The effect of adding randomly weighted edges

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

May 31, 2021

Abstract

We consider the following question. We have a dense regular graph G with degree αn, where $\alpha > 0$ is a constant. We add $m = o(n^2)$ random edges. The edges of the augmented graph $G(m)$ are given independent edge weights $X(e), e \in E(G(m))$. We estimate the minimum weight of some specified combinatorial structures. We show that in certain cases, we can obtain the same estimate as is known for the complete graph, but scaled by a factor α^{-1}. We consider spanning trees, shortest paths and perfect matchings in (pseudo-random) bipartite graphs.

1 Introduction

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 [34] on the performance of the simplex algorithm, see also Vershynin [37] and Dadush and Huiberts [13].

The paper [34] inspired the following model of Bohman, Frieze and Martin [10]. They consider adding random edges to an arbitrary member G of $\mathcal{G}(\alpha)$. Here α is a positive constant and $\mathcal{G}(\alpha)$ is the set of graphs with vertex set $[n]$ and minimum degree at least αn. They show that adding $O(n)$ random edges to G is enough to create a Hamilton cycle w.h.p. This is in contrast to the approximately $\frac{1}{2}n \log n$ edges needed if we rely only on the random edges. Research on this model and its variations has been quite substantial, see for example [11], [27], [35], [25], [26], [12], [5], [29], [9], [21], [14], [33], [16], [31].

Anastos and Frieze [4] introduced a variation on this theme by adding color to the edges. They consider rainbow Hamiltonicity and rainbow connection in the context of a randomly colored dense graph with the addition of randomly colored edges. Aigner-Horev and Hefetz [1] strengthened the Hamiltonicity result of [4].

In this paper we introduce another variation. We start with a dense graph in which each edge e has a random weight $X(e)$ and add randomly weighted random edges. We study the effect on the minimum value of various

*Research supported in part by NSF grant DMS1952285

combinatorial structures. We will for simplicity restrict our attention to what we will call $\mathcal{G}_{reg}(\alpha)$, the graphs in $\mathcal{G}(\alpha)$ that are αn regular.

1.1 Spanning Trees

We start with spanning trees. Suppose that $G \in \mathcal{G}_{reg}(\alpha)$ and each edge e of G is given an independent random weight $X(e)$ chosen uniformly from $[0,1]$. Let $mst(G)$ denote the expected minimum weight of a spanning tree of G, assuming it is connected. Beveridge, Frieze and McDiarmid [6] and Frieze, Ruszinko and Thoma [20] show that assuming certain connectivity conditions on G,

$$mst(G) \approx \frac{\zeta(3)}{\alpha} \text{ as } n \to \infty. \tag{1}$$

where for a positive integer $k \geq 2$ we have $\zeta(k) = \sum_{n=1}^{\infty} n^{-k}$.

Here $A_n \approx B_n$ if $A_n = (1+o(1))B_n$ as $n \to \infty$ and $A_n \lesssim B_n$ if $A_n \leq (1+o(1))B_n$ as $n \to \infty$ and $A_n \gg B_n$ if $A_n/B_n \to \infty$.

Now let $G(m)$ be obtained from G by adding m random edges to G. Each added random edge also has an independent uniform $[0,1]$ weight. Also, let $G(p)$ be obtained from G by independently adding randomly weighted copies of edges not in G, with probability p. We let R_m, R_p denote the added edges. Our first theorem is a simple extension of (1).

Theorem 1. *Suppose that $G \in \mathcal{G}_{reg}(\alpha)$ and $n \log n \ll m \ll n^{5/3}$ and the edges of $G(m)$ have independent weights chosen uniformly from $[0,1]$. Then w.h.p.*

$$mst(G(m)) \approx \frac{\zeta(3)}{\alpha} \text{ as } n \to \infty. \tag{2}$$

In addition, if $\alpha > 1/2$ then (2) holds without the use of random edges.

This theorem is very easy to prove. One simply verifies that certain conditions in [6] hold w.h.p. On the other hand it sets the stage for what we are trying to prove in other scenarios. The upper bound on m is not essential, we could most likely replace it by $o(n^2)$, but this would require us to re-do the calculations in [6].

Without the addition of random edges, all that can be claimed (assuming G is connected) is that

$$\frac{\zeta(3)}{\alpha} \lesssim mst(G) \lesssim \frac{\zeta(3)+1}{\alpha}. \tag{3}$$

See [20].

Conjecture: The $+1$ in (3) can be replaced by $+1/2$ (which is best possible).

The example giving $1/2$ is a collection of n/r copies of $H = K_r - e, r = \alpha n$ where there is a perfect matching on the vertices of degree $r-2$ added so that the copies of H are connected in a cycle by bridges.

1.2 Shortest paths

We turn our attention next to shortest paths. Janson [23] considered the following scenario: the edges of K_n are given independent exponential mean one random lengths, denoted by $E(1)$. Let $d_{i,j}$ denote the shortest distance between vertex i and vertex j. He shows that w.h.p.

$$d_{1,2} \approx \frac{\log n}{n}, \quad \max_{j\in[n]} d_{1,j} \approx \frac{2\log n}{n}, \quad \max_{i,j} d_{i,j} \approx \frac{3\log n}{n}.$$

Bhamidi and van der Hofstad [7] proved an equivalent expression for $d_{1,2}$ for a much wider class of distribution. They actually determined an asymptotic limiting distribution. (See also Bhamidi, van der Hofstad and Hooghiemstra [8].) We prove the following:

Theorem 2. *Suppose that $n^2/\log n \ll m \ll n^2$ and that $G \in G_{reg}(\alpha)$ and the edges of $G(m)$ are given independent exponential mean one random lengths. Let $d_{i,j}$ denote the shortest distance between vertex i and vertex j. Then w.h.p.*

$$d_{1,2} \approx \frac{\log n}{\alpha n}, \quad \max_{j\in[n]} d_{1,j} \approx \frac{2\log n}{\alpha n}, \quad \max_{i,j\in[n]} d_{i,j} \approx \frac{3\log n}{\alpha n}.$$

In addition, if $\alpha > 1/2$ then (2) *holds without the use of random edges.*

1.3 Bipartite matchings

We turn our attention next to bipartite matchings. For background consider the following well-studied problem: each edge of the complete bipartite graph $K_{n,n}$ is given an independent edge weight $X(e)$. Let C_n denote the minimum weight of a perfect matching in this context. Walkup [38] considered the case where $X(e)$ is uniform $[0,1]$ and proved that $\mathbb{E}(C_n) \leq 3$. Later Karp [24] proved that $\mathbb{E}(C_n) \leq 2$. Aldous [2, 3] proved that if the $X(e)$ are independent exponential mean one random variables then $\lim_{n\to\infty} \mathbb{E}(C_n) = \zeta(2) = \sum_{k=1}^{\infty} \frac{1}{k^2}$. Parisi [32] conjectured that in fact $\mathbb{E}(C_n) = \sum_{k=1}^{n} \frac{1}{k^2}$. This was proved independently by Linusson and Wästlund [28] and by Nair, Prabhakar and Sharma [30]. A short elegant proof was given by Wästlund [40, 41].

We now consider $G(m)$. G is an αn regular bipartite graph with vertex set $A \cup B, |A| = |B| = n$. Unfortunately, our proof only works if G is *pseudo-random*, as defined by Thomason [36]. By this we mean that for some $0 < \varepsilon < 1$ we have

$$|co-degree(u,v) - \alpha^2 n| \leq \mu = O(n^{1-\varepsilon}) \quad \text{for all } u,v \in A. \tag{4}$$

Here, as usual, $co-degree(u,v) = |\{w \in B : (u,w),(v,w) \in E(G)\}$.

