Minimum-cost matching in a random graph with random costs

Alan Frieze[*] Tony Johansson[†]
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh PA 15213
U.S.A.

November 8, 2015

Abstract

Let $G_{n,p}$ be the standard Erdős-Rényi-Gilbert random graph and let $G_{n,n,p}$ be the random bipartite graph on $n + n$ vertices, where each $e \in [n]^2$ appears as an edge independently with probability p. For a graph $G = (V, E)$, suppose that each edge $e \in E$ is given an independent uniform exponential rate one cost. Let $C(G)$ denote the random variable equal to the length of the minimum cost perfect matching, assuming that G contains at least one. We show that w.h.p. if $d = np \gg (\log n)^2$ then w.h.p. $\mathbf{E}\left[C(G_{n,n,p})\right] = (1 + o(1))\frac{\pi^2}{6p}$. This generalises the well-known result for the case $G = K_{n,n}$. We also show that w.h.p. $\mathbf{E}\left[C(G_{n,p})\right] = (1 + o(1))\frac{\pi^2}{12p}$ along with concentration results for both types of random graph.

1 Introduction

There are many results concerning the optimal value of combinatorial optimization problems with random costs. Sometimes the costs are associated with n points generated uniformly at random in the unit square $[0, 1]^2$. In which case the most celebrated result is due to Beardwood, Halton and Hammersley [3] who showed that the minimum length of a tour through the points a.s. grew as $\beta n^{1/2}$ for some still unknown β. For more on this and related topics see Steele [19].

The optimisation problem in [3] is defined by the distances between the points. So, it is defined by a random matrix where the entries are highly correlated. There have been many examples considered where the matrix of costs contains independent entries. Aside from the Travelling Salesperson Problem, the most studied problems in combinatorial optimization are perhaps, the shortest path problem; the minimum spanning tree problem and the matching problem. As a first example, consider the shortest path problem in the complete graph K_n where the edge lengths are independent exponential random variables with rate 1. We denote the exponential random variable with rate λ by $E(\lambda)$. Thus $\mathbf{Pr}(E(\lambda) \geq x) = e^{-\lambda x}$ for $x \in \mathbf{R}$. Janson [9] proved (among other things) that if $X_{i,j}$ denotes the shortest distance between vertices i, j in this model then $\mathbf{E}\left[X_{1,2}\right] = \frac{H_n}{n}$ where $H_n = \sum_{i=1}^n \frac{1}{i}$.

As far as the spanning tree problem is concerned, the first relevant result is due to Frieze [6]. He showed that if the edges of the complete graph are given independent uniform $[0, 1]$ edge weights, then the (random) minimum length of a spanning tree L_n satisfies $\mathbf{E}[L_n] \to \zeta(3) = \sum_{i=1}^{\infty} \frac{1}{i^3}$ as $n \to \infty$. Further results on this question can be found in Steele [18], Janson [8], Beveridge, Frieze and McDiarmid [4], Frieze, Ruszinko and Thoma [7] and Cooper, Frieze, Ince, Janson and Spencer [5].

In the case of matchings, the nicest results concern the the minimum cost of a matching in a randomly edge-weighted copy of the complete bipartite graph $K_{n,n}$. If C_n denotes the (random) minimum cost of a perfect matching when edges are given independent exponential $E(1)$ random variables then the story begins with Walkup [20] who proved that $\mathbf{E}[C_n] \le 3$. Later Karp [10] proved that $\mathbf{E}[C_n] \le 2$. Aldous [1], [2] proved that $\lim_{n\to\infty} \mathbf{E}[C_n] = \zeta(2) = \sum_{k=1}^{\infty} \frac{1}{k^2}$. Parisi [13] conjectured that in fact $\mathbf{E}[C_n] = \sum_{k=1}^{n} \frac{1}{k^2}$. This was proved independently by Linusson and Wästlund [11] and by Nair, Prabhakar and Sharma [12]. A short elegant proof was given by Wästlund [16], [17].

In the paper [4] on the minimum spanning tree problem, the complete graph was replaced by a d-regular graph G. Under some mild expansion assumptions, it was shown that if $d \to \infty$ then $\zeta(3)$ can be replaced asymptotically by $n\zeta(3)/d$.

Now consider a d-regular bipartite graph G on $2N$ vertices. Here $d = d(N) \to \infty$ as $N \to \infty$. Each edge e is assigned a cost $w(e)$, each independently chosen according to the exponential distribution $E(1)$. Denote the total cost of the minimum-cost perfect matching by $C(G)$.

We conjecture the following (under some possibly mild restrictions):

Conjecture 1. *Suppose $d = d(N) \to \infty$ as $N \to \infty$. For any d-regular bipartite G,*

$$\boldsymbol{E}[C(G)] = (1 + o(1)) \frac{N}{d} \frac{\pi^2}{6}.$$

Here the $o(1)$ term goes to zero as $N \to \infty$.

In this paper we prove the conjecture for random graphs and random bipartite graphs. Let $G_{n,n,p}$ be the random bipartite graph on $n+n$ vertices, where each $e \in [n]^2$ appears as an edge independently with probability p. Suppose that each edge e is given an independent uniform exponential rate one cost.

Theorem 1. *If $d = np = \omega(\log n)^2$ where $\omega \to \infty$ then w.h.p. $\boldsymbol{E}[C(G_{n,n,p})] \approx \frac{\pi^2}{6p}$.*

Here the statement is that for almost all graphs $G = G_{n,n,p}$ we have $\mathbf{E}[C(G)] \approx \frac{\pi^2}{6p}$. We will in fact show that in this case $C(G)$ will be highly concentrated around $\frac{\pi^2}{6p}$.

Here $A_n \approx B_n$ iff $A_n = (1 + o(1))B_n$ as $n \to \infty$ and the event $\mathcal{E}_n$ occurs with high probability (w.h.p.) if $\mathbf{Pr}(\mathcal{E}_n) = 1 - o(1)$ as $n \to \infty$.

In the case of $G_{n,p}$ we prove

Theorem 2. *If $d = np = \omega(\log n)^2$ where $\omega \to \infty$ then w.h.p. $\boldsymbol{E}[C(G_{n,p})] \approx \frac{\pi^2}{12p}$.*

Applying results of Talagrand [14] we can prove the following concentration result.

Theorem 3. *Let $\varepsilon > 0$ be fixed, then*

$$\mathbf{Pr}\left(\left|C(G_{n,n,p}) - \frac{\pi^2}{6p}\right| \geq \frac{\varepsilon}{p}\right) \leq n^{-K}, \quad \mathbf{Pr}\left(\left|C(G_{n,p}) - \frac{\pi^2}{12p}\right| \geq \frac{\varepsilon}{p}\right) \leq n^{-K}$$

for any constant $K > 0$ and n large enough.

2 Proof of Theorem 1

We find that the proofs in [16], [17] can be adapted to our current situation. Suppose that the vertices of $G = G_{n,n,p}$ are denoted $A = \{a_i, i \in [n]\}$ and $B = \{b_j, j \in [n]\}$. Let $C(n, r)$ denote the cost of the minimum cost matching

$$M_r = \{(a_i, \phi_r(a_i)) : i = 1, 2, \ldots, r\} \text{ of } A_r = \{a_1, a_2, \ldots, a_r\} \text{ into } B.$$

We will prove that w.h.p.

$$\mathbf{E}\left[C(n, r) - C(n, r - 1)\right] \approx \frac{1}{p}\sum_{i=0}^{r-1} \frac{1}{r(n - i)}.$$

for $r = 1, 2, \ldots, n - m$ where

$$m = \left(\frac{n}{\omega^{1/2}\log n}\right).$$

Using this we argue that w.h.p.

$$\mathbf{E}\left[C(G)\right] = \mathbf{E}\left[C(n, n)\right] = \mathbf{E}\left[C(n, n) - C(n, n - m + 1)\right] + \frac{1 + o(1)}{p}\sum_{r=1}^{n-m}\sum_{i=0}^{r-1} \frac{1}{r(n - i)}. \tag{1}$$

We will then show that

$$\sum_{r=1}^{n-m}\sum_{i=0}^{r-1} \frac{1}{r(n - i)} \approx \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}. \tag{2}$$

$$\mathbf{E}\left[C(n, n) - C(n, n - m + 1)\right] = o(p^{-1}) \quad w.h.p. \tag{3}$$

Theorem 1 follows from these two statements.

2.1 Outline of the proof

We first argue (Lemma 1) that $B_r = \phi(A_r)$ is a uniformly random set. This enables us to show (Lemma 2) that w.h.p. vertices $v \in A_r$ have aproximately $(n - r)p$ neighbors in $B \setminus B_r$. Then comes the beautiful idea of adding a vertex b_{n+1} and joining it to every vertex in A by an edge of cost $E(\lambda)$. The heart of the proof is in Lemma 3 that relates $\mathbf{E}\left[C(n, r) - C(n, r - 1)\right]$ in a precise way to the probability that b_{n+1} is covered by M_r^*, the minimum cost matching of A_r into $B^* = B \cup \{b_{n+1}\}$. The proof now focuses on estimating this probability $P(n, r)$. If r is not too close to n then this probability can be estimated (see (6)) by careful conditioning and the use of properties of the exponential distribution. From thereon, it is a matter of analysing the consequences of the estimate for $\mathbf{E}\left[C(n, r) - C(n, r - 1)\right]$ in (7). The final part of the proof involves showing that $\mathbf{E}\left[C(n, n) - C(n - m + 1)\right]$ is insignificant. This essentially boils down to showing that w.h.p. no edge in the minimum cost matching has cost more than $O(\log n/(np))$.

