_2 \) at \( a_1 \) and let \( u_1 = 0 \). The equations \( T \) 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 \( u, v \). \( \square \)

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 \( |u_i|, |v_j| \leq 2\gamma_r \) for all \( i, j \).

Lemma 8. \( \mathbb{P}(\mathcal{E}) = 1 - o(n^{-1}) \).

Proof. 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 \( T \) holds. We also know that if \( \mathcal{T} \) occurs then \( u_{i'} + v_j \leq C(i', j) \) for all \( i' \neq i \). It follows that \( u_i - u_i' \geq C(i,j) - C(i',j) \geq -\gamma_r \) for all \( i' \neq i \). (We have used the fact that we do not need to consider edges of cost greater than \( \gamma_r \) to find \( M_r \), see Lemma 2.) 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 \( i \in r \).

Now because of the optimal matching edges \((i, \phi(i)), i \in [r]\) we have \( u_i + v_{\phi(i)} = C(i, \phi(i)) \), we see that \( |v_j| \leq 2\gamma_r \) for \( j \in [r] \). \( \square \)

The next two lemmas help us to understand the structure of the tree \( T_r \). Fix \( M_r \) and let \( G^*_r(u,v) \) 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 \( \mathcal{T}_r(u,v) \) denote the set of spanning trees of \( G^*_r(u,v) \) that contain the edges of \( M_r \). This is non-empty because \( T_r \in \mathcal{T}_r(u,v) \).

Lemma 9. If \( T \in \mathcal{T}_r(u,v) \) then

\[ \mathbb{P}(T_r = T \mid u, v) = \prod_{(a_i, b_j) \in G^*_r(u,v)} e^{-(u_i + v_j)}, \]

which is independent of \( T \).

Proof. Fixing \( u, v \) 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 \( x^+ = \max \{x, 0\} \). Thus,

\[ \mathbb{P}(T_r = T \mid u, v) = \prod_{(a_i, b_j) \notin E(T)} e^{-(u_i + v_j)^+} \prod_{(a_i, b_j) \in E(T)} e^{-(u_i + v_j)} = \prod_{(a_i, b_j) \in G^*_r(u,v)} e^{-(u_i + v_j)}. \]

Thus

\[ T_r \] is a uniform random member of \( \mathcal{T}_r(u,v) \).

\hfill (7)

The next lemma will show that \( G^*_r \) has a large minimum degree. 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 \( L_{i,-} = \{ j : (i,j) \in E(G) \} \). Then for \( i = 1, 2, \ldots, r \) and \( \eta > 0 \) let \( A_{i,+} = A_{i,+}(\eta) \) be the event that \( \{ j \in L_{i,+} : u_i + v_j \geq 0 \} \leq \eta r \) and let \( A_{j,-} = A_{j,-}(\eta) \) be the event that \( \{ i \in L_{j,-} : u_i + v_j \geq 0 \} \leq \eta r \).
Lemma 10. Fix a spanning tree \( T \) of \( G^*_r \) that contains \( M_r \). Then there exists a small positive constant \( \eta \) such that
\[
P(\mathcal{A}_{i,+}(\eta) \lor \mathcal{A}_{j,-}(\eta) \mid T_r = T) = O(e^{-\Omega(\varepsilon r^2)}) \text{ for } i, j = 1, 2, \ldots, r.
\]

Proof. We assume that \( C(i, j) \leq \gamma_r \) for \((a_i, b_j) \in T \). The justification for this is Lemma 6 in that we can solve the assignment problem, only using edges of cost at most \( \gamma_r \).

The number of edges in \( G_r \) of cost at most \( \gamma_r \) incident with a fixed vertex is dominated by \( \text{Bin}(r, \gamma_r) \) and so with probability \( 1 - e^{-\Omega(\varepsilon^2)} \) 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 \( \mathcal{B} \) be the event that \( C(i, j) > u_i + v_j \) for all \((a_i, b_j) \notin E(T) \). Note that if \( \mathcal{B} \) occurs then \( T_r = T \).