Theorem 3. *Let G be a pseudo-random αn regular bipartite graph with vertex set $A \cup B, |A| = |B| = n$. Suppose that $n^{49/25} \ll m = o(n^2)$. Let C_n denote the minimum weight of a perfect matching when the weights of the edges of $G(m)$ are independent exponential mean one random variables. (To be clear, the m added random edges have endpoints in the different vertex classes A and B, so that $G(m)$ is bipartite.) Then*

$$\mathbb{E}(C_n) \approx \frac{\zeta(2)}{\alpha} = \frac{\pi^2}{6\alpha}. \tag{5}$$

Conjecture: equation (5) holds for $G(m)$, $m = o(n^2)$ growing sufficiently quickly, but without the assumption of pseudo-randomness.

Frieze and Johansson [18] showed that if G is the random bipartite graph $K_{n,n,p}$ where $np \gg \log^2 n$ then

$$\mathbb{E}(C_n) \approx \frac{\pi^2}{6p}. \tag{6}$$

That paper also conjectured that if (G_n) is a sequence of $r = r(n)$-regular bipartite graphs with $n+n$ vertices then $\mathbb{E}(C_n) \approx \frac{n\pi^2}{6r}$ as $r, n \to \infty$. This conjecture is false. Instead we have:
Conjecture: $\mathbb{E}(C_n) \approx \frac{n\pi^2}{6r}$ if the connectivity of G_n tends to infinity. Also, in general $\mathbb{E}(C_n) \lesssim \frac{n}{r}\left(\frac{\pi^2}{6} + \frac{1}{2}\right)$.

The 1/2 here is best possible in general. We take n/r copies of $H = K_{r,r} - e$ where there is a perfect matching on the vertices of degree $r-1$ added so that the copies of H are connected in a cycle by bridges.

In what follows we will sometimes treat large values as integers when strictly speaking we should round up or down. In all cases the choice of up or down has negligible effect on the proof.

2 Spanning Trees

Theorem 2 of Beveridge, Frieze and McDiarmid [6] yields the following. Suppose that

$$\alpha n \leq \delta(G) \leq \Delta(G) \leq \alpha(1 + O(n^{-1/3})))n. \tag{7}$$

Let $S : \bar{S}$ denote the set of edges of G with exactly one endpoint in S. Then (2) holds if

$$\frac{|S : \bar{S}|}{|S|} \geq n^{2/3} \log^{3/2} n \text{ for all } S \subseteq [n], \frac{\alpha n}{2} \leq |S| \leq \frac{n}{2}. \tag{8}$$

Now if we add m random edges satisfying the conditions of the theorem then all degrees will be $\alpha n + o(n^{2/3})$ and this will satisfy (7).

So, to prove Theorem 1, all we need to do is to verify (8). Now let $p = \frac{m}{\binom{n}{2}} \gg \frac{\log n}{n}$. The probability that $G(p)$ contains a set failing to satisfy (8) can be bounded by

$$\sum_{s=\alpha n/2}^{n/2} \binom{n}{s} \mathbb{P}(Bin(sn/2, p) \leq sn^{2/3} \log^{3/2} n) \leq \sum_{s=\alpha n/2}^{n/2} \left(\frac{ne}{s}\right)^s e^{-snp/10} = o(1), \tag{9}$$

where we have just looked at the edges R_p to satisfy (8). The property described in (8) is monotone increasing and so the $o(1)$ upper bound in (9) holds in $G(m)$ as well, see for example Lemma 1.3 of [19].

Finally note that if $\alpha > 1/2$ and S is as in (8) then each $v \in S$ has at least εn neighbors in $\bar{S}$. And therefore $|S : \bar{S}|/|S| \geq \varepsilon n$. This completes the proof of Theorem 1.

3 Shortest Paths

We use the ideas of Janson [23]. Sometimes we make a small tweak and in one case we shorten his proof considerably. The case $\alpha > 1/2$ will be discussed at the end of this section. We note that the lower bounds hold a fortiori if we do not have random edges R_p.

3.1 $d_{1,2}$

We set $S_1 = \{1\}$ and $d_1 = 0$ and consider running Dijkstra's shortest path algorithm [15]. At the end of Step k we will have computed $S_k = \{1 = v_1, v_2, \ldots, v_k\}$ and $0 = d_1, d_2, \ldots, d_k$ where d_i is the minimum length of a path from 1 to $v_i, i = 1, 2, \ldots, k$. Let there be ν_k edges from S_k to $[n] \setminus S_k$. Arguing as in [23] we see that $d_{k+1} - d_k = Z_k$ where Z_k is the minimum of ν_k independent exponential mean one random variables, independent of d_k. We note that

$$\mathbb{E}(Z_k \mid \nu_k) = \frac{1}{\nu_k} \text{ and } \mathbb{V}\mathrm{ar}(Z_k \mid \nu_k) = \frac{1}{\nu_k^2}. \tag{10}$$

Suppose now that

$$m = \frac{\omega n^2}{\log n} \text{ where } 1 \ll \omega \ll \log n.$$

It follows that w.h.p. $\delta(G(m)) \approx \Delta(G(m)) \approx \alpha n$. Now

$$k\delta - 2\binom{k}{2} \leq \nu_k \leq k\Delta(G(m))$$

and so

$$\text{w.h.p. } \nu_k \approx k\alpha n \text{ for } k = o(n). \tag{11}$$

Conditioning on the set of added edges and taking expectations with respect to edge weights, we see that if $1 \ll k = o(n)$ then

$$\mathbb{E}(d_k) = \mathbb{E}\left(\sum_{i=1}^{k-1} \frac{1}{\nu_i}\right) \approx \sum_{i=1}^{k-1} \frac{1}{i\alpha n} \approx \frac{\log k}{\alpha n}. \tag{12}$$

By the same token,

$$\mathbb{V}\mathrm{ar}(d_k) \approx \sum_{i=1}^{k-1} \frac{1}{i^2\alpha^2 n^2} = O(n^{-2}). \tag{13}$$

3.1.1 Upper Bound

If $k_0 = n^{1/2}\omega^{1/2}$ then w.h.p. $d_k \lesssim \frac{\log n}{2\alpha n}$ for $0 \leq k \leq k_0$. Now execute Dijkstra's algorithm from vertex 2 and let $\bar{d}_k, T_k$ correspond to d_k, S_k. If $S_{k_0} \cap T_{k_0} \neq \emptyset$ then we already have $d_{1,2} \lesssim \frac{\log n}{\alpha n}$. If S_{k_0}, T_{k_0} are disjoint then we use the random edges R_m or R_p. Let $p = m/\binom{n}{2} \approx 2\omega/\log n$. Then,

$$\begin{aligned}
\mathbb{P}\left(\nexists e \in R_p \cap (S_{k_0} : T_{k_0}) : X(e) \leq \frac{\log n}{\omega n}\right) &\leq \left(1 - p\left(1 - \exp\left\{-\frac{\log n}{\omega n}\right\}\right)\right)^{k_0^2} \\
&= \left(1 - (1 + o(1))\frac{p\log n}{\omega n}\right)^{k_0^2} \leq \exp\left\{-\frac{k_0^2 p \log n}{2\omega n}\right\} = e^{-\omega}.
\end{aligned} \tag{14}$$

So, in this case we see too that w.h.p.

$$d_{1,2} \leq (1 + o(1))\left(\frac{\log n}{2\alpha n} + \frac{\log n}{2\alpha n}\right) + \frac{\log n}{\omega n} \approx \frac{\log n}{\alpha n}.$$