2.2 Proof details

Let $B_r = \{\phi_r(a_i) : i = 1, 2, \ldots, r\}$.

Lemma 1. B_r *is a random* r-*subset of* B.

Proof. Let L denote the $n \times n$ matrix of edge costs, where $L(i, j) = W(a_i, b_j)$ and $L(i, j) = \infty$ if edge (a_i, b_j) does not exist in G. For a permutation π of B let L_π be defined by $L_\pi(i, j) = L(i, \pi(j))$. Let X, Y be two distinct r-subsets of B and let π be any permutation of B that takes X into Y. Then we have

$$\mathbf{Pr}(B_r(L) = X) = \mathbf{Pr}(B_r(L_\pi) = \pi(X)) = \mathbf{Pr}(B_r(L_\pi) = Y) = \mathbf{Pr}(B_r(L) = Y),$$

where the last equality follows from the fact that L and L_π have the same distribution. □

We use the above lemma and the Chernoff bounds to bound degrees. For reference we use the following: Let $B(n, p)$ denote the binomial random variable with parameters n, p. Then for $0 \leq \varepsilon \leq 1$ and $\alpha > 0$,

$$\mathbf{Pr}(B(n, p) \leq (1 - \varepsilon)np) \leq e^{-\varepsilon^2 np/2}. \tag{4}$$

$$\mathbf{Pr}(B(n, p) \leq (1 + \varepsilon)np) \leq e^{-\varepsilon^2 np/3}. \tag{5}$$

$$\mathbf{Pr}(B(n, p) \geq \alpha np) \leq \left(\frac{e}{\alpha}\right)^{\alpha np}.$$

For $v \in A$ let $d_r(v) = |\{w \in B \setminus B_r : (v, w) \in E(G)\}|$. Then we have the following lemma:

Lemma 2.

$$|d_r(v) - (n - r)p| \leq \omega^{-1/5}(n - r)p \ w.h.p. \ for \ v \in A, \ 0 \leq r \leq n - m.$$

Proof. This follows from Lemma 1 i.e. $B \setminus B_r$ is a random set and the Chernoff bounds (4), (5) with $\varepsilon = \omega^{-1/5}$ viz.

$$\mathbf{Pr}(\exists v : |d_r(v) - (n - r)p| \geq \omega^{-1/5}(n - r)p) \leq 2ne^{-\omega^{-2/5}(n-r)p/3}$$
$$\leq 2n^{1 - \omega^{1/10}/3}.$$

□

We can now use the ideas of [16], [17]. We add a special vertex b_{n+1} to B, with edges to all n vertices of A. Each edge adjacent to b_{n+1} is assigned an $E(\lambda)$ cost independently, $\lambda > 0$. We now consider M_r to be a minimum cost matching of A_r into $B^* = B \cup \{b_{n+1}\}$. We denote this matching by M_r^* and we let B_r^* denote the corresponding set of vertices of B^* that are covered by M_r^*.

Define $P(n, r)$ as the normalized probability that v_{n+1} participates in M_r^*, i.e.

$$P(n, r) = \lim_{\lambda \to 0} \frac{1}{\lambda} \mathrm{Pr}\{b_{n+1} \in B_r^*\}.$$

Its importance lies in the following lemma:

Lemma 3.

$$\boldsymbol{E}[C(n,r) - C(n, r-1)] = \frac{1}{r} P(n,r).$$

Proof. Let $X = C(n,r)$ and let $Y = C(n, r-1)$. Fix $i \in [r]$ and let w be the cost of the edge (a_i, b_{n+1}), and let I denote the indicator variable for the event that the cost of the cheapest A_r-assignment that contains this edge is smaller than the cost of the cheapest A_r-assignment that does not use b_{n+1}. In other words, I is the indicator variable for the event $\{Y + w < X\}$.

If $(a_i, b_{n+1}) \in M_r^*$ then $w < X - Y$. Conversely, if $w < X - Y$ and no other edge from b_{n+1} has cost smaller than $X - Y$, then $(a_i, b_{n+1}) \in M_r^*$, and when $\lambda \to 0$, the probability that there are two distinct edges from b_{n+1} of cost smaller than $X - Y$ is of order $O(\lambda^2)$.

Since w is $E(\lambda)$ distributed, as $\lambda \to 0$ we have,

$$\mathbf{E}[X - Y] = \frac{d}{d\lambda} \mathbf{E}[I]\bigg|_{\lambda=0} = \lim_{\lambda \to 0} \frac{1}{\lambda} \Pr\{w < X - Y\} = \frac{1}{r} P(n,r).$$

The factor $1/r$ comes from each $i \in [r]$ being equally likely to be incident to the matching edge containing b_{n+1}, if it exists. $\qquad\square$

We now proceed to estimate $P(n,r)$.

Lemma 4. *Suppose $r < n - m$. Then*

$$\mathbf{Pr}(b_{n+1} \in B_r^* \mid b_{n+1} \notin B_{r-1}^*) = \frac{\lambda}{p(n - r + 1)(1 + \varepsilon_r) + \lambda} \tag{6}$$

where $|\varepsilon_r| \leq \omega^{-1/5}$.

Proof. Assume that $b_{n+1} \notin B_{r-1}^*$. M_r^* is obtained from M_{r-1}^* by finding an augmenting path $P = (a_r, \ldots, a_\sigma, b_\tau)$ from a_r to $B^* \setminus B_{r-1}^*$ of minimum additional cost. Let $\alpha = W(\sigma, \tau)$. We condition on (i) σ, (ii) the lengths of all edges other than $(a_\sigma, b_j), b_j \in B^* \setminus B_{r-1}^*$ and (iii) $\min\{W(\sigma, j) : b_j \in B^* \setminus B_{r-1}^*\} = \alpha$. With this conditioning $M_{r-1} = M_{r-1}^*$ will be fixed and so will $P' = (a_r, \ldots, a_\sigma)$. We can now use the following fact: Let $X_1, X_2, \ldots, X_M$ be independent exponential random variables of rates $\alpha_1, \alpha_2, \ldots, \alpha_M$. Then the probability that X_i is the smallest of them is $\alpha_i/(\alpha_1 + \alpha_2 + \cdots + \alpha_M)$. Furthermore, the probability stays the same if we condition on the value of $\min\{X_1, X_2, \ldots, X_M\}$. Thus

$$\mathbf{Pr}(b_{n+1} \in B_r^* \mid b_{n+1} \notin B_{r-1}^*) = \frac{\lambda}{d_{r-1}(a_\sigma) + \lambda}.$$

$\qquad\square$

Corollary 1.

$$P(n,r) = \frac{1}{p} \left(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{n-r+1} \right)(1 + \varepsilon_r) \tag{7}$$

where $|\varepsilon_r| \leq \omega^{-1/5}$.

Proof. Let $\nu(j) = p(n-j)(1+\varepsilon_j)$, $|\varepsilon_j| \le \omega^{-1/5}$. Then

$$\mathbf{Pr}(b_{n+1} \in B_r^*) = 1 - \frac{\nu(0)}{\nu(0)+\lambda} \cdot \frac{\nu(1)}{\nu(1)+\lambda} \cdots \frac{\nu(r-1)}{\nu(r-1)+\lambda}$$

$$= 1 - \left(1 + \frac{\lambda}{\nu(0)}\right)^{-1} \cdots \left(1 + \frac{\lambda}{\nu(r-1)}\right)^{-1}$$

$$= \left(\frac{1}{\nu(0)} + \frac{1}{\nu(1)} + \cdots + \frac{1}{\nu(r-1)}\right)\lambda + O(\lambda^2)$$

$$= \frac{1}{p}\left(\frac{1}{n(1+\varepsilon_0)} + \frac{1}{(n-1)(1+\varepsilon_1)} + \cdots + \frac{1}{(n-r+1)(1+\varepsilon_{r-1})}\right)\lambda + O(\lambda^2)$$

and each error factor satisfies $|1 - 1/(1+\varepsilon_j)| \le \omega^{-1/5}$. Letting $\lambda \to 0$ gives the lemma. $\square$

Lemma 5. *If $r \le n - m$ then*

$$\boldsymbol{E}[C(n,r) - C(n,r-1)] = \frac{1+o(1)}{rp}\sum_{i=0}^{r-1}\frac{1}{n-i}$$

where $|\varepsilon_k| \le \omega^{-1/5}$.