We now condition on the set \( E_T \) of edges (and the associated costs) of \( \{(a_i, b_j) \notin E(T)\} \) such that \( C(i, j) \geq 2\gamma_r \). Let \( \mathcal{F}_T = \{(a_i, b_j) \notin E(T)\} \setminus E_T \). Note that \( |\mathcal{F}_T| \) is dominated by \( \text{Bin}(r^2, 1 - e^{-2\gamma_r}) \) and so \( |\mathcal{F}_T| \leq 3r^2\gamma_r \) 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 \( \mathcal{A}_{s,+} \land \mathcal{E} \). We write,
\[
P(\mathcal{A}_{s,+} \mid \mathcal{B}) = P(\mathcal{A}_{s,+} \land \mathcal{E} \mid \mathcal{B}) = \frac{\int \delta_1(Y)P(\mathcal{B} \mid Y)dC}{\int P(\mathcal{B} \mid Y)dC} \tag{8}
\]

Then we note that since \((a_i, b_j) \notin \mathcal{F}_T \cup E(T)\) satisfies the condition \[4\],
\[
P(\mathcal{B} \mid Y) = \prod_{(a_i, b_j) \in \mathcal{F}_T} \exp \left\{-u_i(Y) + v_j(Y)\right\} = e^{-W}, \tag{9}
\]
where \( W = W(Y) = \sum_{(a_i, b_j) \in \mathcal{F}_T} u_i(Y) + v_j(Y) \leq 12r^2\gamma^2 = 12r^{4\varepsilon} \). Then we have
\[
\int Y \delta_1(Y)P(\mathcal{B} \mid Y) dC = \int Y e^{-W} \delta_1(Y) dC \\
\leq \left(\int Y e^{-2W} dC\right)^{1/2} \times \left(\int Y \delta_1(Y)^2 dC\right)^{1/2} \\
= e^{-E(W)} \left(\int Y e^{-2(W-E(W))} dC\right)^{1/2} \times P(\mathcal{A}_{s,+} \mid \mathcal{E})^{1/2} \\
\leq e^{-E(W)} e^{12r^{4\varepsilon}} P(\mathcal{A}_{s,+} \mid \mathcal{E})^{1/2}. \tag{10}
\]

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 \( T \).

It then follows from \[8\], \[10\] and \[11\] that
\[
P(\mathcal{A}_{s,+} \mid \mathcal{B}) \leq e^{12r^{4\varepsilon}} P(\mathcal{A}_{s,+} \mid \mathcal{E})^{1/2}. \tag{12}
\]

Equation \[12\] removes the conditioning on \( \mathcal{B} \) at the expense of inflating our probability estimate by \( e^{12\gamma^2 \log^2 n} \). For the remainder of the lemma we assume that the \( C(i, j) \) for \((a_i, b_j) \in T\) satisfy \( C \leq \gamma^* \) and that \( \mathcal{E} \) holds.
Denote this conditioning by $\mathcal{F}$. Note that if $B$ occurs and (3) holds then $T^* = T$. Let $b_j$ be a neighbor of $a_s$ in $G^*_r$ and let $P_j = (i_1 = s, j_1, j_2, \ldots, i_k, j_k = j)$ define the path from $a_s$ to $b_j$ in $T$. Then it follows from (3) that $v_{i_{j_1}} = v_{i_{j_1-1}} - C(i_{j_1-1}, i_{j_1})$. 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 $\text{EXP}(1)$ that are conditionally bounded above by $\gamma_r$. Given $\mathcal{E}$ we can assume that the partial sums $S_i$ satisfy $|S_i| \leq 2\gamma_r$ for $i = 1, 2, \ldots, k - 1$. Assume for the moment that $k \geq 4$ and let $x = u_{i_{k-3}} \in [-2\gamma_r, 2\gamma_r]$. Given $x$ we see that there is some positive probability $p_0 = p_0(x)$ that $S_k > 0$. Indeed,

$$p_0 = \mathbb{P}(S_k > 0 \mid \mathcal{E}) \geq \mathbb{P}(x + Z_1 - Z_2 > 0) - \mathbb{P}(\neg \mathcal{E}), \quad (13)$$