3.1.2 Lower Bound

We now consider a lower bound for $d_{1,2}$. Let $k_1 = n^{1/2}/\log n$. We observe that because w.h.p. all vertices have degree $\approx \alpha n$ and because the edge joining v_{k+1} to S_k is uniform among $S_k : \bar{S}_k$ edges, we see that $\mathbb{P}(2 \in S_{k_1}) = O(k_1/n) = o(1)$. By the same token, $\mathbb{P}(T_{k_1} \cap S_{k_1} \neq \emptyset) = O(k_1^2/n) = o(1)$. It follows that w.h.p.

$$d_{1,2} \gtrsim 2\frac{\log k_1}{\alpha n} \approx \frac{\log n}{\alpha n}.$$

3.2 $\max_j d_{1,j}$

3.2.1 Lower Bound

For this we run Dijkstra's algorithm until all vertices have been included in the shortest path tree. We can therefore immediately see that if $k_2 = n/\log n$ then

$$\mathbb{E}(\max_j d_{1,j}) \gtrsim \sum_{i=1}^{k_2} \frac{1}{i\alpha n} + \sum_{i=n-k_2+1}^{n-1} \frac{1}{(n-i)\alpha n} \approx \frac{2\log n}{\alpha n}. \tag{15}$$

The second sum in (15) is the contribution from adding the final k_2 vertices and uses $\nu_{n-i} \approx (n-i)\alpha n$ w.h.p. for $i = o(n)$. Equation (13) allows us to claim the lower bound w.h.p.

3.2.2 Upper Bound

For an upper bound we use the fact that w.h.p. there are approximately $i(n-i)p$ R_p edges between S_i and $\bar{S}_i$ in order to show that if $k_2 = n/\omega$ then

$$\begin{aligned}\mathbb{E}(\max_j d_{1,j}) &\lesssim \left(\frac{2\log n}{\alpha n} + \sum_{i=k_2+1}^{n-k_2} \frac{1}{i(n-i)p}\right) \\ &\approx \frac{2\log n}{\alpha n} + \frac{\log n}{2\omega n}\sum_{i=k_2+1}^{n-k_2}\left(\frac{1}{i} + \frac{1}{n-i}\right) = \frac{2\log n}{\alpha n}\left(1 + \frac{(\alpha+o(1))\log\omega}{2\omega}\right) \approx \frac{2\log n}{\alpha n}.\end{aligned} \tag{16}$$

Equations (15) and (16) imply that $\mathbb{E}(\max_j d_{1,j}) \approx \frac{2\log n}{\alpha n}$ and we can use equation (13) to get concentration around the mean.

3.3 $\max_{i,j} d_{i,j}$

3.3.1 Lower Bound

Our proof here is somewhat shorter than that in [23], but it is based on the same idea. We begin with a lower bound. Let $Y_v = \min\{X(e) : e = \{v,w\} \in G(m)\}$. Let $A = \left\{v : Y_v \geq \frac{(1-\varepsilon)\log n}{\alpha n}\right\}$. Then, given that all vertex degrees are asymptotically equal to αn, we have that for $v \in [n]$,

$$\mathbb{P}(v \in A) = \exp\left\{-(\alpha n + o(n))\frac{(1-\varepsilon)\log n}{\alpha n}\right\} = n^{-1+\varepsilon+o(1)}. \tag{17}$$

An application of the Chebyshev inequality shows that $|A| \approx n^{\varepsilon+o(1)}$ w.h.p. and we can assume the existence of $a_1 \neq a_2 \in A$. Now the expected number of paths from a_1 to a_2 of length at most $\frac{(3-2\varepsilon)\log n}{\alpha n}$ can be bounded by

$$n^{2\varepsilon+o(1)} \times n^2 \times n^{-3\varepsilon+o(1)} \times \frac{\log^2 n}{\alpha^2 n^2} = n^{-\varepsilon+o(1)}. \tag{18}$$

Explanation for (18)**:** The first factor $n^{2\varepsilon+o(1)}$ is the expected number of pairs of vertices $a_1, a_2 \in A$. The second factor is a bound on the number of choices b_1, b_2 for the neighbors of a_1, a_2 on the path. The third factor F_3 is a bound on the expected number of paths of length at most $\frac{\beta \log n}{\alpha n}$ from b_1 to b_2, $\beta = 1-3\varepsilon$. This factor comes from

$$F_3 \le \sum_{\ell \ge 0} ((\alpha+o(1)n)^\ell \left(\frac{\beta \log n}{\alpha n}\right)^{\ell+1} \frac{1}{(\ell+1)!}.$$

Here ℓ is the number of internal vertices on the path. There will be $((\alpha+o(1))n)^\ell$ choices for the sequence of vertices on the path. We then use the fact that the exponential mean one random variable stochastically dominates the uniform $[0,1]$ random variable U. The final two factors are the probability that the sum of $\ell+1$ independent copies of U sum to at most $\frac{\beta \log n}{\alpha n}$. Continuing we have

$$F_3 \le \sum_{\ell \ge 0} \frac{\beta \log n}{\alpha n(\ell+1)} \left(\frac{e^{1+o(1)} \beta \log n}{\ell}\right)^\ell \le \frac{\beta \log n}{\alpha n} \left(\sum_{\ell=0}^{10 \log n} n^{\beta+o(1)} + \sum_{\ell > 10 \log n} e^{-\ell}\right) = n^{-1+\beta+o(1)} = n^{-3\varepsilon+o(1)}.$$

The final factor in (18) is a bound on the probability that $X_{a_1 b_1} + X_{a_2 b_2} \le \frac{(2+\varepsilon) \log n}{\alpha n}$. For this we use the fact that $X_{a_i b_i}, i = 1, 2$ is distributed as $\frac{(1-\varepsilon) \log n}{\alpha n} + E_i$ where E_1, E_2 are independent exponential mean one. Now $\mathbb{P}(E_1 + E_2 \le t) \le (1-e^{-t})^2 \le t^2$ and taking $t = \frac{3\varepsilon \log n}{\alpha n}$ justifies the final factor of (18).

It follows from (18) and the Markov inequality that the shortest distance between a pair of vertices in A is at least $\frac{(3-2\varepsilon) \log n}{\alpha n}$ w.h.p., completing our proof of the lower bound in Theorem 2.

3.3.2 Upper Bound

We now consider the upper bound. Let $Y_1 = d_{k_3}$ where d_k is from Section 3.1 and $k_3 = n^{1/2} \log n$. For $t < 1 - \frac{1+o(1)}{\alpha n}$ we have that w.h.p. over our choice of R_m, that

$$\mathbb{E}(e^{t\alpha n Y_1}) = \mathbb{E}\left(\exp\left\{\sum_{i=1}^{k_3} \alpha t n Z_i\right\}\right) = \prod_{i=1}^{k_3} \left(1 - \frac{(1+o(1))t}{i}\right)^{-1},$$

where the Z_i are as in (10).

Then for any $\beta > 0$ we have

$$\begin{aligned} \mathbb{P}\left(Y_1 \ge \frac{\beta \log n}{\alpha n}\right) &\le \mathbb{E}(e^{t\alpha n Y_1 - t\beta \log n}) \le e^{-t\beta \log n} \prod_{i=1}^{k_3} \left(1 - \frac{(1+o(1))t}{i}\right)^{-1} \\ &= e^{-t\beta \log n} \exp\left\{\sum_{i=1}^{k_3} \frac{(1+o(1))t}{i} + O\left(\frac{t}{i^2}\right)\right\} = \exp\left\{\left(\frac{1}{2} + o(1) - \beta\right) t \log n\right\}. \end{aligned}$$

It follows, on taking $\beta = 3/2 + o(1)$ that w.h.p.

$$Y_j \le \frac{(3+o(1)) \log n}{2\alpha n} \text{ for all } j \in [n].$$

Letting T_j be the set corresponding to S_{k_3} when we execute Dijkstra's algorithm starting at j, then we have that for $j \ne k$ where $T_j \cap T_k = \emptyset$,

$$\mathbb{P}\left(\nexists e \in R_p \cap (T_j : T_k) : X(e) \le \frac{\log n}{\omega n}\right) \le \exp\left\{-\frac{(1+o(1)) k_3^2 p \log n}{\omega n}\right\} = e^{-(2+o(1)) \log^2 n} = o(n^{-2}) \tag{19}$$

and this is enough to complete the proof of Theorem 2, except for when $\alpha > 1/2$ and we do not add random edges.

