Analysis of a Simple Greedy Matching Algorithm on Random Cubic Graphs

Alan Frieze* A.J. Radcliffe† Stephen Suen
Department of Mathematics,
Carnegie Mellon University,
Pittsburgh PA15213, U.S.A.

June 14, 1995

Abstract

We consider the performance of a simple greedy matching algorithm MINGREEDY when applied to random cubic graphs. We show that if λ_n is the expected number of vertices not matched by MINGREEDY, then there are positive constants c_1 and c_2 such that $c_1 n^{1/5} \leq \lambda_n \leq c_2 n^{1/5} \log n$.

1 Introduction

There have been a number of papers analysing simple greedy-type algorithms for finding large matchings in graphs, e.g. Korte and Hausmann [6], Karp and Sipser [7], Tinhofer [8], Dyer and Frieze [3], Goldschmidt and Hochbaum [5] and Dyer, Frieze and Pittel [4]. Most of these deal with the expected performance of various algorithms on random graphs. In this paper we discuss the algorithm MINGREEDY, given below. It is simple and can easily be implemented in linear time.

We use the following notation in the description of MINGREEDY. $\Gamma_G(v)$ denotes the set of neighbours of the vertex v in the graph G; often G is understood and the subscript is omitted. $G \setminus \{u, v\}$ is the graph obtained from G by deleting the vertices u and v, all edges incident with them and, in addition, any new isolated vertices.

*Supported by NSF grant CCR-9024935
†Current address: Department of Mathematics, Georgia Institute of Technology, Atlanta GA30332

MINGREEDY

> **Input** G;
> **begin**
> $M \leftarrow \emptyset$;
> **while** $E(G) \neq \emptyset$ **do**
> **begin**
> ***A***: Choose $v \in V$ uniformly at random from among the vertices of minimum degree in G;
> ***B***: Choose $u \in \Gamma(v)$ uniformly at random and set $e = \{u, v\}$;
> $G \leftarrow G \setminus \{u, v\}$;
> $M \leftarrow M \cup \{e\}$
> **end**;
> Output M as our matching
> **end**

The performance of MINGREEDY, at least on random cubic graphs, is remarkably good. In Theorem 1.1 below we prove that the number of vertices left unmatched is usually comparatively small, but it is instructive to see the algorithm's performance in "practice". We have done a limited amount of computation and some results are given in Table 1. We have compared it to two other greedy matching algorithms: GREEDY, which chooses an edge at random from those available, and MODIFIED GREEDY, which randomly chooses a vertex v and then an edge incident with it. The algorithms were run on random n-vertex cubic graphs, up to $n = 10^6$. The difference in performance is quite dramatic. It makes good sense to use an algorithm like MINGREEDY as a "front end" to an optimising algorithm.

n	i_M	i_R	i_V
10^2	1.2	12.0	11.8
10^3	2.2	116.3	102.2
10^4	3.5	1193.3	1049.9
10^5	6.2	11892.7	10472.1
10^6	10.4	119091.9	104546.8

Table 1: Performance of greedy type matching algorithms in 20 Monte Carlo runs, where i_M (respectively i_R, i_V) is the average number of vertices left isolated when using MINGREEDY (respectively GREEDY, MODIFIED GREEDY).

This paper is concerned with the performance of MINGREEDY on random graphs. For a graph G we let $\lambda(G)$ denote the expected number of vertices left unmatched by a run of MINGREEDY[1]. Let Φ_n denote the set of 3-regular graphs with vertex set $[n] = \{1, 2, \ldots, n\}$.

[1] MINGREEDY is a randomising algorithm and the expectation here is with respect to the algorithm's random choices

Theorem 1.1 *Let G_n be chosen uniformly at random from Φ_n. Then there exist constants $c_1, c_2 > 0$ such that*

$$c_1 n^{1/5} \leq \mathbf{E}(\lambda(G_n)) \leq c_2 n^{1/5} \ln n.$$

It seems likely that the $\ln n$ factor in the upper bound is unnecessary.

Note, in contrast to this result, that the algorithm which at every stage chooses an edge randomly from among those available (analysed in [3]), is very likely to leave at least αn vertices unmatched for some absolute constant $\alpha > 0$. A similar estimate holds if, at every stage, one chooses a random vertex v and then a random neighbour of v. These calculations are borne out by the computational results in Table 1; it is crucial to the success of our algorithm that we always choose a vertex of minimum degree.