where $Z_1 = Z_{1,1} + Z_{1,2} + Z_{1,3}$ and $Z_2 = Z_{2,1} + Z_{2,2}$ are the sums of independent $\text{EXP}(1)$ random variables, each conditioned on being bounded above by $\gamma_r$ and such that $|x + \sum_{j=1}^t (Z_{1,j} - Z_{2,j})| \leq 2\gamma_r$ for $t = 1, 2$ and that $|x + Z_1 - Z_2| \leq 2\gamma_r$. The absolute constant $\eta_0 = p_0(-2\gamma_r) > 0$ is such that $\min \left\{ x \geq -2\gamma_r : p_0(x) \right\} \geq \eta_0$.

We now partition (most of) the neighbors of $a_s$ into $N_0, N_1, N_2$ where $N_t = \{ b_j : k \geq 3, k \mod 3 = t \}$, $k$ being the number of edges in the path $P_j$ from $a_s$ to $b_j$. Now because $T$ has maximum degree $2r^{2e}$, as observed at the beginning of the proof of this lemma, we know that there exists $t$ such that $|N_t| \geq (\alpha_0 r - (2r^{2e})^3)/3 \geq \alpha r/4$, where $\alpha_0 \sim \alpha$ as in Lemma $[\text{4}]$. It then follows from (13) that $|L_{s,+}|$ dominates $\text{Bin}(\alpha r/4, \eta_0 - o(1))$ and then $\mathbb{P}(|L_{s,+}| \leq \alpha \eta_0/10) = O(e^{-\Omega(r)})$ follows from the Chernoff bounds. Similarly for $L_{s,-}$. Applying the union bound over $r$ choices for $s$ and applying (12) gives

$$\mathbb{P}(\exists s : A_{s,+} \lor A_{s,-}) \leq re^{12r^{2e} - \Omega(r)} = O(e^{-\Omega(r^{r/2})}).$$

Thus the lemma holds with $\eta = \eta_0/10$. \hfill $\Box$

### 4.1.2 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 $\bar{G}_r$ 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 $P_r = (x_1 = a_{r+1}, y_1, x_2, \ldots, x_r, y_r = b_{r+1})$ from $a_{r+1}$ to $b_{r+1}$ in $\bar{G}_r$. 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

$$u_{r+1} = \min \{ C(r+1, j) - v_j(T_r) : j \in [r] \} \quad \text{and} \quad v_{r+1} = \min \{ C(r+1, r+1) - u_{r+1}, \min \{ C(i, r+1) - u_i(T_r) : i \in [r] \} \}. \quad (14)$$

We use costs $\tilde{C}(i, j) = C(i, j) - u_i - v_j$ in our search for a shortest augmenting path. Our choice of $u_{r+1}, v_{r+1}$ and (3), (4) implies that $\tilde{C}(i, j) \geq 0$ and that matching edges have cost zero. This idea for making edge costs non-negative is well known, see for example Kleinberg and Tardos [20]. The $\tilde{C}$ 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 $u, v \in U = \{ u_i, v_i : |u_i|, |v_i| \leq 2\gamma_r \}$ and that for all $i$, neither $A_{i,+}$ nor $A_{i,-}$ of Lemma 10 hold. The constraints (3), (4)
on the \( C(i, j) \) become that
\[
C(i, \phi(i)) = u_i + v_{\phi(i)} \quad \text{for } i = 1, 2, \ldots, r \\
C(i, j) \geq u_i + v_j, \quad \text{otherwise.} \tag{15}
\]

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 \( C^+ \) that will only be needed in Section 4.1.5 when we deal with non-basic edges. Not only will we fix \( M_r \), but we will also fix \( T_r \) and \( u, v \in U \).

\subsection*{4.1.3 Dijkstra’s algorithm}

We let \( \Gamma^* = \Gamma^*_r(u, v) \) 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 \( G^*_r(u, v) \). Of course, \( T_r \) only contain at most one of such a pair. It follows from \( \{7\} \) that given \( u,v, \widehat{T}_r \) is a uniform random spanning tree of \( \Gamma^*_r \).