3.4 $\alpha > 1/2$

The R_p edges are needed for (14), (16) and (19). In each case we are two sets S, T of size $s = o(n)$ say and we need to argue for a short edge between them. In our case we look for a short path of length two. So, let X denote the number of triples a, b, x where $a \in S, b \in T$ and $x \notin S \cup T$ and the lengths of edges $\{a, x\}, \{b, x\}$ are both at most $p = \frac{\log n}{\omega n}$. Let $\mathcal{T}$ denote the set of such triples, so that $X = |\mathcal{T}|$. The lengths of candidate edges will not be conditioned by the history of the process. We use Janson's inequality [22].

Each pair $a \in S, b \in T$ have at least $2\varepsilon n$ common neighbors. It follows that

$$\mathbb{E}(X) \geq s^2 \varepsilon n p^2.$$

We then estimate

$$\Delta = \sum_{(a,b,x)\sim(a',b',x')} \mathbb{P}((a, b, x), (a', b', x') \in \mathcal{T}),$$

where $(a, b, x) \sim (a', b', x')$ if $\{a, x\} = \{a', x'\}$ or $\{b, x\} = \{b', x'\}$.

Then,

$$\Delta \leq \mathbb{E}(X) + 2s^2 np^2 \times sp$$

Then Janson's inequality implies

$$\mathbb{P}(X = 0) \leq \exp\left\{-\frac{\mathbb{E}(X)^2}{2\Delta}\right\} \leq \exp\left\{-\frac{s^4\varepsilon^2 n^2 p^4}{s^2np^2 + 4s^3np^3}\right\} = e^{-\Omega(snp)} = e^{-\Omega(s \log n/\omega)}. \tag{20}$$

In all cases considered, $s \geq n^{1/2+o(1)}$ an so the RHS of (20) is $o(n^{-1})$, completing the proof of Theorem 2 for the case where $\alpha > 1/2$.

4 Bipartite matchings

We find, just as in [18], that the proofs in [40], [41] can be adapted to our current situation. Suppose that the vertices of G are denoted $A = \{a_i, i \in [n]\}$ and $B = \{b_j, j \in [n]\}$. We will need to assume that

$$a_1, a_2, \ldots, a_n \text{ constitutes a random ordering of the vertices in } A.$$

We will use the notation (a, b) for edges of G, where $a \in A$ and $b \in B$. We will let $w(a, b)$ denote the weight of (a, b). Let $A_r = \{a_1, a_2, \ldots, a_r\}$ and let $C(n, r)$ denote the weight of the minimum weight matching of M_r of A_r into B. (M_r is unique with probability one.) Suppose also that ϕ_r is defined by $M_r = \{(a_i, \phi_r(a_i)) : i = 1, 2, \ldots, r\}$. Let $B_r = \{\phi_r(a_i) : i = 1, 2, \ldots, r\}$.

We will prove that

$$\mathbb{E}(C(n, r) - C(n, r - 1)) \approx \frac{1}{\alpha}\sum_{i=1}^{r} \frac{1}{r(n - i + 1)}. \tag{21}$$

for $r = 1, 2, \ldots, n - o(n)$.

Using this and a simple argument for $r \geq n - o(n)$ we argue that

$$\mathbb{E}(C_n) = \mathbb{E}(C(n, n)) \approx \frac{1}{\alpha}\sum_{r=1}^{n}\sum_{i=1}^{r} \frac{1}{r(n - i + 1)} \approx \frac{1}{\alpha}\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6\alpha}. \tag{22}$$

4.1 Proof details

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 $E(\lambda)$ weight independently, $\lambda > 0$. Here $E(\lambda)$ is an exponential random variable of rate λ i.e. $\mathbb{P}(E(\lambda) \geq x) = e^{-\lambda x}$. We now consider M_r to be the minimum weight matching of A_r into $B^* = B \cup \{b_{n+1}\}$. (As $\lambda \to 0$ it becomes increasingly unlikely that any of the extra edges are actually used in the minimum weight matching.) 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 b_{n+1} participates in M_r^*, i.e.

$$P(n,r) = \lim_{\lambda \to 0} \frac{\mathbb{P}(b_{n+1} \in B_r^*)}{\lambda}. \tag{23}$$

Its importance lies in the following lemma:

Lemma 4.

$$\mathbb{E}(C(n,r) - C(n,r-1)) = \frac{P(n,r)}{r}. \tag{24}$$

Proof. Choose i randomly from $[r]$ and let $\widehat{B}_i \subseteq B_r$ be the B-vertices in the minimum weight matching of $(A_r \setminus \{a_i\})$ into B^*. Let $X = C(n,r)$ and let $Y = C(n,r-1)$. Let w_i be the weight of the edge (a_i, b_{n+1}), and let I_i denote the indicator variable for the event that the minimum weight of an A_r matching that contains this edge is smaller than the minimum weight of an A_r matching that does not use b_{n+1}. We can see that I_i is the indicator variable for the event $\{Y_i + w_i < X\}$, where Y_i is the minimum weight of a matching from $A_r \setminus \{a_i\}$ to B. Indeed, if $(a_i, b_{n+1}) \in M_r^*$ then $w_i < X - Y_i$. Conversely, if $w_i < X - Y_i$ and no other edge from b_{n+1} has weight smaller than $X - Y_i$, 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 weight smaller than $X - Y_i$ is of order $O(\lambda^2)$. Indeed, let $\mathcal{F}$ denote the existence of two distinct edges from b_{n+1} of weight smaller than X and let $\mathcal{F}_{i,j}$ denote the event that (a_i, b_{n+1}) and $a_j, b_{n+1})$ both have weight smaller than X.

Then,

$$\mathbb{P}(\mathcal{F}) \leq n^2 \mathbb{E}_X(\max_{i,j} \mathbb{P}(\mathcal{F}_{i,j} \mid X)) = n^2 \mathbb{E}((1 - e^{-\lambda X})^2) \leq n^2 \lambda^2 \mathbb{E}(X^2), \tag{25}$$

and since $\mathbb{E}(X^2)$ is finite and independent of λ, this is $O(\lambda^2)$.

Note that Y and Y_i have the same distribution. They are both equal to the minimum weight of a matching of a random $(r-1)$-set of A into B. As a consequence, $\mathbb{E}(Y) = \mathbb{E}(Y_i) = \frac{1}{r} \sum_{j \in A_r} \mathbb{E}(Y_j)$. Since w_i is $E(\lambda)$ distributed, as $\lambda \to 0$ we have from (25) that

$$P(n,r) = \lim_{\lambda \to 0} \left(\frac{1}{\lambda} \sum_{j \in A_r} \mathbb{P}(w_j < X - Y_j) + O(\lambda) \right) = \lim_{\lambda \to 0} \mathbb{E} \left(\frac{1}{\lambda} \sum_{j \in A_r} \left(1 - e^{-\lambda(X - Y_j)}\right) \right)$$

$$= \sum_{j \in A_r} \mathbb{E}(X - Y_i) = r\mathbb{E}(X - Y).$$

□

We now proceed to estimate $P(n,r)$. Fix r and assume that $b_{n+1} \notin B_{r-1}^*$. Suppose that 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 weight. 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) $\min\{w(a_\sigma, b_j) : b_j \in B \setminus B_{r-1}\}$. 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 $\lambda_1, \lambda_2, \ldots, \lambda_M$. Then the probability that X_i is the smallest of them is $\lambda_i/(\lambda_1+\lambda_2+\cdots+\lambda_M)$. Furthermore, the probability stays the same if we condition on the value of $\min\{X_1, X_2, \ldots, X_M\}$. Thus

$$\mathbb{P}(b_{n+1} \in B^*_r \mid b_{n+1} \notin B^*_{r-1}) = \mathbb{E}\left(\frac{\lambda}{\delta_r + \lambda}\right)$$

where $\delta_r = d_{r-1}(a_\sigma)$ is the number of neighbors of a_σ in $B \setminus B_{r-1}$.