We would like to extend Theorem 1.1 to random r-regular graphs and to sparse random graphs but it seems difficult to carry through the analysis. It should be noted however that MINGREEDY is likely to perform well. It is closest in spirit to Algorithm 2 of [7], which, when run on sparse random graphs, usually leaves only $o(n)$ more vertices unmatched than the minimum possible. (This is a difficult proof and there is no estimate given of the $o(n)$ error term.)

2 Overview of the Proof

Our proof of Theorem 1.1 is rather long and so we now give a brief overview. For functions f and g, we use $f(n) \approx g(n)$ to denote that $f(n) = (1 + o(1))g(n)$ as $n \to \infty$.

Let t denote the number of iterations of the algorithm so far, i.e. the number of executions of Step $\boldsymbol{A}$, and let $n_i = n_i(t)$ be the number of vertices of degree $i = 1, 2, 3$ in the graph $G(t)$ at the end of iteration t. (So $n_1(0) = n_2(0) = 0$, $n_3(0) = n$.) Similarly, write $m(t) = \frac{1}{2}(n_1(t) + 2n_2(t) + 3n_3(t))$ for the number of edges of $G(t)$. Let $\xi(t)$ denote the number of new isolated vertices created at time t. Then the number of unmatched vertices at the end is given by

$$\lambda(G_n) = \sum_{t \geq 0} \xi(t).$$

Our ability to analyse the algorithm rests on the following fact.

Claim 2.1 *Suppose that $G = G(t)$ has n_i vertices of degree i, $i = 1, 2, 3$. Then G is equally likely to be any member of $\mathcal{G}(n_1, n_2, n_3)$, the set of all graphs with vertex set $V(G)$ and n_i vertices of degree i, $i = 1, 2, 3$.*

The claim allows us to treat the progress of the algorithm as a Markov chain $\mathcal{M}_1$ on $\mathbf{N}^3$ (where $\mathbf{N} = \{0, 1, 2, \ldots\}$), the state at time t being $(n_1(t), n_2(t), n_3(t))$. Analysis of $\mathcal{M}_1$ yields the following fact.

Claim 2.2 *Conditional on $n_1(t), n_2(t)$ and $n_3(t)$, the expected number $\xi(t)$ of isolated vertices created at iteration t is $\Theta(n_1(t)/m(t))$.*

We will see that, in a typical execution of the algorithm (except near the end of the process), $n_3(t)$ and $m(t)$ almost surely satisfy

$$n_3(t) \approx \left(\frac{2}{3}\right)^{3/2} \left(\frac{m(t)^{3/2}}{n^{1/2}}\right), \tag{2.1}$$

and n_1 is usually small compared to n_2 and n_3. The expected behaviour of n_1 is "controlled" by the size of n_3 until $m(t) \approx m_0 = n^{3/5}$. Indeed, conditional on $n_3(t)$ satisfying (2.1), we can bound the progress of $n_1(t)$ by a random walk on $\mathbf{N}$ where the particle's expected distance from the origin is $O(m(t)/n_3(t))$ for $m(t) \geq m_0$. Thus from Claim 2.2 we see that the expected number of isolated vertices created before $m \approx m_0$ is $O(\sum_{m(t)\geq m_0} n_3(t)^{-1})$. Hence the contribution to $\mathbf{E}[\lambda(G_n)]$ before $m(t) \approx m_0$ is

$$O\left(\sum_{m(t)\geq m_0} \frac{n^{1/2}}{m^{3/2}}\right) = O(n^{1/5}). \tag{2.2}$$

When $m < m_0$, we are near the end of the process and it is good enough for our purposes to bound n_1 by a symmetric random walk on $\mathbf{Z}$ (the set of integers) in which the particle moves up or down by one when it moves, and stays where it is with probablity $1 - O(n_3/m)$. Applying standard results on random walks, we obtain that for $m < m_0$, $\mathbf{E}[n_1] = O(n^{1/5})$. Claim 2.2 now gives that the contribution to $\mathbf{E}[\lambda(G_n)]$ after $m(t) \leq m_0$ is $O(n^{1/5} \ln n)$, which together with (2.2) yield the upper bound in the theorem.