Proof. This follows from Lemma 3 and Corollary 1. $\square$

This confirms (1) and we turn to (2). We use the following expression from Young [21].

$$\sum_{i=1}^{n}\frac{1}{i} = \log n + \gamma + \frac{1}{2n} + O(n^{-2}), \qquad \text{where } \gamma \text{ is Euler's constant.}$$

Let $m_1 = \omega^{1/4}m$. Observe first that

$$\sum_{i=0}^{m_1}\frac{1}{n-i}\sum_{r=i+1}^{n-m}\frac{1}{r} \le O\left(\frac{\log n}{n^{1/4}}\right) + \sum_{i=n^{3/4}}^{m_1}\frac{1}{n-i}\sum_{r=i+1}^{n-m}\frac{1}{r}$$

$$\le o(1) + \frac{1}{n-m_1}\sum_{i=n^{3/4}}^{m_1}\left(\log\left(\frac{n}{i}\right) + \frac{1}{2(n-m)} + O(n^{-3/2})\right)$$

$$\le o(1) + \frac{2}{n}\log\left(\frac{n^{m_1}}{m_1!}\right)$$

$$\le o(1) + \frac{2m_1}{n}\log\left(\frac{ne}{m_1}\right)$$

$$= o(1).$$

Then,

$$\sum_{r=1}^{n-m}\sum_{i=0}^{r-1}\frac{1}{r(n-i)} = \sum_{i=0}^{n-m-1}\frac{1}{n-i}\sum_{r=i+1}^{n-m}\frac{1}{r},$$

$$= \sum_{i=m_1}^{n-m-1}\frac{1}{n-i}\sum_{r=i+1}^{n-m}\frac{1}{r} + o(1),$$

$$= \sum_{i=m_1}^{n-m-1}\frac{1}{n-i}\left(\log\left(\frac{n-m}{i}\right) + \frac{1}{2(n-m)} - \frac{1}{2i} + O(i^{-2})\right) + o(1),$$

$$= \sum_{i=m_1}^{n-m-1}\frac{1}{n-i}\log\left(\frac{n-m}{i}\right) + o(1),$$

$$= \sum_{j=m+1}^{n-m_1}\frac{1}{j}\log\left(\frac{n-m}{n-j}\right) + o(1), \tag{8}$$

$$= \int_{x=m+1}^{n-m_1}\frac{1}{x}\log\left(\frac{n-m}{n-x}\right)dx + o(1).$$

We can replace the sum in (8) by an integral because the sequence of summands is unimodal and the terms are all $o(1)$.

Continuing, we have

$$\int_{x=m+1}^{n-m_1}\frac{1}{x}\log\left(\frac{n-m}{n-x}\right)dx$$

$$= -\int_{x=m+1}^{n-m_1}\frac{1}{x}\log\left(1 - \frac{x-m}{n-m}\right)dx$$

$$= \sum_{k=1}^{\infty}\int_{x=m+1}^{n-m_1}\frac{1}{x}\frac{(x-m)^k}{k(n-m)^k}dx$$

$$= \int_{y=1}^{n-m-m_1}\frac{1}{y+m}\frac{y^k}{k(n-m)^k}dy. \tag{9}$$

Observe next that

$$\int_{y=1}^{n-m-m_1}\frac{1}{y+m}\frac{y^k}{k(n-m)^k}dy \le \int_{y=1}^{n-m-m_1}\frac{y^{k-1}}{k(n-m)^k}dy \le \frac{1}{k^2}.$$

So,

$$0 \le \sum_{k=\log n}^{\infty}\int_{x=m+1}^{n-m_1}\frac{1}{x}\frac{(x-m)^k}{k(n-m)^k}dx \le \sum_{k=\log n}^{\infty}\frac{1}{k^2} = o(1). \tag{10}$$

If $1 \le k \le \log n$ then we write

$$\int_{y=1}^{n-m-m_1}\frac{1}{y+m}\frac{y^k}{k(n-m)^k}dy = \int_{y=1}^{n-m-m_1}\frac{(y+m)^{k-1}}{k(n-m)^k}dy + \int_{y=1}^{n-m-m_1}\frac{y^k - (y+m)^k}{(y+m)k(n-m)^k}dy.$$

Now

$$\int_{y=1}^{n-m-m_1}\frac{(y+m)^{k-1}}{k(n-m)^k}dy = \frac{1}{k^2}\frac{(n-m_1)^k - (m+1)^k}{(n-m)^k} = \frac{1}{k^2} + O\left(\frac{1}{k\omega^{1/4}\log n}\right). \tag{11}$$

If $k = 1$ then

$$\int_{y=1}^{n-m-m_1} \frac{(y+m)^k - y^k}{(y+m)k(n-m)^k} dy \leq \frac{m \log(n-m_1)}{n-m} = o(1).$$

And if $2 \leq k \leq \log n$ then

$$\int_{y=1}^{n-m-m_1} \frac{(y+m)^k - y^k}{(y+m)k(n-m)^k} dy = \sum_{l=1}^{k} \int_{y=1}^{n-m-m_1} \binom{k}{l} \frac{y^{k-l} m^l}{(y+m)k(n-m)^k} dy$$

$$\leq \sum_{l=1}^{k} \int_{y=0}^{n-m-m_1} \binom{k}{l} \frac{y^{k-l-1} m^l}{k(n-m)^k} dy$$

$$= \sum_{l=1}^{k} \binom{k}{l} \frac{m^l (n-m-m_1)^{k-l}}{k(k-l)(n-m)^k}$$

$$= O\left(\frac{km}{k(k-1)n}\right) = O\left(\frac{1}{k\omega^{1/2} \log n}\right).$$

It follows that

$$0 \leq \sum_{k=1}^{\log n} \int_{y=1}^{n-m-m_1} \frac{(y+m)^k - y^k}{(y+m)k(n-m)^k} dy = o(1) + O\left(\sum_{k=2}^{\log n} \frac{1}{k\omega^{1/2} \log n}\right) = o(1). \qquad (12)$$

Equation (2) now follows from (9), (10), (11) and (12).

Turning to (3) we prove the following lemma:

Lemma 6. *If $r \geq n - m$ then $0 \leq C(n, r+1) - C(n, r) = O\left(\frac{\log n}{np}\right)$.*

This will prove that

$$0 \leq \mathbf{E}\left[C(n, n) - C(n-m+1)\right] = O\left(\frac{m \log n}{np}\right) = O\left(\frac{n}{\omega^{1/2} np}\right) = o\left(\frac{1}{p}\right)$$

and complete the proof of (3) and hence Theorem 1.

2.3 Proof of Lemma 6

Let $w(e)$ denote the weight of edge e in G. Let $V_r = A_{r+1} \cup B$ and let G_r be the subgraph of G induced by V_r. For a vertex $v \in V_r$ order the neighbors $u_1, u_2, \ldots,$ of v in G_r so that $w(v, u_i) \leq w(v, u_{i+1})$. Define the k-neighborhood $N_k(v) = \{u_1, u_2, \ldots, u_k\}$.

Let the k-neighborhood of a set be the union of the k-neighborhoods of its vertices. In particular, for $S \subseteq A_{r+1}, T \subseteq B$,

$$N_k(S) = \{b \in B : \exists a \in S : y \in N_k(a)\},$$
$$N_k(T) = \{a \in A_{r+1} : \exists b \in T : a \in N_k(b)\}.$$

Given a function ϕ defining a matching M of A_r into B, we define the following digraph: let $\vec{\Gamma}_r = (V_r, \vec{X})$ where $\vec{X}$ is an orientation of

$$X =$$
$$\{\{a, b\} \in G : a \in A_{r+1}, b \in N_{40}(a)\} \cup \{\{a, b\} \in G : b \in B, a \in N_{40}(b)\} \cup \{(\phi(a_i), a_i) : i = 1, 2, \ldots, r\}.$$

An edge $e \in M$ is oriented from B to A and has weight $w_r(e) = -w(e)$. The remaining edges are oriented from A to B and have weight equal to their weight in G.

The arcs of directed paths in $\vec{\Gamma}_r$ are alternately forwards $A \to B$ and backwards $B \to A$ and so they correspond to alternating paths with respect to the matching M. It helps to know (Lemma 7, next) that given $a \in A_{r+1}, b \in B$ we can find an alternating path from a to b with $O(\log n)$ edges. The ab-diameter will be the maximum over $a \in A_{r+1}, b \in B$ of the length of a shortest path from a to b.

Lemma 7. *W.h.p., for every ϕ, the (unweighted) ab-diameter of $\vec{\Gamma}_r$ is at most $k_0 = \lceil 3\log_4 n\rceil$.*

Proof. For $S \subseteq A_{r+1}, T \subseteq B$, let

$$N(S) = \{b \in B : \exists a \in S \text{ such that } (a,b) \in \vec{X}\},$$
$$N(T) = \{a \in A_{r+1} : \exists b \in T \text{ such that } (a,b) \in \vec{X}\}.$$

We first prove an expansion property: that **whp**, for all $S \subseteq A_{r+1}$ with $|S| \leq \lceil n/5\rceil$, $|N(S)| \geq 4|S|$. (Note that $N(S), N(T)$ involve edges oriented from A to B and so do not depend on ϕ.)

$$\mathbf{Pr}(\exists S : |S| \leq \lceil n/5\rceil, |N(S)| < 4|S|) \leq o(1) + \sum_{s=1}^{\lceil n/5\rceil} \binom{r+1}{s}\binom{n}{4s}\left(\frac{\binom{4s}{40}}{\binom{n}{40}}\right)^s$$

$$\leq \sum_{s=1}^{\lceil n/5\rceil} \left(\frac{ne}{s}\right)^s \left(\frac{ne}{4s}\right)^{4s}\left(\frac{4s}{n}\right)^{40s}$$

$$= \sum_{s=1}^{\lceil n/5\rceil}\left(\frac{e^5 4^{36} s^{35}}{n^{35}}\right)^s$$

$$= o(1).$$

Explanation: *The $o(1)$ term accounts for the probability that each vertex has at least 40 neighbors in $\vec{\Gamma}_r$. Condition on this. Over all possible ways of choosing s vertices and $4s$ "targets", we take the probability that for each of the s vertices, all 40 out-edges fall among the $4s$ out of the n possibilities.*

Similarly, w.h.p., for all $T \subseteq B$ with $|T| \leq \lceil n/5\rceil$, $|N(T)| \geq 4|T|$. Thus by the union bound, w.h.p. both these events hold. In the remainder of this proof we assume that we are in this "good" case, in which all small sets S and T have large vertex expansion.