Lemma 5.

$$P(n, r) = \mathbb{E}\left(\frac{1}{\delta_1} + \frac{1}{\delta_2} + \cdots + \frac{1}{\delta_r}\right). \tag{26}$$

Proof.

$$\begin{aligned}
\lim_{\lambda\to 0} \lambda^{-1}\mathbb{P}(b_{n+1} \in B^*_r) &= \lim_{\lambda\to 0} \lambda^{-1}\mathbb{E}\left(1 - \frac{\delta_1}{\delta_1+\lambda}\cdot\frac{\delta_2}{\delta_2+\lambda}\cdots\frac{\delta_r}{\delta_r+\lambda}\right)\\
&= \lim_{\lambda\to 0} \lambda^{-1}\mathbb{E}\left(1 - \left(1+\frac{\lambda}{\delta_1}\right)^{-1}\cdots\left(1+\frac{\lambda}{\delta_r}\right)^{-1}\right)\\
&= \lim_{\lambda\to 0} \lambda^{-1}\mathbb{E}\left(\left(\frac{1}{\delta_1}+\frac{1}{\delta_2}+\cdots+\frac{1}{\delta_r}\right)\lambda + O(\lambda^2)\right)\\
&= \mathbb{E}\left(\frac{1}{\delta_1}+\frac{1}{\delta_2}+\cdots+\frac{1}{\delta_r}\right).
\end{aligned} \tag{27}$$

∎

It is this point we need to assume that G is pseudo-random. We have used this to control the values of the δ_i. We now state (part of) Theorem 2 of Thomason [36] in terms of our notation. Assume that $G(m)$ is as in Theorem 3.

Theorem 6. *If $X \subseteq A, Y \subseteq B$ and $\alpha|X| > 1$ and $x = |X|, y = |Y|$, then*

$$|e(X,Y) - \alpha xy| \le (xy(\alpha n + \mu x))^{1/2}.$$

where $e(X,Y)$ is the number of edges with one end in X and the other in Y.

4.1.1 Upper bound

We begin with an upper bound estimate for (27). This means finding lower bounds for the δ_i. Let

$$r_0 = n^\beta, \quad \omega = n^\gamma, \quad \theta = \frac{1}{\omega^2}, \quad \varepsilon = \frac{1}{\omega}, \quad k = \omega^3. \tag{28}$$

We will establish bounds on $0 < \beta, \gamma < 1$ as we proceed through the proof. We will then choose suitable values for these parameters.

We have the trivial bound $\delta_r \ge \alpha n - r$ which implies that

$$\sum_{r=1}^{r_0} \frac{1}{r} \sum_{i=1}^{r} \frac{1}{\delta_i} \le \sum_{r=1}^{r_0} \frac{1}{\alpha n - r_0} = o(1). \tag{29}$$

Now suppose that $r \geq r_0$ and let

$$\mathcal{E}_r = \{\exists S \subseteq [r, r+\theta r] : \ |S| = k, \delta_i \leq \alpha(1-\varepsilon)(n-r-i) \text{ for } i \in S\}.$$

We claim that Theorem 6 implies that $\mathcal{E}_r$ cannot occur for $r \leq n - r_0$. Indeed, suppose that $\mathcal{E}_r$ occurs. Then

$$e(S, B_{r+\theta r}) \geq \alpha \sum_{i \in S} (r + i + \varepsilon(n-r-i)) \geq \alpha k(r + \varepsilon(n-r)). \tag{30}$$

On the other hand, Theorem 6 implies that

$$e(S, B_{r+\theta r}) \leq \alpha k(r+\theta r) + (k(r+\theta r)(\alpha n + \mu k))^{1/2}. \tag{31}$$

Plugging in the values from (28) into (30) and (31) we see that after subtracting αkr the RHS of (30) is $\Omega\left(\frac{kn^\beta}{\omega}\right)$ and the RHS of (31) is $O\left(\frac{kn}{\omega^{3/2}}\right)$, a contradiction, assuming

$$\frac{n^\beta}{n^\gamma} \gg \frac{n}{n^{3\gamma/2}} \text{ or } \beta > 1 - \frac{\gamma}{2}. \tag{32}$$

Let

$$p = \frac{m}{\binom{n}{2}} = \frac{1}{n^\eta}. \tag{33}$$

Let ζ_a denote the number of times that vertex a takes the role of a_σ. We will show that with probability $1 - o(n^{-1})$ that

$$\zeta_a \leq \nu_2 = \frac{\log^3 n}{p}, \text{ for all } a \in A. \tag{34}$$

We will prove below that if $r \leq n - r_0$ then with probability $1 - O(n^{-2\eta})$

$$\delta_r \geq \nu_1 = \frac{n^{9\beta/10}}{\nu_2^2} = n^{9\beta/10 - 2\eta + o(1)} \text{ except for at most } \frac{n^{9\beta/10+2\eta+o(1)}}{\nu_2} \text{ indices } r. \tag{35}$$

Let $\xi(r)$ be the indicator for the exceptions in (35).

Let $I_1, I_2, \ldots, I_s, s = \left\lceil \frac{n-2r_0}{\theta n} \right\rceil$ be an equitable partition of $[r_0, n-r_0]$ into consecutive intervals of length $\approx \theta n$. By equitable we mean that $|I_k - I_l| \leq 1$ for all $k \neq l$. Given that $\mathcal{E}_r$ doesn't occur and (35) we see that with probability $1 - O(n^{-3\eta/2})$ we have

$$\sum_{j \in I_\ell} \frac{1}{\delta_j} \leq \frac{1}{\alpha} \sum_{j \in I_\ell} \frac{1}{(1-\varepsilon)(n-j)} + \frac{k\nu_2}{\nu_1} + \xi(I_t). \tag{36}$$

Consequently, if $\gamma_t = |I_1| + \cdots + |I_t|$ and

$$\frac{k\nu_2}{\theta \nu_1 r_0} = n^{5\gamma + 3\eta - 2\beta + o(1)} = o(1), \tag{37}$$

then with probability $1 - O(n^{-3\eta/2})$ we have

$$\begin{aligned} C(n, n-r_0) &\leq o(1) + \frac{1+o(1)}{\alpha} \sum_{t=1}^{s} \sum_{r=r_0+\gamma_{t-1}+1}^{r_0+\gamma_t} \frac{1}{r} \sum_{j=1}^{r} \frac{1}{n-j+1} + \frac{k\nu_2 s}{\nu_1 r_0} + \frac{\nu_2 n^{9\beta/10+2\eta+o(1)}}{\nu_2 r_0} \\ &= o(1) + \frac{1+o(1)}{\alpha} \sum_{r=r_0}^{n-r_0} \frac{1}{r} \sum_{j=1}^{r} \frac{1}{n-j+1} \approx \frac{\pi^2}{6\alpha}, \end{aligned} \tag{38}$$

assuming that

$$\eta < \frac{\beta}{20}. \tag{39}$$

Explanation: The first $o(1)$ term in (38) comes from (29). For a proof of the final estimate in (38) we refer the reader to Section A in an appendix. (The calculation is taken from [18].)