To prove the lower bound we consider only those times when $An^{3/5} \geq m \geq Bn^{3/5}$ for some large constants A, B. We use coupling arguments to bound $n_1(t)$ from below by a random walk, and to estimate the probability that this walk "reaches" its steady state before too long. We then deduce that for some constants $B < A$ we have $\mathbf{E}[n_1(t)] \geq \epsilon n^{1/5}$ for all m between $Bn^{3/5}$ and $An^{3/5}$. This is enough to give the lower bound in the theorem.

We shall establish Claim 2.1 in the next section. The transition probabilities of the chain are then described in §4. In order to bound $\mathbf{E}[n_1/m]$ it is necessary to study the behaviour of n_3, which is done in §5. The lower and upper bounds in Theorem 1.1 are proved in §§6 and 7 respectively.

3 The Graph Chain

We next establish Claim 2.1 by using induction on t. Since $G(0) = G_n$ which is a random member of $\mathcal{G}(0,0,n)$, we need only establish that for $x, y \in \mathbf{N}^3$ and for $G \in \mathcal{G}(y_1, y_2, y_3)$, the number of triples in

$$\begin{aligned} I_G &= \{(H, v, w) : H \in \mathcal{G}(x_1, x_2, x_3), v \text{ is of minimum degree in } H, \\ &\qquad w \text{ is a neighbour of } v \text{ and } G = H \setminus \{v, w\}\} \end{aligned}$$

depends only on x, y. But to construct a triple in I_G, we

(a) choose $v, w \in [n] \setminus V(G)$;

(b) choose $v_1, v_2, \dots v_\ell \in [n] \setminus (V(G) \cup \{v, w\})$, where $\ell = (x_1 + x_2 + x_3) - (y_1 + y_2 + y_3) - 2$ is the number of new isolated vertices created in going from x to y;

(c) add appropriate edges incident with $v, w, v_1, v_2, \dots v_\ell$ to make a graph in $\mathcal{G}(x_1, x_2, x_3)$.

The number of choices in (a),(b) (trivially) depends only on x, y and the same is true for the number of choices in (c), which is fixed once the degree sequence of G is fixed. (Just observe that if one chooses A as the set of edges in (c) and changes G without changing the degree sequence then A remains a valid choice.) This verifies Claim 2.1.

4 The Degree Chain

In this section we derive the transition probabilities of the chain $\mathcal{M}_1$ whose state at time t is $n(t) = (n_1(t), n_2(t), n_3(t))$. We use the notation

$$p[x : y] = Pr(n(t+1) = y | n(t) = x).$$

We shall require the *configuration model* of Bollobás [2] which is a simple and useful description of that used by Bender and Canfield [1].

Suppose we are given a degree sequence $1 \le d_1, d_2, \dots d_\nu \le \Delta$. Let $W_i = \{i\} \times [d_i]$ for $i \in [\nu]$ and $W = \bigcup_{i=1}^{\nu} W_i$. A configuration is a partition of W into $\mu = \sum_{i=1}^{\nu} d_i/2$ pairs. Let Ω_ν be the set of all configurations, and let F_ν be chosen uniformly from Ω_ν. Then let $\gamma(F_\nu)$ be the multigraph with vertex set $[\nu]$ and edges $\{i, j\}$ for each occurrence of $\{(i, x), (j, y)\} \in F_\nu$ for some $x \in [d_i], y \in [d_j]$.

The properties that we need of this model are that: first, conditional on $\gamma(F_\nu)$ being simple, it is equally likely to be any simple graph with the given degree sequence; second, if Δ is an absolute constant (here of course $\Delta = 3$ would suffice), then

$$\mathbf{Pr}(\gamma(F_n) \text{ is simple}) = (1 + O(\mu^{-1})) \exp\left\{-\frac{\kappa}{2} - \frac{\kappa^2}{4}\right\}, \tag{4.1}$$

where $\kappa = \mu^{-1} \sum_{i=1}^{\nu} \binom{d_i}{2}$. (Note that

$$\mathbf{E}[\lambda(G_n)] = O(\mathbf{E}[\lambda(\gamma(F_n))]),$$

where F_n is chosen uniformly at random from Ω_n with $d_i = 3$ for all $i \in [n]$. Thus for proving the upper bound in Theorem 1.1 we need only consider this multigraph distribution together with a corresponding multigraph version of Claim 2.1.)