Now, choose an arbitrary $a \in A_{r+1}$, and define $S_0, S_1, S_2, \ldots$ as the endpoints of all alternating paths starting from a and of lengths $0, 2, 4, \ldots$. That is,

$$S_0 = \{a\} \text{ and } S_i = \phi^{-1}(N(S_{i-1})).$$

Since we are in the good case, $|S_i| \geq 4|S_{i-1}|$ provided $|S_{i-1}| \leq n/5$, and so there exists a smallest index i_S such that $|S_{i_S-1}| > n/5$, and $i_S - 1 \leq \log_4(n/5) \leq \log_4 n - 1$. Arbitrarily discard vertices from S_{i_S-1} to create a smaller set S'_{i_S-1} with $|S'_{i_S-1}| = \lceil n/5\rceil$, so that $S'_{i_S} = N(S'_{i_S-1})$ has cardinality $|S'_{i_S}| \geq 4|S'_{i_S-1}| \geq 4n/5$.

Similarly, for an arbitrary $b \in B$, define $T_0, T_1, \ldots$, by

$$T_0 = \{b\} \text{ and } T_i = \phi(N(T_{i-1})).$$

9

Again, we will find an index $i_T \leq \log_4 n$ whose modified set has cardinality $|T'_{i_T}| \geq 4n/5$.

With both $|S'_{i_S}|$ and $|T'_{i_T}|$ larger than $n/2$, there must be some $a' \in S'_{i_S}$ for which $b' = \phi(a') \in T'_{i_T}$. This establishes the existence of an alternating walk and hence (removing any cycles) an alternating path of length at most $2(i_S + i_T) \leq 2\log_4 n$ from a to b in $\vec{\Gamma}_r$. $\qquad\square$

We will need the following lemma,

Lemma 8. *Suppose that $k_1 + k_2 + \cdots + k_M \leq a \log N$, and $X_1, X_2, \ldots, X_M$ are independent random variables with Y_i distributed as the k_ith minimum of N independent exponential rate one random variables. If $\mu > 1$ then*

$$\mathbf{Pr}\left(X_1 + \cdots + X_M \geq \frac{\mu a \log N}{N - a \log N}\right) \leq N^{a(1 + \log \mu - \mu)}.$$

Proof. Let $Y_{(k)}$ denote the kth smallest of $Y_1, Y_2, \ldots, Y_N$, where we assume that $k = O(\log N)$. Then the density function $f_k(x)$ of $Y_{(k)}$ is

$$f_k(x) = \binom{N}{k} k (1 - e^{-x})^{k-1} e^{-x(N-k+1)}$$

and hence the ith moment of $Y_{(k)}$ is given by

$$\begin{aligned}
\mathbf{E}\left[Y_{(k)}^i\right] &= \int_0^\infty \binom{N}{k} k x^i (1 - e^{-x})^{k-1} e^{-x(N-k+1)} dx \\
&\leq \int_0^\infty \binom{N}{k} k x^{i+k-1} e^{-x(N-k+1)} dx \\
&= \binom{N}{k} k \frac{(i+k-1)!}{(N-k+1)^{i+k}} \\
&\leq \left(1 + O\left(\frac{k^2}{N}\right)\right) \frac{k(k+1) \cdots (i+k-1)}{(N-k+1)^i}.
\end{aligned}$$

Thus, if $0 \leq t < N - k + 1$,

$$\mathbf{E}\left[e^{tY_{(k)}}\right] \leq \left(1 + O\left(\frac{k^2}{N}\right)\right) \sum_{i=0}^\infty \left(\frac{-t}{N-k+1}\right)^i \binom{-k}{i} = \left(1 + O\left(\frac{k^2}{N}\right)\right) \left(1 - \frac{t}{N-k+1}\right)^{-k}.$$

If $Z = X_1 + X_2 + \cdots + X_M$ then if $0 \leq t < N - a \log N$,

$$\mathbf{E}\left[e^{tZ}\right] = \prod_{i=1}^M \mathbf{E}\left[e^{tX_i}\right] \leq \left(1 - \frac{t}{N - a \log N}\right)^{-a \log N}.$$

It follows that

$$\mathbf{Pr}\left(Z \geq \frac{\mu a \log N}{N - a \log N}\right) \leq \left(1 - \frac{t}{N - a \log N}\right)^{-a \log N} \exp\left\{-\frac{t \mu a \log N}{N - a \log N}\right\}.$$

We put $t = (N - a \log N)(1 - 1/\mu)$ to minimise the above expression, giving

$$\mathbf{Pr}\left(Z \geq \frac{\mu a \log N}{N - a \log N}\right) \leq (\mu e^{1-\mu})^{a \log N}.$$

$\qquad\square$

Lemma 9. *W.h.p., for all ϕ, the weighted ab-diameter of $\vec{\Gamma}_r$ is at most $c_1 \frac{\log n}{np}$ for some absolute contant $c_1 > 0$.*

Proof. Let

$$Z_1 = \max\left\{ \sum_{i=0}^{k} w(x_i, y_i) - \sum_{i=0}^{k-1} w(y_i, x_{i+1}) \right\},$$

where the maximum is over sequences $x_0, y_0, x_1, \ldots, x_k, y_k$ where (x_i, y_i) is one of the 40 shortest arcs leaving x_i for $i = 0, 1, \ldots, k \le k_0 = \lceil 3 \log_4 n \rceil$, and (y_i, x_{i+1}) is a backwards matching edge.

We compute an upper bound on the probability that Z_1 is large. For any $\zeta > 0$ we have

$$\mathbf{Pr}\left(Z_1 \ge \zeta \frac{\log n}{np} \right) \le o(n^{-4}) + \sum_{k=0}^{k_0} ((r+1)n)^{k+1} \left(\frac{1+o(1)}{n} \right)^{k+1} \times p^{k-1} \times$$

$$\int_{y=0}^{\infty} \left[\frac{1}{(k-1)!} \left(\frac{y \log n}{np} \right)^{k-1} \sum_{\rho_0 + \rho_1 + \cdots + \rho_k \le 40(k+1)} q(\rho_0, \rho_1, \ldots, \rho_k; \zeta + y) \right] dy$$

where

$$q(\rho_0, \rho_1, \ldots, \rho_k; \eta) = \mathbf{Pr}\left(X_0 + X_1 + \cdots + X_k \ge \eta \frac{\log n}{np} \right),$$

$X_0, X_1, \ldots, X_k$ are independent and X_j is distributed as the ρ_jth minimum of r independent exponential random variables. (When $k = 0$ there is no term $\frac{1}{(k-1)!} \left(\frac{y \log n}{n} \right)^{k-1}$.)

Explanation: *The $o(n^{-4})$ term is for the probability that there is a vertex in V_r that has fewer than $(1 - o(1))np$ neighbors in V_r. We have at most $((r+1)n)^{k+1}$ choices for the sequence $x_0, y_0, x_1, \ldots, x_k, y_k$. The term $\frac{1}{(k-1)!} \left(\frac{y \log n}{np} \right)^{k-1} dy$ bounds the probability that the sum of k independent exponentials, $w(y_0, x_1) + \cdots + w(y_{k-1}, x_k)$, is in $\frac{\log n}{np}[y, y + dy]$. (The density function for the sum of k independent exponentials is $\frac{x^{k-1} e^{-x}}{(k-1)!}$.) We integrate over y.*
$\frac{(1+o(1))p}{np}$ *is the probability that (x_i, y_i) is and edge of G and is the ρ_ith shortest edge leaving x_i, and these events are independent for $0 \le i \le k$. The factor p^{k-1} is the probability that the B to A edges of the path exist. The final summation bounds the probability that the associated edge lengths sum to at least $\frac{(\zeta + y) \log n}{np}$.*

It follows from Lemma 8 with $a \le 3, N = (1 + o(1))np, \mu = (\zeta + y)/a$ that if ζ is sufficiently large then, for all $y \ge 0$,

$$q(\rho_1, \ldots, \rho_k; \zeta + y) \le (np)^{-(\zeta + y) \log n / (2 \log np)} = n^{-(\zeta + y)/2}.$$

Since the number of choices for $\rho_0, \rho_1, \ldots, \rho_k$ is at most $\binom{41k+40}{k+1}$ (the number of positive integral

solutions to $a_0 + a_1 + \ldots + a_{k+1} \leq 40(k+1)$) we have

$$
\mathbf{Pr}\left(Z_1 \geq \zeta \frac{\log n}{np}\right) \leq o(n^{-4}) + 2n^{2-\zeta/2} \sum_{k=0}^{k_0} \frac{(\log n)^{k-1}}{(k-1)!} \binom{41k+40}{k+1} \int_{y=0}^{\infty} y^{k-1} n^{-y/2} dy
$$

$$
\leq o(n^{-4}) + 2n^{2-\zeta/2} \sum_{k=0}^{k_0} \frac{(\log n)^{k-1}}{(k-1)!} 2^{41k+40} \left(\frac{2}{\log n}\right)^{k-2} \int_{z=0}^{\infty} z^{k-1} e^{-z} dz
$$

$$
= o(n^{-4}) + 2^{41} n^{2-\zeta/2} \log n \sum_{k=0}^{k_0} 2^{41k}
$$

$$
= o(n^{-4}),
$$

for ζ sufficiently large. $\qquad\square$

Lemma 9 shows that with probability $1 - o(n^{-4})$ in going from M_r to M_{r+1} we can find an augmenting path of weight at most $\frac{c_1 \log n}{np}$. This completes the proof of Lemma 6 and Theorem 1. (Note that to go from w.h.p. to expectation we use the fact that w.h.p. $w(e) = O(\log n)$ for all $e \in A \times B$.) $\qquad\square$

Notice also that in the proof of Lemmas 7 and 9 we can certainly make the failure probability less than n^{-K} for any constant $K > 0$.