We show later that with probability $1 - o(n^{-1})$ we have

$$C(n,n) - C(n, n - r_0) = o(1). \tag{40}$$

To get an upper bound on $\mathbb{E}(C_n)$ we have to deal with the possibility of failure of (35). So what we do is to think of adding R_p as the union of two copies of $R_{p/2}$, a red copy and a blue copy. With just the red copy, we see that with probability $1 - O(n^{-3\eta/2})$ we have $C_n \lesssim \frac{\pi^2}{6\alpha}$ and we use (6) for the blue copy. Thus,

$$\mathbb{E}(C_n) \leq \frac{\pi^2}{6\alpha} + O(n^{-2\eta} \times p^{-1}) \approx \frac{\pi^2}{6\alpha}.$$

4.1.2 Lower bound

Let $\ell = n^{1/3}$ and $n_1 = n - r_0$ and $s = \left\lceil \frac{n_1}{\ell} \right\rceil \approx n^{2/3}$ and equitably partition $[n_1]$ into intervals $I_j, j = 1, 2, \ldots, \ell$ of length $\approx s$ and define $\gamma_t = |I_1| + \cdots + |I_t|$ as before. Fix j and let $I = I_j$. Next let $S_i, i \leq \nu_2$ denote the set of elements of A that appear i times as a_σ in I and let $s_i = |S_i|$. Let T_i denote the subset of I_j corresponding to S_i. Partition $T_i = U_1 \cup \cdots \cup U_i$ into i copies of S_i in a natural way. Then it follows from Theorem 6 that if $s_i\alpha > 1$ then for $1 \leq k \leq i$,

$$\left| \sum_{j \in U_k} \delta_j - s_i\alpha(n - (j-1)s) \right| \leq (s_i(n - (j-1)s)(\alpha n + \mu s_i))^{1/2}.$$

Therefore,

$$\begin{aligned} \sum_{i: s_i > 1/\alpha} \left(\sum_{j \in T_i} \delta_j - i s_i \alpha(n - (j-1)s) \right) &\leq \sum_{i: s_i > 1/\alpha} i(s_i(n - (j-1)s)(\alpha n + \mu s_i))^{1/2} \\ &\leq n^{1/2}(n - (j-1)s)^{1/2} \sum_{i: s_i > 1/\alpha} i s_i^{1/2} \leq \nu_2^2 s^{1/2} n^{1/2} (n - (j-1)s)^{1/2}. \end{aligned}$$

It follows then that

$$\sum_{k \in I_j} \delta_k \leq \alpha s(n - (j-1)s) + \nu_2^2 s^{1/2} n^{1/2} (n - (j-1)s)^{1/2} + \alpha^{-1}\nu_2(n - (j-1)s). \tag{41}$$

We have from (41) and the fact that the harmonic mean is at most the arithmetic mean that if

$$\eta < \frac{1}{6} \text{ which implies that } \nu_2^2 = o(s^{1/2}) \tag{42}$$

then, assuming

$$\beta > \frac{2}{3}, \tag{43}$$

$$\sum_{i\in I_j}\frac{1}{\delta_i}\geq\frac{s^2}{\sum_{i\in I_j}\delta_i}\geq\frac{s}{\alpha(n-(j-1)s)\left(1+\frac{\nu_2^2}{\alpha}\left(\frac{n}{s(n-(j-1)s)}\right)^{1/2}+\frac{\nu_2}{\alpha s}\right)}$$

$$=\frac{s}{\alpha(n-(j-1)s)}\left(1+O\left(\frac{\nu_2^2}{s^{1/2}\left(1-\frac{(j-1)s}{n}\right)}\right)\right)=$$

$$\frac{1}{\alpha}\sum_{i\in I_j}\frac{1}{n-i+1}\left(1+O\left(\frac{\nu_2^2}{s^{1/2}\left(1-\frac{(j-1)s}{n}\right)}\right)+O\left(\frac{s}{n-(j-1)s}\right)\right)\approx\frac{1}{\alpha}\sum_{i\in I_j}\frac{1}{n-i+1}.$$

Therefore,

$$\mathbb{E}(C(n,n_1))=\mathbb{E}\left(\sum_{r=1}^{n_1}\frac{1}{r}\sum_{i=1}^{r}\frac{1}{\delta_i}\right)=\mathbb{E}\left(\sum_{j=1}^{\ell}\sum_{r\in I_j}\frac{1}{r}\sum_{i=1}^{r}\frac{1}{\delta_i}\right)\gtrsim\frac{1}{\alpha}\sum_{j=1}^{\ell}\sum_{r\in I_j}\frac{1}{r}\sum_{i=1}^{\gamma_{j-1}}\frac{1}{n-i+1}\geq$$

$$\frac{1}{\alpha}\sum_{r=1}^{n_1}\frac{1}{r}\sum_{i=1}^{r}\frac{1}{n-i+1}-\frac{s}{n-s}-\sum_{j=2}^{\ell}\frac{s}{(j-1)s(n-js)}=\frac{1}{\alpha}\sum_{r=1}^{n_1}\frac{1}{r}\sum_{i=1}^{r}\frac{1}{n-i+1}-o(1)\approx\frac{\pi^2}{6\alpha}. \quad (44)$$

Explanation: the term $-\sum_{j=2}^{\ell}\frac{s}{(j-1)s(n-js)}$ accounts for the third summation in the last term of the first line only going as far as γ_{j-1} instead of γ_j. The term $-\frac{s}{n-s}$, which is small due to (43), accounts for the $j=1$ summand in the last term of the first line being missing, due to the same thing.

This gives the correct lower bound for Theorem 3. The proof of the final estimate in (44) is proved in the same way as the final estimate in (38).

We now have to verify (35), (40). These claims rest on a bound on the maximum weight of an edge in the minimum weight perfect matching.

4.1.3 No long edges

The aim of this section is to show that w.h.p. no edges of weight more than $w_1=2w_0\log n$ (where $w_0=\frac{c_1\log n}{np}$) are used in the construction of M_n. Here c_1 is a sufficiently large constant. For a set $S\subseteq A, 2\leq|S|$ we let

$$N_0(S)=\{b\in B:(a,b)\in R_p\text{ and }w(a,b)\leq w_0\text{ for some }a\in S\}.$$

And for $a\in A$ let

$$N_0(a)=\{b\in B:(a,b)\in E(G)\text{ and }w(a,b)\leq w_0\}$$

Let

$$r_1=p^{-1}\log^{1/2}n;\qquad r_2=\frac{n}{10c_1\log n};\qquad r_3=\frac{n}{2000};\qquad r_4=n-\frac{4000n}{c_1\log n}.$$

Lemma 7. *W.h.p. we have*

$$|N_0(a)| \geq 2r_1 \quad \textit{for all } a \in A. \tag{45}$$

$$|N_0(S)| \geq \frac{c_1|S|\log n}{4} \quad \textit{for all } S \subseteq A, r_1 < |S| \leq r_2. \tag{46}$$

$$|N_0(S)| \geq \frac{n}{40} \quad \textit{for all } S \subseteq A, r_2 < |S| \leq r_3. \tag{47}$$

$$|N_0(S)| \geq n - \frac{3000n}{c_1 \log n} \quad \textit{for all } S \subseteq A, r_3 < |S|. \tag{48}$$

$$n - |N_0(S)| \leq \frac{1}{2}(n - |S|) \quad \textit{for all } S \subseteq A, |S| \geq r_4. \tag{49}$$

Proof. We first observe that $|N_0(a)|$ is distributed as $Bin(\alpha n, 1 - e^{-w_0})$ and $\alpha n(1 - e^{-w_0}) \gtrsim \alpha r_1 \log^{1/2} n$ and so the Chernoff bounds imply that

$$\mathbb{P}(\exists a : |N_0(a)| \leq 2r_1) \leq ne^{-\alpha n w_0/4} = o(n^{-1}).$$

We next observe that for a fixed $S \subseteq A$ we have

$$|N_0(S)| \sim Bin(n, q) \qquad \text{where } q = 1 - (1 - p(1 - e^{-w_0}))^{|S|} = 1 - (1 - (1 + o(1))w_0 p)^{|S|}.$$

(Here $\sim$ is used to indicate the distribution of $|N_0(S)|$.)