We use Dijkstra’s algorithm to find the shortest augmenting path from \( a_{r+1} \) to \( b_{r+1} \) in the digraph \( \widehat{G}_r \). 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 \( \Gamma_r \) with vertex set \( A_{r+1} \). The edges of \( \Gamma_r \) are derived from paths \((a_i, \phi(a_j), a_j)\) in \( \widehat{G}_r \). (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 \( \widehat{G}_r \). 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 \( \widehat{G}_r \).

The Dijkstra algorithm applied to \( \widehat{G}_r \) 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 \( d_i \) for \( a_i \notin S_k \) is our current estimate for the cost of a shortest path from \( a_{r+1} \) to \( a_i \). The algorithm chooses \( a_j \notin S_k \) to add to \( S_k \) to create \( S_{k+1} \). Here \( l^* \) minimises \( d_i + \widehat{C}(i, j) \) over \( a_i \in S_k, a_j \notin S_k \). It then updates the \( d_i, a_i \notin S_{k+1} \) 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 \( u, v \) and \( T_r \) the only restriction on \( C(i, \ell) \) for \( (i, \ell) \) 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 \( C^+ \), the non-basic/non-tree values \( C(i, \ell) \) are independently distributed as follows:

\[
\text{If } \theta_{i, \ell} \geq 0 \text{ then } d_i + \widehat{C}(i, \ell) - d_k \text{ is distributed as } EXP(1). \\
\text{Otherwise, } d_i + \widehat{C}(i, \ell) - d_k \text{ is distributed as } -\theta_{i, \ell} + EXP(1) \leq u_\ell - u_i + EXP(1). \tag{16}
\]

\subsection*{4.1.4 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 \( \widehat{M}_{r, r} \) is the matching obtained from \( M_r \) by adding the edges \((x_k, y_k), k = 1, 2, \ldots, \tau \) and deleting the edges \((x_{k+1}, y_{k+1}), k = 1, 2, \ldots, \tau - 1 \). Suppose now that \( x_r = a_i \) and \( y_r = b_j \). Observe that vertex \( i \) is the head of a path, \( Q \) say, in the set of paths and cycles \( A_{\widehat{M}_{r, r}} \). (\( Q \) is directed towards \( j \).) We say that vertex \( x_r \) creates a short cycle if \( j \) lies on \( Q \) and the sub-path of \( Q \) from \( j \) to \( i \) has length at most \( \ell_1 := n^{1e} \). In this case we also
say that the edge $(i, j)$ creates a short cycle. Extending the notation, we say that $x_\alpha$ creates a short cycle if $r + 1$ ($y_\alpha = b_{r+1}$) is the tail of $Q$ and the length of $Q$ is at most $\ell_1$. 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 $\chi_r$ denote the number of virgin short cycles created in iteration $r$. We then have that

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

(17)

Here $n/\ell_1$ bounds the number of large cycles induced by $M_n$, i.e. those of length greater than $\ell_1$. 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 $M_a$. 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

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

(18)

Assume (18) for the moment. Then we have,

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

(19)

Lemma 3 now follows from the Markov inequality. It only remains to prove (18).

4.1.5 Proof of (18)

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 $\xi_i$ 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. $\xi_j \leq r\gamma_r$ for all $j \leq r$. To create the shortest augmenting path from $a_{r+1}$ to $b_{r+1}$ we must find the minimum $\mu^*$ of the $C(a_{r+1,j}) + \xi_j$. There are at least $a_0 r\varepsilon$ indices $j$ for which the edge $(a_{r+1}, b_j)$ exists in $G_r$, see Lemma 4. It follows that w.h.p. $\mu^* \leq \min_j C(a_{r+1,j}) + r\gamma_r \leq 2r\gamma_r$. There are at most $\ell_1$ indices $j$ that would lead to the creation of a short cycle and for these the probability that $C(a_{r+1,j}) + j \leq 2r\gamma_r$ is at most $2\gamma_r$. Thus in expectation, edges incident with $a_{r+1}$ in this context, only contribute $O(\ell_1 r\gamma_r)$ 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 $\hat{C}$ 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 $\ell_1$ 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 $G_r$ pointing into $a_i$.