3 Proof of Theorem 2

Just as the proof method for $K_{n,n}$ in [16], [17] can be modified to apply to $G_{n,n,p}$, the proof for K_n in [15] can be modified to apply to $G_{n,p}$.

3.1 Outline of the proof

This has many similarities with the proof of Theorem 1. The differences are subtle. The first is to let M_r^* be the minimum cost matching of size r after ading a special vertex v_{n+1}. It is again important (Lemma 10) to estimate the probability that $v_{n+1} \in M_r^*$. The approach is similar to that for Theorem 1, except that we now need to prove separate lower and upper bounds for this probability $P(n, r)$.

3.2 Proof details

Consider $G = G_{n,p}$, and denote the vertex set by $V = \{v_1, v_2, \ldots, v_n\}$. Add a special vertex v_{n+1} with $E(\lambda)$-cost edges to all vertices of V, and let G^* be the extended graph on $V^* = V \cup \{v_{n+1}\}$. Say that $v_1, \ldots, v_n$ are *ordinary*. Let M_r^* be the minimum cost r-matching (one of size r) in G^*, unique with probability one. (Note the change in definition.) Define $P(n, r)$ as the normalized probability that v_{n+1} participates in M_r^*, i.e.

$$
P(n, r) = \lim_{\lambda \to 0} \frac{1}{\lambda} \Pr\{v_{n+1} \text{ participates in } M_r^*\}
$$

Let $C(n, r)$ denote the cost of the cheapest r-assignment of G. To estimate $C(n, r)$, we will again need to estimate $P(n, r)$, by the following lemma.

Lemma 10.
$$\boldsymbol{E}[C(n, r) - C(n - 1, r - 1)] = \frac{1}{n} P(n, r)$$

Proof. Let $X = C(n, r)$ and $Y = C(n - 1, r - 1)$. Fix $i \in [n]$ and let w be the cost of the edge (v_i, v_{n+1}), and let I denote the indicator variable for the event that the cost of the cheapest r-assignment that contains this edge is smaller than the cost of the cheapest r-assignment that does not use v_{n+1}. The rest of the proof is identical to the proof of Lemma 3, except that there are now n choices for i as opposed to r in the previous lemma. $\square$

In this case, unlike the bipartite case, we are unable to directly find an asymptotic expression for $P(n, r)$, as we did in Lemma 6 and Corollary 1. Here we will have to turn to bounding $P(n, r)$ from below and above.

3.3 A lower bound for $P(n, r)$

We will consider an algorithm that finds a set $A_s \subseteq V^*$ which contains the set B_s of vertices participating in M_r^*, $s = |A_s| \geq |B_s| = 2r$. Call A_s the set of *exposed* vertices.

Initially let $A_s = B_s = \emptyset$ and $r = s = 0$. At stage s of the algorithm we condition on

> A_s, B_s and the existence and cost of all edges within A_s.
> In particular, we condition on r and the minimum r-matching M_r^*.

Given a minimum matching M_r^*, we decide how to build a proposed $(r+1)$-matching by comparing the following numbers and picking the smallest.

(a) z_a equals the cost of the cheapest edge between a pair of unexposed vertices.

(b) $z_b = \min\{c_1(v) : v \in A_s \setminus B_s\}$, where $c_1(v)$ is the cost of the cheapest edge between v and a vertex $\tau_1(v) \notin A_s$.

(c) $z_c = \min\{c_1(v) + c_1(u) + \delta(u, v) : u, v \in B_s\}$ where $\delta(u, v)$ denotes the cost of the cheapest alternating path from u to v, with the cost of edges in M_r^* taken as the negative of the actual value.

Let
$$z_{\min} = \min\{z_a, z_b, z_c\}.$$

If $z_{\min} = z_a$ then we reveal the edge $\{v, w\}$ and add it to M_r^* to form M_{r+1}^*. Once v, w have been determined, they are added to A_s and B_s, and we move to the next stage of the algorithm, updating $s \leftarrow s + 2, r \leftarrow r + 1$.

If $z_{\min} = z_b$ then let $v \in A_s \setminus B_s$ be the vertex with the cheapest $c_1(v)$. We reveal $w = \tau_1(v)$ and add w to A_s and to B_s while adding v to B_s. Now $M_{r+1}^* = M_r^* \cup \{v, w\}$. We move to the next stage of the algorithm, updating $s \leftarrow s + 1, r \leftarrow r + 1$.

If $z_{\min} = z_c$ then reveal $w_1 = \tau_1(u), w_2 = \tau_1(v)$. If $w_1 = w_2$, we say that we have a *collision*. In this case, the vertex w_1 is added to A_s (but not B_s), and we move to the next stage with $s \leftarrow s + 1$. If there is no collision, we update M_r^* by the augmenting path $w_1, u, \ldots, v, w_2$ to form M_{r+1}^*. We add w_1, w_2 to A_s and B_s, and move on to the next stage with $s \leftarrow s + 2$ and $r \leftarrow r + 1$.

It follows that $A_s \setminus B_s$ consists of unmatched vertices that have been the subject of a collision.

It will be helpful to define A_s for all s, so in the cases where two vertices are added to A_s, we add them sequentially with a coin toss to decide the order.

The possibility of a collision is the reason that not all vertices of A_s participate in M_r^*. However, the probability of a collision at v_{n+1} is $O(\lambda^2)$, and as $\lambda \to 0$ this is negligible. In other words, as $\lambda \to 0$,

$$\Pr\{v_{n+1} \in M_r^*\} \geq \Pr\{v_{n+1} \in B_{2r}\} = \Pr\{v_{n+1} \in A_{2r}\} - O(\lambda^2)$$

and we will bound $\Pr\{v_{n+1} \in A_{2r}\}$ from below.

Lemma 11. *Conditioning on $v_{n+1} \notin A_s$, A_s is a random s-subset of V.*

Proof. Trivial for $s = 0$. Suppose A_{s-1} is a random $(s-1)$-subset of V. Define $N_s(v) = \{w \notin A_s : (v, w) \in E\}$. In stage s, if we condition on $d_s(v) = |N_s(v)|$, then under this conditioning $N_s(v)$ is a random $d_s(v)$-subset of $V \setminus A_s$. This is because the constructon of A_s does not require the edges from A_s to $V \setminus A_s$ to be exposed. So, if $A_s \setminus A_{s-1} = \{w\}$ where w is added due to being the cheapest unexposed neighbor of an exposed v, then w is a random element of $N_s(v)$ and hence a random element of $V \setminus A_s$.

If we are in case (a), i.e. M_{r+1}^* is formed by adding an edge between two ordinary unexposed vertices v, w, then since we only condition on the size of the set $\{(v, w) : v, w \notin A_s\}$, all pairs $v, w \in V \setminus A_s$ are equally likely, and after a coin toss this can be seen as adding two random elements sequentially. We conclude that A_s is a random s-subset of V. $\square$

Recall that $m = n/(\omega^{1/2} \log n)$.

Corollary 2. *W.h.p., for all $0 \leq s \leq n - m$ and all $v \in V$,*

$$|d_s(v) - (n - s)p| \leq \omega^{-1/5}(n - s)p.$$

Proof. This follows from the Chernoff bounds as in Lemma 2. $\square$

We now bound the probability that $A_s \setminus A_{s-1} = \{v_{n+1}\}$ from below. There are a few different ways this may happen.

We now have to address some cost conditioning issues. Suppose that we have just completed an iteration. First consider the edges between vertices not in A_s. For such an edge e, all we know is $w(e) \geq \zeta$ where $\zeta = z_{\min}$ of the just completed iteration. So the conditional cost of such an edge can be expressed as $\zeta + E(1)$ or $\zeta + E(\lambda)$ in the case where e is incident with v_{n+1}. The exponentials are independent. We only need to compare the exponential parts of each edge cost here to decide the probability that an edge incident with v_{n+1} is chosen.