If $r_1 < |S| \leq r_2$ then $q \geq w_0 p|S|/2$. So,

$$\mathbb{P}(\neg(46)) \leq \sum_{s=r_1}^{r_2} \binom{n}{s} \mathbb{P}\left(Bin(n, q) \leq \frac{c_1 s \log n}{4}\right) \leq \sum_{s=r_1}^{r_2} \left(\frac{ne}{s}\right)^s e^{-c_1 s \log n/4} = \sum_{s=r_1}^{r_2} \left(\frac{n^{1-c_1/4}e}{s}\right)^s = o(1).$$

If $r_2 < |S| \leq r_3$ then $q > 1/20$. So,

$$\mathbb{P}(\neg(47)) \leq \sum_{s=r_2}^{r_3} \binom{n}{s} \mathbb{P}\left(Bin\left(n, \frac{1}{20}\right) \leq \frac{n}{40}\right) \leq \sum_{s=r_2}^{r_3} \left(\frac{ne}{s}\right)^s e^{-n/160} \leq 2(2000)^{n/2000} e^{-n/160} = o(1).$$

If $r_3 < |S| \leq r_4$ then $q \geq 1 - n^{-c_1/3000}$. So,

$$\mathbb{P}(\neg(48)) \leq 2^n n^{-(n-n_0)c_1/3000} = o(1).$$

If $r_4 < |S|$, let $t = n - |S|$. Then, $q \geq 1 - n^{-c_1/2}$ and so

$$\mathbb{P}(\neg(49)) \leq \sum_{t=1}^{n-r_4} \binom{n}{t}\binom{n}{t/2}(1 - q)^{t/2} \leq \sum_{t=1}^{n-r_4} \left(\frac{ne}{t}\right)^{2t} n^{-c_1 t/2} = o(1).$$

□

Lemma 8. *W.h.p., no edge of length at least w_1 appears in any $M_r, r \leq n$*

Proof. We first consider $r = 1, 2, \ldots, r_1 = p^{-1}\log^{1/2} n$. If $a \in A_r$ and $w(a, \phi_r(a)) > w_0$ then (45) implies that there are at least r_1 choices of $b \in B \setminus \phi(A_r)$ such that we can reduce the matching cost by replacing $(a, \phi_r(a))$ by (a, b).

We now consider $r > r_1$. Choose $a \in A_r$ and let $S_0 = \{a\}$ and let an alternating path $P = (a = u_1, v_1, \ldots, v_{k-1}, u_k, \ldots)$ be *acceptable* if (i) $u_1, \ldots, u_k, \ldots \in A$, $v_1, \ldots, v_{k-1}, \ldots \in B$, (ii) $(u_{i+1}, v_i) \in M_r, i = 1, 2, \ldots$ and (iii) $w(u_i, v_i) \leq w_0$, $i = 1, 2, \ldots$.

Now consider the sequence of sets $S_0 = \{a_0\}, S_1, S_2, \ldots, S_i, \ldots$ defined as follows:

Case (a): $N_0(S_i) \subseteq \phi(A_r)$. In this case we define $S_{i+1} = \phi_r^{-1}(T_i)$, where $T_i = N_0(S_i)$. By construction then, every vertex in $S_j, j \leq i+1$ is the endpoint of some acceptable alternating path.

Case (b): $T_i \backslash \phi(A_r) \neq \emptyset$. In this case there exists $b \in T_i$ which is the endpoint of some acceptable augmenting path.

It follows from (46) applied to S_i that w.h.p. there exists $k = O\left(\frac{\log n}{\log \log n}\right)$ such that $|N_0(S_k)| > r$ and so Case (b) holds. This implies that if $r_1 \leq r \leq r_2$ then $w(a, \phi_r(a)) \leq w_0 \log n$ for all $a \in A_r$. For if $w(a, \phi_r(a)) > w_0 \log n$ then there are at least $\Omega(r \log n)$ choices of $b \in B \setminus \phi(A_r)$ such that we can reduce the matching cost by deleting $(a, \phi_r(a))$ and changing M_r via an acceptable augmenting path from a to b. The extra cost of the edges added in this path is $o(w_0 \log n)$.

Now consider $r_2 < r \leq r_3 = n/100$. We know that w.h.p. there is $k = o(\log n)$ such that $|S_k| > r_2$ and that by (47) we have that w.h.p. $|N_0(S_{k+1})| > n/40 > r$ and we are in Case (b) and there is a low cost augmenting path for every a, as in the previous case. When $r_3 < |S_k| \leq r_4$ we use the same argument and find by (48) we have w.h.p. $N_0(S_{k+1}) > r_4 \geq r$ and there is a low cost augmenting path. Similarly for $r > r_4$, using (49).

Finally note that the number of edges in the augmenting paths we find is always at most $o(\log n) + \log_2 n \leq 2 \log n$. □

This also proves that

$$\mathbb{E}(C(n,n) - C(n, n - r_0) = O\left(n^\beta w_1\right) = o(1),$$

provided

$$\beta < 1 - \eta. \tag{50}$$

This verifies (40).

To prove (34) we argue

$$\begin{aligned} \mathbb{P}\left(\exists a \in A : |\{e : a \in e, X_e \leq w_1\}| \geq \frac{\log^3 n}{p}\right) &\leq \mathbb{P}\left(Bin\left(\alpha n, w_1\right) \geq \frac{\log^3 n}{p}\right) \\ &\leq \binom{\alpha n}{p^{-1} \log^3 n} w_1^{p^{-1} \log^3 n} \leq \left(\frac{e w_1}{\log^3 n}\right)^{p^{-1} \log^2 n} = o(1). \end{aligned} \tag{51}$$

This verifies (34).

We finally consider (35). Consider how a vertex $a \in A$ loses neighbors in $B \setminus B_r$. It can lose up to ν_2 for the times when $a = a_\sigma$. Otherwise, it loses a neighbor when $a_\sigma \neq a$ chooses a common neighbor with a. The important point here is that this choice depends on the structure of G, but not on the weights of edges incident with a. It follows that the cheapest neighbors at any time are randomly distributed among the current set of available neighbors. To get to the point where $a_\sigma = a$ and $\delta_r \leq \nu_1$, we must have at least one of the ν_2 original cheapest neighbors occuring in a random ν_1 subset of a set of size $\approx \mu_r = \min\{\alpha n, n - r\}$. This has probability $O(\nu_1 \nu_2 / \mu_r)$ and (35) follows from the Markov inequality.

We finally choose β, γ, η such that (32), (37), (39), (42), (43) and (50) hold. We let $\beta = 5/6 + \eta$ and then we choose $\gamma = 1/3 - \eta, \eta = 1/25$.