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

$$\mathbb{P}((a_i, x) \in \tilde{T}_r) = O\left(\frac{1}{r}\right)$$

from which we can deduce that

$$\mathbb{P}(\text{an added basic edge is bad}) = O\left(\frac{\ell_1}{r}\right),$$

where $\text{bad}$ means that the edge creates a short cycle.

To prove (20) 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 $\mathbb{P}(e \in T) = R_{\text{eff}}(a, b)$ where $R_{\text{eff}}$ denotes effective resistance, see for example [34]; (ii) $R_{\text{eff}}(a, b) = \frac{\tau(a, b) + \tau(b, a)}{2|E(G)|}$ 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 (20), we may have exposed some edges of $T_r$. Fortunately, edge inclusion in a random spanning tree is negatively correlated i.e. $\mathbb{P}(e \in T_r | f_1, \ldots, f_s \in T_r) \leq \mathbb{P}(e \in T_r)$, see for example [34].

Given (i) and (ii) and Lemma [9] it only remains to show that with $G = \Gamma^*_r = \Gamma^*_r(u, v)$ that $\tau(a, x) = O(r)$, for $a, x \in A_r$. For this we only have to show that the mixing time for a random walk on $\Gamma_r$ is sufficiently small. After this we can use the fact that the expected time to visit a vertex $a$ from stationarity is $1/\pi_a \leq r/\eta\alpha$ where $\eta$ is from Lemma [10] and where $\pi$ denotes the stationary distribution, see for example [33]. We estimate the mixing time of a walk by its conductance.

Let $deg(v) \geq \eta r$ denote degree in $\Gamma^*_r$. For $S \subseteq A_r$, let $\Phi_S = e(S, \tilde{S})/deg(S)$ where $e(S, \tilde{S})$ is the number of edges of $\Gamma^*_r(u, v)$ 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 $\deg(\tilde{S}) \geq \deg(A_r)/2 \geq \eta r^2/2$ which implies that $|\tilde{S}| \geq \eta r/2$ and so $|S| \leq (1 - \eta/2)r$.

Assume first that $|S| \leq \eta r/2$. Then

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

If $\eta r/2 \leq |S| \leq (1 - \eta/2)r$ then we use the random edges $R$. We sum over the $2^{O(r)}$ choices for $S$ and the $r^{O(r)}$ choices for $\tilde{T}_r$. Then, as in the final paragraph of the proof of Lemma [10] we see via the Chernoff bounds that with probability $1 - e^{-\Omega(r^{2-\varepsilon})}$ there are at least $\eta(1 - \eta/2)r^{2-\varepsilon}/3$ edges in $R$ from $S$ to $\tilde{S}$. The failure probability $e^{-\Omega(r^{2-\varepsilon})}$ is small enough to handle the $r^{O(r)}$ choices of $S, T_r$. So,

$$\Phi_S \geq \frac{\eta(1 - \eta/2)r^{2-\varepsilon}/3}{r^{2}/2} = \frac{2\eta(1 - \eta/2)r^{-\varepsilon}}{3}.$$

It then follows that after $r$ steps of the random walk the total variation distance between the walk and the steady state is $e^{-\Omega(r^{1-2\varepsilon})}$, see for example [33]. This completes our verification of (20) and hence (21).

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 11. The diameter of $\hat{T}_r$ is $O(r^{1/2+3\varepsilon})$ with probability $1-o(r^{-2})$.

Proof. Let $A$ be the adjacency matrix of $\Gamma^*_r$ and let $D$ be the diagonal matrix of degrees $\text{deg}(v), v \in A$, 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

$$\sigma \leq 1 - \frac{1}{2} \left( \frac{2\eta(1-\eta/2)r^{-\varepsilon}}{3} \right)^2 \leq 1 - \frac{\eta^2}{20r^{2\varepsilon}}. \quad (23)$$

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

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

It is shown in [11] that

$$\mathbb{P}(\text{diam}(T) \geq 2(\rho_0 + js)) \leq \frac{r}{2^{j-2}}. \quad (24)$$

Putting $j = 5\log r$ into (24) yields the lemma.