For $x \in \mathbf{N}^3$, let $\nu = x_1 + x_2 + x_3$ and $\mu = (x_1 + 2x_2 + 3x_3)/2$. Also, without loss of generality, assume that $d_i = 1, 1 \leq i \leq x_1, d_i = 2, x_1 < i \leq x_1 + x_2, d_i = 3, x_1 + x_2 > i$. Now choose F_ν randomly from Ω_ν and apply one step of Algorithm MINGREEDY to $\gamma(F_\nu)$. For $y \in \mathbf{N}^3$, let $\mathcal{E}_y$ be the event that the multigraph remaining has y_i vertices of degree i $(i = 1, 2, 3)$. We shall first prove that

$$p[x : y] = \mathbf{Pr}(\mathcal{E}_y) + O(1/\nu). \tag{4.2}$$

Note that from Claim 2.1, all we need to show is that

$$\mathbf{Pr}(\mathcal{E}_y \mid \gamma(F_\nu) \text{ is simple}) = \mathbf{Pr}(\mathcal{E}_y) + O(1/\nu). \tag{4.3}$$

But

$$\mathbf{Pr}(\mathcal{E}_y \mid \gamma(F_\nu) \text{ is simple}) = \frac{\mathbf{Pr}(\gamma(F_\nu) \text{ is simple} \mid \mathcal{E}_y)\mathbf{Pr}(\mathcal{E}_y)}{\mathbf{Pr}(\gamma(F_\nu) \text{ is simple})}$$

and so we need only show

$$\mathbf{Pr}(\gamma(F_\nu) \text{ is simple} \mid \mathcal{E}_y) = \mathbf{Pr}(\gamma(F_\nu) \text{ is simple}) + O(1/\nu)$$

or

$$\mathbf{Pr}(\gamma(F_\nu) \text{ is simple} \mid D) = \mathbf{Pr}(\gamma(F_\nu) \text{ is simple}) + O(1/\nu) \tag{4.4}$$

where D is the set of pairs of points deleted from F_ν in one step and D does not contain loops nor multiple edges. But $|D| \leq 5$ and (4.4) follows easily from (4.1).

We next write down the transition probabilities $p[x : y]$. Now many of the transition probabilities are small (i.e. $O(1/m)$) and do not have a significant effect on our analysis. We will only give transition probabilities up to this level of accuracy. Suppose that $n(t) = x$ and thus $2m = x_1 + 2x_2 + 3x_3$. Let $p_i = ix_i/2m$ for $i = 1, 2, 3$. The following table gives the significant transitions. There is an $O(1/m)$ term to be added to each probability. Any transitions not mentioned will have total probability $O(1/m)$ of occuring. (These are associated with triangles close to the chosen edge.) The probabilities in the table are only accurate for m sufficiently large. (Once m is small the remainder of the process cannot effect the number of uncovered vertices very much.) In Table 2 below, the first row corresponds to the initial state, the next seven rows correspond to cases where there are no vertices of degree one and the last ten rows cover the cases where there are! vertices of degree one.

These probabilities are derived by using (4.2). For example consider the transition from (x_1, x_2, x_3) to $(x_1 - 2, x_2 + 1, x_3 - 2)$. Here v in Step $\boldsymbol{A}$ is of degree 1 and u is of degree 3, which accounts for one factor p_3 in the transition probability. The other 2 neighbours of u are of degree 1 and 3 respectively. This accounts for the factor $2p_1p_3$. The remaining probabilities can be checked in the same manner.

x	y	$p[x:y]$
$(0,0,x_3)$	$(0,4,x_3-6)$	1
$(0,x_2,x_3)$	$(0,x_2+2,x_3-4)$	p_3^4
$(0,x_2,x_3)$	$(1,x_2,x_3-3)$	$3p_2p_3^3$
$(0,x_2,x_3)$	$(2,x_2-2,x_3-2)$	$3p_2^2p_3^2$
$(0,x_2,x_3)$	$(3,x_2-4,x_3-1)$	$p_2^3p_3$
$(0,x_2,x_3)$	$(0,x_2,x_3-2)$	$p_2p_3^2$
$(0,x_2,x_3)$	$(1,x_2-2,x_3-1)$	$2p_2^2p_3$
$(0,x_2,x_3)$	$(2,x_2-4,x_3)$	p_2^3
(x_1,x_2,x_3)	(x_1-1,x_2+2,x_3-3)	p_3^3
(x_1,x_2,x_3)	(x_1,x_2,x_3-2)	$2p_2p_3^2$
(x_1,x_2,x_3)	(x_1+1,x_2-2,x_3-1)	$p_2^2p_3$
(x_1,x_2,x_3)	(x_1-1,x_2-1,x_3-1)	$2p_1p_2p_3$
(x_1,x_2,x_3)	(x_1-2,x_2+1,x_3-2)	$2p_1p_3^2$
(x_1,x_2,x_3)	(x_1-3,x_2,x_3-1)	$p_1^2p_3$
(x_1,x_2,x_3)	(x_1-1,x_2,x_3-1)	p_2p_3
(x_1,x_2,x_3)	(x_1,x_2-2,x_3)	p_2^2
(x_1,x_2,x_3)	(x_1-2,x_2-1,x_3)	p_1p_2
(x_1,x_2,x_3)	(x_1-2,x_2,x_3)	p_1