We can now consider case (a). Suppose that an edge $\{u, v\}$ between unexposed vertices is added to A_{s-1}. By Corollary 2, there are at most $p\binom{n-s+1}{2}(1 + \omega^{-1/5})$ ordinary such edges. There are

14

$n - s$ edges between v_{n+1} and $V \setminus A_s$, each at rate λ. As $\lambda \to 0$, the probability that one of the endpoints of the edge chosen in case (a) is v_{n+1} is therefore at least

$$\frac{\lambda(n - s)}{\lambda(n - s) + p\binom{n-s+1}{2}(1 + \omega^{-1/5})} \geq \frac{1}{p}\frac{2\lambda}{n - s}(1 - \omega^{-1/5}) + O(\lambda^2)$$

We toss a fair coin to decide which vertex in the edge $\{u, v\}$ goes in A_s. Hence the probability that $A_s \setminus A_{s-1} = \{v_{n+1}\}$ in case (a) is at least

$$\frac{1}{p}\frac{\lambda}{n - s}(1 - \omega^{-1/5}) + O(\lambda^2).$$

We may also have $A_s \setminus A_{s-1} = \{v_{n+1}\}$ if case (a) occurs at stage $s - 2$ and v_{n+1} loses the coin toss, in which case the probability is at least

$$\frac{1}{p}\frac{\lambda}{n - s + 1}(1 - \omega^{-1/5}) + O(\lambda^2).$$

Now consider case (b). Here only one vertex is added to A_{s-1}, the cheapest unexposed neighbor w of some $v \in A_{s-1} \setminus B_{s-1}$. The cost conditioning here is the same as for case (a), i.e. that the cost of an edge is $\zeta + E(1)$ or $\zeta + E(\lambda)$. By Corollary 2, this v has at most $p(n - s + 1)(1 + \omega^{-1/5})$ ordinary unexposed neighbors, so the probability that $w = v_{n+1}$ is at least

$$\frac{\lambda}{p(n - s + 1)(1 + \omega^{-1/5}) + \lambda} = \frac{1}{p}\frac{\lambda}{n - s + 1}(1 - \omega^{-1/5}) + O(\lambda^2).$$

Finally, consider case (c). To handle the cost conditioning, we condition on the values $c_1(v)$ for $v \in B_s$. By well-known properties of independent exponential variables, the minimum is located with probability proportional to the rates of the corresponding exponential variables. A collision at v_{n+1} has probability $O(\lambda^2)$, so assume we are in the case of two distinct unexposed vertices w_1, w_2. Suppose that w_1 is revealed first. Exactly as in (b), the probability that $w_1 = v_{n+1}$ is at least

$$\frac{1}{p}\frac{\lambda}{n - s + 1}(1 - \omega^{-1/5}) + O(\lambda^2).$$

If $w_1 \neq v_{n+1}$, the probability that $w_2 = v_{n+1}$ (i.e. $A_{s+1} \setminus A_s = \{v_{n+1}\}$) is at least

$$\frac{1}{p}\frac{\lambda}{n - s}(1 - \omega^{-1/5}) + O(\lambda^2),$$

so by considering the possibility that v_{n+1} is the second vertex added from A_{s-2}, we again have probability at least

$$\frac{1}{p}\frac{\lambda}{n - s + 1}(1 - \omega^{-1/5}) + O(\lambda^2).$$

We conclude that no matter which case occurs, the probability is at least

$$\frac{1}{p}\frac{\lambda}{n - s + 1}(1 - \omega^{-1/5}) + O(\lambda^2).$$

So

$$P(n, r) \geq \lim_{\lambda \to 0}\frac{1}{\lambda}\sum_{s=1}^{2r}\Pr\{A_s \setminus A_{s-1} = \{v_{n+1}\}\} \geq \frac{1 - \omega^{-1/5}}{p}\sum_{s=1}^{2r}\frac{1}{n - s + 1}.$$

Write

$$L(n, r) = \sum_{s=1}^{2r}\frac{1}{n - s + 1}.$$

15

3.4 An upper bound for $P(n, r)$

We now alter the algorithm above in such a way that $A_{2r} = B_{2r}$. We do not consider A_s for odd s here. At a stage with $s = 2r$, we condition on

> A_s, and the appearance and cost of all edges within A_s.
> In particular, we condition on r and the minimum r-matching M_r^*.   A

set $C_s \subseteq A_s$, where each $v \in C_s$ has been involved in a collision.

This changes how we calculate a candidate for M_{r+1}^*. We now take the minimum of

(a) z_a equals the cost of the cheapest edge between unexposed vertices.

(b) $z_b = \min\{c_1(u) + c_1(v) + \delta(u, v) : u, v \in A_s, |\{u, v\} \cap C_s| \leq 1\}$, where c_1 and δ are as defined in Section 3.3.

(c) $z_c = \min\{c_1(u) + c_2(v) + \delta(u, v) : u, v \in C_s, \tau_1(u) = \tau_1(v)\}$, where τ_1 is defined in Section 3.3 and $c_2(v)$ is the cost of the second cheapest edge between v and a vertex $\tau_2(v) \notin A_s$.

Let

$$z_{\min} = \min\{z_a, z_b, z_c\}.$$

If $z_{\min} = z_a$ then we reveal the edge $\{v, w\}$ and add it to M_r^* to form M_{r+1}^*. Once v, w have been determined, they are added to A_s and we move to the next stage of the algorithm, updating $s \leftarrow s + 2$.

If $z_{\min} = z_b$ then reveal $w_1 = \tau_1(u), w_2 = \tau_1(v)$. If $w_1 = w_2$ then we add u, v to C_s and go to the next stage of the algorithm without changing s. (The probability that $\tau_1(u) = \tau_1(v) = v_{n+1}$ is $O(\lambda^2)$, and we can safely ignore this as $\lambda \to 0$). If at some later stage w_1 is added to A_s and u say is still in A_s then we remove u from C_s. If $w_1 \neq w_2$ then we update M_r^* by the augmenting path $w_1, u, \ldots, v, w_2$ to form M_{r+1}^*. We add w_1, w_2 to A_s, and move on to the next stage with $s \leftarrow s + 2$.

If $z_{\min} = z_c$ then we update M_r^* by the augmenting path $w_1 = \tau_1(u), u, \ldots, v, w_2 = \tau_2(v)$ to form M_{r+1}^*. We add w_1, w_2 to A_s, and move on to the next stage with $s \leftarrow s + 2$.

Eventually we will construct M_{r+1}^* since case (b) with $\tau_1(u) = \tau_1(v)$ can happen at most s times before $C_s = A_s$.

The cost conditioning is the same as we that for computing the lower bound in Section 3.3, except for the need to deal with $c_2(v), v \in C_s$. For this we condition on $c_2(v)$ and argue that the probability $\delta_2(v) = x$ is proportional to the exponential rate for the edge (v, x). At this point we know that $\delta_1(v) \neq v_{n+1}$, since we are assuming λ is so small that this possibility can be ignored. So, in this case, we can only add v_{n+1} as $\delta_2(v)$ for some $v \in C_s$.

To analyze this algorithm we again need to show that A_{2r} is a uniformly random subset of V.

Lemma 12. *Conditioning on $v_{n+1} \notin A_{2r}$, A_{2r} is a random $2r$-subset of V.*

Proof. Let L denote the $n \times n$ matrix of edge costs, where $L(i, j) = w(v_i, v_j)$ and $L(i, j) = \infty$ if edge (v_i, v_j) does not exist in G. For a permutation π of V let L_π be defined by $L_\pi(i, j) = L(\pi(i), \pi(j))$.

Let X, Y be two distinct $2r$-subsets of V and let π be any permutation of V that takes X into Y. Then we have

$$\mathbf{Pr}(A_{2r}(L) = X) = \mathbf{Pr}(A_{2r}(L_\pi) = \pi(X)) = \mathbf{Pr}(A_{2r}(L_\pi) = Y) = \mathbf{Pr}(A_{2r}(L) = Y),$$

where the last equality follows from the fact that L and L_π have the same distribution. This shows that A_{2r} is a random $2r$-subset of V. $\qquad\square$

Let $d_{2r}(v) = |\{w \notin A_{2r} : (v, w) \in E\}|$.

Corollary 3. *W.h.p., for all $0 \leq r \leq (n - m)/2$ and all $v \in V$,*

$$|d_{2r}(v) - p(n - 2r)| \leq \omega^{-1/5} p(n - 2r).$$

Proof. The proof is again via Chernoff bounds, see Lemma 2. $\qquad\square$

We bound the probability that $v_{n+1} \in A_{2r} \setminus A_{2r-2}$ from above. Suppose we are at a stage where a collisionless candidate for M_r^* has been found.