5 Final remarks

We have shown that adding sufficiently many random edges is enough to "smooth out" the optimal value in certain optimization problems. There are several questions that remain. The first is to remove the pseudo-random requirement from Theorem 3. The problem is to control the sizes of the δ_r. Another possibility is to consider matchings and 2-factors in arbitrary regular graphs, not just bipartite ones. Then one can consider the Travelling Salesperson problem. We could also consider relaxing α to be $o(1)$ and we can consider more general distributions than $E(1)$.

References

[1] E. Aigner-Horev and D. Hefetz, Rainbow Hamilton cycles in randomly coloured randomly perturbed dense graphs.

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

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

[4] M. Anastos and A.M. Frieze, How many randomly colored edges make a randomly colored dense graph rainbow Hamiltonian or rainbow connected?, *Journal of Graph Theory* 92 (2019) 405-414.

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

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

[7] S. Bhamidi and R. van der Hofstadt, Weak disorder in the stochastic mean field model of distance, *Annals of Applied Probability* 22 (2012) 29-69.

[8] S. Bhamidi, R. van der Hofstadt and G. Hooghiemstra, First passage percolation on random graphs with finite mean degrees, *Annals of Probability* 20 (2010) 1907-1965.

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

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

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

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

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

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

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

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

[17] C.M. Fortuin, P.W. Kasteleyn and J. Ginibre, Correlation inequalities on some partially ordered sets, *Communications in Mathematical Physics* 22 (1971) 89-103,

[18] A.M. Frieze and T. Johansson, Minimum-cost matching in a random graph with random costs, *SIAM Journal on Discrete Mathematics* 31 (2017) 489-510.

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

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

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

[22] S. Janson, Poisson approximation for large deviations, *Random Structures and Algorithms* 1 (1990) 221-230.

[23] 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.

[24] 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.

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

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

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

[28] 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.

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

[30] 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.

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

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

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

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

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

[36] A. Thomason, Dense expanders and pseudo-random bipartite graphs, *Discrete Mathematics* 75 (1989) 381-386.

[37] R. Vershynin, Beyond Hirsch Conjecture: Walks on Random Polytopes and Smoothed Complexity of the Simplex Method, *SIAM Journal on Computing* 39 (2009) 646678.

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

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

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

[41] J. Wästlund, An easy proof of the $\zeta(2)$ limit in the random assignment problem, *Electronic Communications in Probability* **14** (2009) 261-269.

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

A Proof of final estimate from (38)

We use the following expression from Young [42].

$$\sum_{i=1}^{n} \frac{1}{i} = \log n + \gamma + \frac{1}{2n} + O(n^{-2}), \qquad \text{where } \gamma \text{ is Euler's constant.} \tag{52}$$

$$\begin{aligned}
&\sum_{r=r_0}^{n-r_0} \frac{1}{r} \sum_{j=1}^{r} \frac{1}{n-j+1} \sum_{j=r_0}^{n-r_0} \frac{1}{n-j+1} \sum_{r=j}^{n-r_0} \frac{1}{r}, \\
&\qquad = \sum_{j=r_0}^{n-r_0} \frac{1}{n-j+1} \left(\log\left(\frac{n-r_0}{j}\right) + \frac{1}{2(n-r_0)} - \frac{1}{2j} + O(j^{-2}) \right) + o(1), \\
&\qquad = \sum_{j=r_0}^{n-r_0} \frac{1}{n-j+1} \log\left(\frac{n-r_0}{j}\right) + o(1), \\
&\qquad = \sum_{i=r_0}^{n-r_0} \frac{1}{i+1} \log\left(\frac{n-r_0}{n-i}\right) + o(1), \qquad (53) \\
&\qquad = \int_{x=r_0}^{n-r_0} \frac{1}{x} \log\left(\frac{n-r_0}{n-x}\right) dx + o(1).
\end{aligned}$$

We can replace the sum in (53) by an integral because the sequence of summands is unimodal and the terms are all $o(1)$.

Continuing, we have

$$\begin{aligned}
&\int_{x=r_0}^{n-r_0} \frac{1}{x} \log\left(\frac{n-r_0}{n-x}\right) dx \\
&= -\int_{x=r_0}^{n-r_0} \frac{1}{x} \log\left(1-\frac{x-r_0}{n-r_0}\right) dx \\
&= \sum_{k=1}^{\infty} \int_{x=r_0}^{n-r_0} \frac{1}{x} \frac{(x-r_0)^k}{k(n-r_0)^k} dx \\
&= \sum_{k=1}^{\infty} \int_{y=0}^{n-2r_0} \frac{1}{y+r_0} \frac{y^k}{k(n-r_0)^k} dy. \qquad (54)
\end{aligned}$$

Observe next that for every $k \geq 1$

$$\int_{y=0}^{n-2r_0} \frac{1}{y+r_0} \frac{y^k}{k(n-r_0)^k} dy \leq \int_{y=0}^{n-2r_0} \frac{y^{k-1}}{k(n-r_0)^k} dy \leq \frac{1}{k^2}.$$

So,

$$0 \leq \sum_{k=\log n}^{\infty} \int_{x=r_0}^{n-r_0} \frac{1}{x} \frac{(x-r_0)^k}{k(n-r_0)^k} dx \leq \sum_{k=\log n}^{\infty} \frac{1}{k^2} = o(1). \qquad (55)$$

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

$$\int_{y=0}^{n-2r_0} \frac{1}{y+r_0} \frac{y^k}{k(n-r_0)^k} dy = \int_{y=0}^{n-2r_0} \frac{(y+r_0)^{k-1}}{k(n-r_0)^k} dy + \int_{y=0}^{n-2r_0} \frac{y^k-(y+r_0)^k}{(y+r_0)k(n-r_0)^k} dy.$$

Now

$$\int_{y=0}^{n-2r_0} \frac{(y+r_0)^{k-1}}{k(n-r_0)^k} dy = \frac{1}{k^2} \frac{(n-r_0)^k-(r_0+1)^k}{(n-r_0)^k} = \frac{1}{k^2} + O(n^{-\beta k/2}). \qquad (56)$$

If $k=1$ then our choice of r_0 implies that

$$\int_{y=0}^{n-2r_0} \frac{(y+r_0)^k-y^k}{(y+r_0)k(n-r_0)^k} dy \leq \frac{r_0\log(n-2r_0)}{n-r_0} = o(1).$$

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

$$\begin{aligned}
\int_{y=0}^{n-2r_0} \frac{(y+r_0)^k-y^k}{(y+r_0)k(n-r_0)^k} dy &= \sum_{l=1}^{k} \int_{y=0}^{n-2r_0} \binom{k}{l} \frac{y^{k-l} r_0^l}{(y+r_0)k(n-r_0)^k} dy \\
&\leq \sum_{l=1}^{k} \int_{y=0}^{n-2r_0} \binom{k}{l} \frac{y^{k-l-1} r_0^l}{k(n-r_0)^k} dy \\
&= \sum_{l=1}^{k} \binom{k}{l} \frac{r_0^l (n-2r_0)^{k-l}}{k(k-l)(n-r_0)^k} \qquad (57) \\
&= O\left(\frac{kr_0}{k(k-1)n}\right) = O\left(\frac{1}{kn^{1-\beta}}\right). \qquad (58)
\end{aligned}$$

To go from (57) to (58) we argue that if the summand in (57) is denoted by u_l then $u_{l+1}/u_l = O(r_0/n)$ for $2 \leq l \leq \log n$. Hence the sum is $O(u_1)$.

It follows that

$$0 \leq \sum_{k=1}^{\log n} \int_{y=0}^{n-2r_0} \frac{(y+r_0)^k - y^k}{(y+r_0)k(n-r_0)^k} dy = o(1) + O\left(\sum_{k=2}^{\log n} \frac{1}{kn^{1-\beta}} \right) = o(1). \tag{59}$$

Equations (55), (56) and (59) complete the argument.