(Unfortunately, there are no equation references for (24). 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 $\Gamma^*_r$ and in our context we can drop the $\lambda_{n-1}$ term.)

Non-Basic Edges Each $a_i \in DT_k$ corresponds to an alternating path $P_i$. As such there are at most $\ell_1$ choices of $\ell$ such that $(i, \ell)$ 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 $\text{EXP}(1)$ subject to [15], as Dijkstra’s algorithm progresses, we learn lower bounds on $C(i, \ell)$ through (16). For this part of the argument we condition as for $C_+$. Note that assuming $u, v \in U$ 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 $(i, \ell)$ “competing” to be the next edge added to $DT_k$. This makes the choice of a bad edge unlikely. Examining (16) we see that for there to be any chance that an edge $(i, \ell)$ 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 $i = i_1, i_2, \ldots, i_r$ and that $u_{j_1} < u_{j_2} < \cdots < u_{j_r}$ and that $j = j_1, j_2, \ldots, j_r$. For $r_0 < j \leq r$ let

$$F(i, j) = |\Phi(i, j)|$$

where $\Phi(i, j) = \{ t > j : u_i \leq u_{i_t} + \gamma_r \varepsilon_r \}$ where $\varepsilon_r = r^{-30\varepsilon}$.

Let $X_r(i) = \{ j \leq r: F(i, j) \leq r\varepsilon^2_r \}$. If $j \notin X_r(i)$ let $\tilde{\Phi}(i, j)$ be the first $r\varepsilon^2_r$ members of $\Phi(i, j)$ in the sequence $i$. Note that $X_r(i)$ is determined by $ij$.

Lemma 12. $|X_r(i)| \leq 4r\varepsilon_r$.

Proof. Assume without loss that $i_t = t$ and replace the notation $\Phi(i, j)$ by $\Phi(u, j)$. 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 $u$ by interchanging $u_k$ and $u_r$. Fix $j < r$. We enumerate the possibilities and show that $F(u, j)$ does not increase.

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

13
So in all cases \( F(u, j) \) 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 \( X_r(i) \) 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 \( X_r(i) \) with \( k > u_{k_1} + \gamma_r \varepsilon_r \) and let \( J_2 = [u_{k_2}, u_{k_2} + \gamma_r \varepsilon_r] \) and so on. Using the fact that \( u \in U \) we see that in this way we cover \( X_r(i) \) with at most \( 4 \varepsilon_r^{-1} \) intervals each containing at most \( r \varepsilon_r^2 \) of the values \( u_j \) for which \( j \in X_r(i) \). □

Now let

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

And for each \( k \not\in K_r \), let \( \Psi_k \) be the first \( r \varepsilon_r^2/4 \) members of \( [k - r \varepsilon_r^2/2, k] \setminus X_r(i) \) in the sequence \( i \).

**Lemma 13.** \( |K_r| \leq 2|X_r(i)| \leq 8r \varepsilon_r. \)

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

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

\[
z \leq \sum_{j \in X_r(i)} r \varepsilon_r^2/2 \leq r \varepsilon_r^2 |X_r(i)|/2.
\]

and the lemma follows from Lemma 12. □

It follows from the definition of \( K_r \) that if \( k \not\in K_r \) then there are at least \( r \varepsilon_r^2/4 \times r \varepsilon_r^2 \) pairs \( (j \in \Psi_k, \ell \in \hat{\Phi}(i, j)) \) 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 \( k \not\in K_r \) and \( r_0 \leq k \leq r \) and \( j \leq k < \ell \leq \varepsilon_r \gamma_r \). The Chernoff bounds imply that w.h.p. \( r^2 \varepsilon_r^4 n^{-5}/5 \gg r \) 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 \( i,j \) and hence of \( \hat{\Psi}_k, \hat{\Phi}(i, j), j \not\in X_r(i) \).) Given this, it follows from the final inequality in 16 that

\[
\mathbb{P}(\text{an added non-basic edge is bad } | C+) \leq \ell_1 \left( \varepsilon_r \gamma_r + \frac{5n^{-\varepsilon}}{r^2 \varepsilon_r^4} \right) \leq 2\ell_1 \varepsilon_r \gamma_r. \quad (25)
\]