In case (a), as in the previous section the probability that v_{n+1} is one of the two unexposed vertices is at most

$$\frac{\lambda(n - 2r + 2)}{\lambda(n - 2r + 2) + p\binom{n-2r+2}{2}(1 - \omega^{-1/5})} = \frac{1}{p}\frac{2\lambda}{n - 2r + 1}(1 + \omega^{-1/5}) + O(\lambda^2)$$

Now suppose we are in case (b) with $u, v \notin C_{2r-2}$. If no collision occurs, the probability that one of $\tau_1(u), \tau_1(v)$ is v_{n+1} is at most

$$\frac{\lambda}{\lambda + d_s(u)} + \frac{\lambda}{\lambda + d_s(v) - 1} \leq \frac{1}{p}\frac{2\lambda}{n - 2r + 1}(1 + \omega^{-1/5}) + O(\lambda^2)$$

Finally, if we find M_{r+1}^* by alternating paths where one exposed vertex uses its second-cheapest edge to an unexposed vertex, the probability of that vertex being v_{n+1} is even smaller at $\lambda/(n - 2r + 1)$. So,

$$P(n, r) = \lim_{\lambda \to 0} \sum_{s=1}^{r} \mathrm{Pr}\left\{v_{n+1} \in A_{2s} \setminus A_{2s-2}\right\} \leq \frac{2(1 + \omega^{-1/5})}{p} \sum_{s=1}^{r} \frac{1}{n - 2s + 1}$$

Write

$$U(n, r) = \sum_{s=1}^{r} \frac{2}{n - 2s + 1}.$$

17

3.5 Calculating $\mathbf{E}\left[C(G_{n,p})\right]$

From Lemma 10 and (3.3) we have

$$\mathbf{E}\left[C(n,(n-m)/2)\right]$$

$$= \sum_{r=1}^{(n-m)/2} \frac{1}{n-r+1} P(n-r+1,(n-m)/2-r+1)$$

$$\geq \frac{1+o(1)}{p} \sum_{r=1}^{(n-m)/2} \frac{1}{n-r+1} L(n-r+1,(n-m)/2-r+1)$$

$$= \frac{1+o(1)}{p} \sum_{r=1}^{(n-m)/2} \frac{1}{n-r+1} \sum_{s=1}^{n-m-2r+2} \frac{1}{(n-r+1)-s+1}$$

$$= \frac{1+o(1)}{p} \sum_{r=1}^{(n-m)/2} \frac{1}{n-r+1} \left(\log\left(\frac{n-r+1}{m+r}\right) + \frac{1}{2(n-r)} - \frac{1}{2(m+r)} + O(m^{-2}) \right), \quad (13)$$

by (2.2).

The correction terms are easily taken care of. First we have

$$\left| \sum_{r=1}^{(n-m)/2} \frac{1}{n-r+1} \left(\frac{1}{2(n-r)} - \frac{1}{2(m+r)} + O(m^{-2}) \right) \right|$$

$$= O\left(\frac{1}{m} \sum_{r=1}^{(n-m)/2} \frac{1}{n-r+1} \right)$$

$$= O\left(\frac{n-m}{mn} \right)$$

$$= o(1).$$

Now we want to replace the $(m+r)$ term in the logarithm in the RHS of (13) by r. For this we let $m_1 = n/(\omega^{1/4} \log n) = m\omega^{1/4}$. Then

$$\left| \sum_{r=1}^{(n-m)/2} \frac{1}{n-r+1} \log\left(\frac{r}{m+r}\right) \right|$$

$$= \sum_{r=1}^{m_1-1} \frac{1}{n-r+1} \log\left(1+\frac{m}{r}\right) + \sum_{r=m_1}^{(n-m)/2} \frac{1}{n-r+1} \log\left(1+\frac{m}{r}\right)$$

$$\leq \log m \sum_{r=1}^{m_1-1} \frac{1}{n-r+1} + \log\left(1+\frac{m}{m_1}\right) \sum_{r=m_1}^{(n-m)/2} \frac{1}{n-r+1}$$

$$\leq \log n \frac{m_1}{n-m_1} + \frac{m}{m_1} \frac{(n-m)/2}{n/2}$$

$$= o(1).$$

So,

$$p \times RHS((13)) = o(1) + \sum_{r=1}^{(n-m)/2} \frac{1}{n-r+1} \log\left(\frac{n-r}{r}\right)$$

$$= o(1) + \int_0^{1/2} \frac{1}{1-\alpha} \log\left(\frac{1-\alpha}{\alpha}\right) d\alpha.$$

Substituting $y = \log(1/\alpha - 1)$ we have

$$\int_0^{1/2} \frac{1}{1-\alpha} \log\left(\frac{1-\alpha}{\alpha}\right) d\alpha = \int_0^\infty \frac{y}{e^y + 1} dy$$

$$= \int_0^\infty \frac{ye^{-y}}{1+e^{-y}} dy$$

$$= \sum_{j=0}^\infty \int_0^\infty ye^{-y}(-e^{-y})^j dy$$

$$= \sum_{j=1}^\infty (-1)^{j+1} \frac{1}{j^2}$$

$$= \frac{1}{2} \sum_{k=1}^\infty \frac{1}{k^2}$$

$$= \frac{\pi^2}{12}.$$

This proves a lower bound for $\mathbf{E}\left[C(n, (n-m)/2)\right]$. It also shows that

$$\sum_{r=1}^{(n-m)/2} \frac{1}{n-r+1} L(n-r+1, (n-m)/2 - r + 1) = (1+o(1)) \frac{\pi^2}{12}. \qquad (14)$$

For an upper bound, note that for $r \le (n-m)/2$,

$$U(n,r) - L(n,r) = \sum_{s=1}^r \left(\frac{1}{n-2s+1} - \frac{1}{n-2s+2}\right)$$

$$= \sum_{s=1}^r \frac{1}{(n-2s+1)(n-2s+2)}$$

$$= O\left(\frac{r}{(n-2r)^2}\right)$$

19

So,

$$\mathbf{E}\left[C(n,(n-m)/2)\right]$$

$$\leq \frac{1+o(1)}{p} \sum_{r=1}^{(n-m)/2} \frac{1}{n-r+1} U(n-r+1,(n-m)/2-r+1)$$

$$= \frac{1+o(1)}{p} \sum_{r=1}^{(n-m)/2} \frac{1}{n-r+1} \left(L(n-r+1,(n-m)/2-r+1) + O\left(\frac{r}{(n-2r)^2}\right) \right)$$

$$= \frac{1+o(1)}{p} \left(o(1) + \sum_{r=1}^{(n-m)/2} \frac{1}{n-r+1} \left(L(n-r+1,(n-m)/2-r+1) \right) \right) \tag{15}$$

$$= \frac{\pi^2}{12p}(1+o(1)). \tag{16}$$

To get from (15) to (16) we use (14).

We show that for n even, $\mathbf{E}\left[C(n,n/2) - C(n,(n-m)/2)\right] = o(1/p)$ to conclude that

$$\mathbf{E}\left[C(G_{n,p})\right] = \mathbf{E}\left[C(n,n/2) - C(n,(n-m)/2)\right] + \mathbf{E}\left[C(n,(n-m)/2)\right] = \frac{\pi^2}{12p}(1+o(1)).$$

As above, this will follow from the following lemma.

Lemma 13. *Suppose n is even. If $(n-m)/2 \leq r \leq n/2$ then $0 \leq \mathbf{E}\left[C(n,r+1) - C(n,r)\right] = O\left(\frac{\log n}{np}\right)$.*

3.6 Proof of Lemma 13

This section will replace Section 2.3. Let $M = \{v, \phi(v), v \in [n]\}, \phi^2(v) = v$ for all $v \in [n]$ be an arbitrary perfect matching of $G_{n,p}$. We let $\vec{M} = \{(u,v) : v = \phi(u)\}$ consist of two oppositely oriented copies of each edge of M. We then randomly orient the edges of $G_{n,p}$ that are not in M and then add $\vec{M}$ to obtain the digraph $\vec{G} = \vec{G}_{n,p}$. Because $np = \omega(\log n)^2$ we have that w.h.p. the minimum in- or out-degree in $\vec{G}_{n,p}$ is at least $\omega(\log n)^2/3$. Let $\mathcal{D}$ be the event that all in- and out-degrees are at least this large. Let the *M-alternating diameter* of $\vec{G}$ be the maximum over pairs of vertices $u \neq v$ of the minimum length of an odd length M-alternating path w.r.t. M between u and v where (i) the edges are oriented along the path in the direction u to v, (ii) the first and last edges are not in M. Given this orientation, we define $\vec{\Gamma}_r$ to be the subdigraph of $\vec{G}$ consisting of the r cheapest non-$\vec{M}$ out-edges from each vertex together with $\vec{M}$. Once we can show that the M-alternating diameter of $\vec{\Gamma}_{20}$ is at most $\lceil 3 \log_3 n \rceil$, the proof follows the proof of Lemma 9 more or less exactly.

Lemma 14. *W.h.p., the alternating diameter of $\vec{\Gamma}_{20}$ is at most $k_0 = \lceil 3 \log_3 n \rceil$.*

Proof. We first consider the relatively simple case where $np \geq n^{1/3} \log n$. Let $N^+(u)$ be the set of out-neighbors of u in $\vec{G}$ and let $N^-(v)$ be the set of in-neighbors of v in $\vec{G}$. If there is an edge of $\vec{M}$ from $N^+(u)$ to $N^-(v)$ then this creates an alternating path of length three. Otherwise, let $N^{++}(u)$ be the other endpoints of the matching edges incident with $N^+(u)$ and define $N^{--}(v)$ analogously.

Note that now we have $N^{++}(u) \cap N^{--}(v) = \emptyset$ and given $\mathcal{D}$, the conditional probability that there is no edge from $N^{++}(u)$ to $N^{--}(v)$ in $\vec{G}$ is at most $(1 + o(1))(1 - p)^{(np/3)^2} \leq e^{-(\log n)^3/10} = o(n^{-2})$. Thus in this case there will be an alternating path of length five.