Table 2: Transition probabilities for the chain $\mathcal{M}_1$

We continue by computing the expected number of new isolated vertices created in a single transition (conditional on the present state of $\mathcal{M}_1$ being x). Denote this by $\iota(x)$. Except in rare cases (i.e. of probability $O(1/m)$) isolated vertices are produced only when $x_1 > 0$. The following table gives the number created in each case. (The case numbers refer to the row numbers of Table 2.)

9	10	11	12	13	14	15	16	17	18
0	0	0	1	1	2	0	0	1	0

Table 3: Number of isolated vertices created in transitions of $\mathcal{M}_1$

Observe that with probability 1, the number of isolated vertices created in each transition equals $O(1)$. Hence, we obtain from the above table that

$$\begin{aligned}\iota(x) &= 2p_1p_2p_3 + 2p_1p_3^2 + 2p_1^2p_3 + p_1p_2 + O(1/m)\\ &= p_1(p_2+2p_3) + O(1/m)\\ &= p_1 + p_1(p_3 - p_1) + O(1/m),\end{aligned}$$

and hence that

$$\mathbf{E}[\lambda(G_n)] \quad \leq \quad 2\sum_{t\geq 0}\mathbf{E}\left(\frac{n_1(t)}{m(t)}\right) + O\left(\sum_{t\geq 0}\frac{1}{m(t)}\right)$$

$$= \quad 2\sum_{t\geq 0} \mathbf{E}\left(\frac{n_1(t)}{m(t)}\right) + O(\ln n). \tag{4.5}$$

In studying the lower bound, we will show, for suitable $t_0 < t_1$, that p_1 is negligible when compared with p_3 whenever t is between t_0 and t_1. This gives

$$\mathbf{E}[\lambda(G_n)] \geq (1 - o(1)) \sum_{t=t_0}^{t_1} \mathbf{E}\left(\frac{n_1(t)}{m(t)}\right) - O(\ln n). \tag{4.6}$$

5 Behaviour of n_3

Note that for each edge $\{u, v\}$ removed by MINGREEDY, one of the end-points, say u, is picked from the vertices of minimal (but non-zero) degrees, or u is determined from a previous edge removal. The other end-point v is chosen randomly from the neighbours of u. Now since almost all cubic graphs are connected, each decrease in n_3, except for the first edge removal, is accounted for exactly once as the end-point v whenever v is of degree 3. (Here, we regard n_3 as function of the number m of edges in the current graph.) Now consider the edge removal when the current graph has n_3 vertices of degree 3 and m edges. The probability that the end-point v in the edge removed is of degree 3 equals $3n_3/(2m)(1 + O(1/m))$ (from applying arguments similar to those used in showing (4.2). Thus the rate of change in n_3 with respect to m should be approximately

$$\frac{dn_3}{dm} \approx \frac{3n_3}{2m},$$

which gives the approximation stated in (2.1). These ideas are made rigorous by the following lemma.

Lemma 5.1 *For any fixed $\epsilon > 0$,*

$$\mathbf{Pr}\left(\exists t \textit{ such that } m(t) \geq n^{1/2}\ln^3 n \textit{ and} \left| n_3\sqrt{n}\left(\frac{3}{2m}\right)^{3/2} - 1\right| \geq \epsilon\right) = O(n^{-2}).$$

Proof. We note first that MINGREEDY destroys edges of G sequentially. Let $h = \lfloor n^{1/4} \rfloor$, and for $i = 0, 1, 2, \ldots$, define

$$m_i = \frac{3n}{2} - ih.$$

Let z_i be the number of vertices with degree three in G at the first time when $m \leq m_i$, and let $\mathcal{E}_i$ be the event that

$$z_i\sqrt{n}\left(\frac{3}{2m_i}\right)^{3/2} - 1 = O\left(\sum_{j=0}^{i-1} \frac{n^{3/8}\ln^{1/2} n}{m_j^{5/4}}\right). \tag{5.1}$$