**Explanation:** There are at most \( \ell_1 \) 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 \( r^2 \varepsilon_r^4 n^{-5}/5 \) 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 \( t_k, k \not\in K_r \) 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

\[
\sum_{t_1 < t_2 < \cdots < t_L \atop x_1, x_2, \ldots, x_L} (2 \varepsilon_r \gamma_r)^L \leq \left( \frac{r}{L} \right)^2 (2 \varepsilon_r \gamma_r)^L \leq \left( \frac{2r^2 \varepsilon_r^2 \gamma_r}{L^2} \right)^L = o(n^{-2}), \quad (26)
\]
if $L^2 \geq 3e^2\varepsilon r^2$ or $L \geq 3e^2r^{1/2-16\varepsilon}$.

**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 (25).

Combining Lemma 11 and (26) we obtain a bound of $r^{1-13\varepsilon}$ on the diameter of $DT_r$. (Each path in $DT_r$ consists of a sequence of non-basic edges separated by paths of $\hat{T}_r$ and so we multiply the two bounds.)

### 4.1.6 Putting it all together

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 have that with $C$ equal to the hidden constant in (21),

$$\sum_{r=r_0}^{n} \sum_{k=1}^{r} \mathbb{P}(e_k \text{ is bad } | C) \zeta_{r,k} \leq C\ell_1 \sum_{r=r_0}^{n} \frac{r^{1-13\varepsilon}}{r} + 2 \sum_{r=r_0}^{n} \sum_{k=1}^{r} \gamma_r \varepsilon_r \zeta_{r,k}.$$  

**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 (21) and our bound on the diameter of $DT_r$. The second sum deals with non-basic edges and uses (25).

Now

$$\ell_1 \sum_{r=r_0}^{n} \frac{r^{1-13\varepsilon}}{r} \leq \ell_1 n^{1-11\varepsilon}$$

and

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

Finally, it follows from Lemma 13 and the fact that only edges of cost at most $\gamma_r$ are added that for any $k \leq r$, $\mathbb{P}(e_k \text{ is bad } | C) \leq \ell_1 \gamma_r$. (There are always at most $\ell_1$ choices of edge that could be bad and the probability they have cost at most $\gamma_r$ is $1 - e^{-\gamma_r} \leq \gamma_r$.) So,

$$\sum_{r=r_0}^{n} \sum_{k \in K_r} \mathbb{P}(e_k \text{ is bad } | C) \zeta_{r,k} \leq \ell_1 \sum_{r=r_0}^{n} |K_r| \gamma_r \leq 8\ell_1 \sum_{r=r_0}^{n} r \gamma_r \varepsilon_r \leq \ell_1 n^{1-20\varepsilon}.$$  

After adding the $O(\ell_1 r^\gamma)$ contribution from the edges incident with $a_{r+1}, b_{r+1}$, this completes the justification for (18) and the proof of Lemma 3.

## 5 Final Remarks

We have extended the proof of the validity of Karp’s patching algorithm to random perturbations of dense graphs with minimum in- and out-degree at least $\alpha n$ and independent $EXP(1)$ edge weights. We can extend the analysis to costs with a density function $f(x)$ that satisfies $f(x) = 1 + O(x)$ as $x \to 0$. Janson [25] describes a nice coupling in the case of shortest paths, see Theorem 7 of that paper.
References


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

A Definitions

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

\( G_r \): 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 \( B_r \).

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

\( \Gamma_r^*(u, v) \): This is the graph obtained from \( G_r^*(u, v) \) by contracting the matching edges.

\( \tilde{G}_r \): 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 \( A_r \).

\( \tilde{\Gamma}_r \): This is the digraph obtained from \( \tilde{G}_r(u, v) \) by contracting the matching edges.

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

\( \hat{T}_r \): 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.