Now assume that $np < n^{1/3} \log n$. In which case we can prove

$$\text{W.h.p. } |S| \leq n^{7/12} \text{ implies that } e(S) \leq 6|S|. \tag{17}$$

This follows from

$$\begin{aligned}
\mathbf{Pr}(\exists S \text{ violating (17)}) &\leq \sum_{s=13}^{n^{7/12}} \binom{n}{s} \binom{\binom{s}{2}}{6s} p^{6s} \\
&\leq \sum_{s=13}^{n^{7/12}} \left(\frac{ne}{s}\right)^s \left(\frac{sep}{12}\right)^{6s} \\
&\leq \sum_{s=13}^{n^{7/12}} \left(\frac{e^7 s^5 (\log n)^6}{12^6 n^3}\right)^s \\
&= o(1).
\end{aligned}$$

Imitating Lemma 7, we prove an expansion property for $\vec{\Gamma}_{20}$:

$$\begin{aligned}
\mathbf{Pr}(\exists S : |S| \leq n^{2/3}, |N_{20}(S)| < 10|S|) &\leq o(1) + \sum_{s=1}^{n^{2/3}} \binom{n}{s} \binom{n-s}{10s} \left(\frac{\binom{11s}{20}}{\binom{n}{20}}\right)^s \\
&\leq \sum_{s=1}^{n^{2/3}} \left(\frac{ne}{s}\right)^s \left(\frac{ne}{10s}\right)^{10s} \left(\frac{11s}{n}\right)^{20s} \\
&= \sum_{s=1}^{n^{2/3}} \left(\frac{e^{11} 11^{20} s^9}{10^{10} n^9}\right)^s \\
&= o(1). \tag{18}
\end{aligned}$$

Fix an arbitrary pair of vertices a, b. Define $S_i, i = 0, 1, \ldots$ to be the set of vertices v such that there exists a directed M-alternating path of length $2i$ in $\vec{\Gamma}_{20}$ from a to v. We let $S_0 = \{a\}$ and given S_i we let $S_i' = N^+(S_i) \setminus \{b\}$ and $S_i'' = \{w \neq b : \exists \{v, w\} \in M : v \in S_i'\}$. Here $N^+(S) = \{w \notin S : \exists v \in S, (v, w) \in E(\vec{\Gamma})\}$ is the set of out-neighbors of S. $N^-(S)$ is similarly defined as the set of in-neighbors of S. It follows from (17) and (18) that w.h.p. $|S_i''| \geq 3|S|$, so long as $|S_i| = o(n^{7/12})$. We therefore let S_{i+1} be a subset of S_i'' of size $3|S_i|$. So w.h.p. there exists an $i_a \leq \log_3 n$ such that $|S_{i_a}| \in [n^{13/24}, 3n^{13/24}]$.

Repeat the procedure with vertex b, letting $T_0 = \{b\}, T_{j+1}' = N^-(T_j)$ etc. By the same argument, there exists an $j_b \leq \log_3 n$ such that T_{j_b} is of size in $[n^{13/24}, 3n^{13/24}]$. Finally, the probability that there is no $S_{i_a} \to T_{j_b}$ edge is at most $(1 - p)^{n^{13/12}} = o(n^{-2})$. This completes the proof of Lemma 14. $\qquad \square$

The remainder of the proof of Lemma 13 is now exactly as in Section 2.3. This concludes the proof of Theorem 2.

4 Proof of Theorem 3

The proof of Lemma 9 allows us to claim that for any constant $K > 0$, with probability $1 - O(n^{-K})$ the maximum length of an edge in the minimum cost perfect matching of G is at most $\mu = c_2 \frac{\log n}{np}$ for some constant $c_2 = c_2(K) > 0$. We can now proceed as in Talagrand's proof of concentration for the assignment problem. We let $\widehat{w}(e) = \min\{w(e), \mu\}$ and let $\widehat{C}(G)$ be the assignment cost using $\widehat{w}$ in place of w. We observe that

$$\mathbf{Pr}(\widehat{C}(G) \neq C(G)) = O(n^{-K}) \tag{19}$$

and so it is enough to prove concentration of $\widehat{C}(G)$.

For this we use the following result of Talagrand [14]: consider a family $\mathcal{F}$ of N-tuples $\boldsymbol{\alpha} = (\alpha_i)_{i \leq N}$ of non-negative real numbers. Let

$$Z = \min_{\boldsymbol{\alpha} \in \mathcal{F}} \sum_{i \leq N} \alpha_i X_i$$

where $X_1, X_2, \ldots, X_N$ are an independent sequence of random variables taking values in $[0, 1]$.

Let $\sigma = \max_{\boldsymbol{\alpha} \in \mathcal{F}} \|\boldsymbol{\alpha}\|_2$. Then if M is the median of Z and $u > 0$, we have

$$\mathbf{Pr}(|Z - M| \geq u) \leq 4 \exp\left\{-\frac{u^2}{4\sigma^2}\right\}. \tag{20}$$

We apply (20) with $N = n^2$ and $X_e = \widehat{w}(e)/\mu$. For $\mathcal{F}$ we take the $n!$ $\{0, 1\}$ vectors corresponding to perfect matchings and scale them by μ. In this way, $\sum_e \alpha_e X_e$ will be the weight of a perfect matching. In this case we have $\sigma^2 \leq n\mu^2$. Applying (20) we obtain

$$\mathbf{Pr}\left(|\widehat{C}(G) - \widehat{M}| \geq \frac{\varepsilon}{p}\right) \leq 4 \exp\left\{-\frac{\varepsilon^2}{4p^2} \cdot \frac{1}{n\mu^2}\right\} = \exp\left\{-\frac{\varepsilon^2 n}{(c_2 \log n)^2}\right\}, \tag{21}$$

where $\widehat{M}$ is the median of $\widehat{C}(G)$. Theorem 2 follows easily from (19) and (21).

5 Final remarks

We have generalised the result of [2] to the random bipartite graph $G_{n,n,p}$ and the result of [15] to the random graph $G_{n,p}$. It would be of some interest to extend the result in some way to random regular graphs. In the absence of proving Conjecture 1 we could maybe extend the results of [2], [15] to some special class of special graphs e.g. to the hypercube.

References

[1] D. Aldous, Asymptotics in the random assignment problem, *Probability Theory and Related Fields* 93 (1992) 507-534.

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

[3] J. Beardwood, J. H. Halton and J. M. Hammersley, The shortest path through many points, *Mathematical Proceedings of the Cambridge Philosophical Society* 55 (1959) 299-327.

[4] A. Beveridge, A.M. Frieze and C.McDiarmid, Random minimum length spanning trees in regular graphs, *Combinatorica* 18 (1998) 311-333.

[5] C. Cooper, A.M. Frieze, N. Ince, S. Janson and J. Spencer, On the length of a random minimum spanning tree, see arxiv.org.

[6] A.M. Frieze, On the value of a random minimum spanning tree problem, *Discrete Applied Mathematics* 10 (1985) 47-56.

[7] A.M. Frieze, M. Ruszinko and L. Thoma, A note on random minimum length spanning trees, *Electronic Journal of Combinatorics* 7 (2000) R41.

[8] S. Janson, The minimal spanning tree in a complete graph and a functional limit theorem for trees in a random graph, *Random Structures Algorithms* 7 (1995) 337-355.

[9] 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.

[10] R.M. Karp, An upper bound on the expected cost of an optimal assignment, *Discrete Algorithms and Complexity: Proceedings of the Japan-US Joint Seminar (D. Johnson et al., eds.)*, Academic Press, New York, 1987, 1-4.

[11] S. Linusson and J. Wästlund, A proof of Parisi's conjecture on the random assignment problem, *Probability Theory and Related Fields* 128 (2004) 419-440.

[12] C. Nair, B. Prabhakar and M. Sharma, Proofs of the Parisi and Coppersmith-Sorkin random assignment conjectures, *Random Structures and Algorithms* 27 (2005) 413-444.

[13] G. Parisi, A conjecture on Random Bipartite Matching, Pysics e-Print archive (1998).

[14] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, *Publications Mathmatiques de L'IHÉS* (1995).

[15] J. Wästlund, Random matching problems on the complete graph, *Electronic Communications in Probability* (2008) 258-265.

[16] J. Wästlund, A simple proof of the Parisi and Coppersmith-Sorkin formulas for the random assignment problem, *Linköping Studies in Mathematics* 6 (2005).

[17] J. Wästlund, An easy proof of the $\zeta(2)$ limit in the random assignment problem, *Elec. Comm. in Probab.* **14** (2009) 261 – 269

[18] J. Michael Steele, On Frieze's $\zeta(3)$ limit for lengths of minimal spanning trees, *Discrete Applied Mathematics* 18 (1987) 99-103.

[19] J. Michael Steele, Probability Theory and Combinatorial Optimization, SIAM CBMS series, 1996.

[20] D.W. Walkup, On the expected value of a random asignment problem, *SIAM Journal on Computing* 8 (1979) 440-442.

[21] R.M. Young, Euler's constant, *Mathematical Gazette* 75 (1991) 187-190.