We shall show that for i such that $m_i \geq n^{1/2} \ln^3 n$,

$$\mathbf{Pr}\left(\exists j \leq i \text{ such that } \bar{\mathcal{E}}_j\right) = O(i/n^4) \tag{5.2}$$

(All big O terms in this section are uniform over i.) Equation (5.2) is trivially true for $i = 0$. Assume as induction hypothesis that equation (5.2) holds for smaller values of i. Suppose next that $m_i \geq n^{1/2} \ln^3 n$. We need to show that

$$\mathbf{Pr}_0(\bar{\mathcal{E}}_i) = O(1/n^4), \tag{5.3}$$

where $\mathbf{Pr}_0$ is the probability conditional on $\mathcal{F}_i = \mathcal{E}_1 \cap \ldots \cap \mathcal{E}_{i-1}$. Let

$$\Delta z_i = z_{i-1} - z_i.$$

Recall that for each edge $\{x, y\}$ removed by MINGREEDY, one of the end-points, say x, is either chosen randomly from vertices of minimum degree or determined by a previous edge removal, while the other end-point, say y, is a degree three vertex with probability $p_3 + O(1/m)$. Then conditional on $\mathcal{F}_i$ and for $m_i \leq m \leq m_{i-1}$, the probability that the vertex y in an edge removal is of degree three is

$$\begin{aligned} p &= \frac{3z_{i-1} + O(h)}{2m_i + O(h)} = z_{i-1}\left(\frac{3}{2m_i} + O\left(\frac{h}{z_{i-1}m_{i-1}}\right)\right) \\ &= z_{i-1}\left(\frac{3}{2m_{i-1}} + O\left(\frac{n^{3/4}}{m_{i-1}^{5/2}}\right)\right). \end{aligned}$$

Thus, if $\Delta' z_i$ is the number of degree three vertices that appear as vertices y in the edge removals during the period when m decreases from m_{i-1} to m_i, then using Chernoff bounds,

$$\mathbf{Pr}_0\left(\mid \Delta' z_i - hp \mid \geq \sqrt{12hp \ln n}\right) = O(1/n^4).$$

However, $\Delta z_i = \Delta' z_i + \eta$, where η denotes the number of degree three vertices that appear as vertices x in the edge removals. Now η is at most one plus the number of components of G_n. Hence we may assume for all i that $\eta \leq 10$, which incurs an error probability of $O(1/n^4)$. Since $hp \to \infty$, we have that

$$\mathbf{Pr}_0\left(\mid \Delta z_i - hp \mid \geq \sqrt{12hp \ln n}\right) = O(1/n^4),$$

giving with conditional probability $1 - O(1/n^4)$ that

$$\begin{aligned} \frac{\Delta z_i}{z_{i-1}} &= \frac{3n^{1/4}}{2m_{i-1}} + O\left(\frac{n}{m_{i-1}^{5/2}}\right) + O\left(\sqrt{\frac{n^{1/4} \ln n}{m_{i-1} z_{i-1}}}\right) \\ &= \frac{3n^{1/4}}{2m_{i-1}} + O\left(\frac{n}{m_{i-1}^{5/2}}\right) + O\left(\sqrt{\frac{n^{3/4} \ln n}{m_{i-1}^{5/2}}}\right) \\ &= \frac{3n^{1/4}}{2m_{i-1}} + O\left(\frac{n^{3/8} \ln^{1/2} n}{m_{i-1}^{5/4}}\right). \end{aligned}$$

As $z_i = z_{i-1} - \Delta z_i$, we have with conditional probability at least $1 - O(1/n^4)$,

$$
\begin{aligned}
z_i &= z_{i-1} - \Delta z_i \\
&= z_{i-1}\left(1 - \frac{3n^{1/4}}{2m_{i-1}} + O\left(\frac{n^{3/8}\ln^{1/2} n}{m_{i-1}^{5/4}}\right)\right) \\
&= z_{i-1}\left(\frac{m_i}{m_{i-1}}\right)^{3/2}\left(1 + O\left(\frac{n^{3/8}\ln^{1/2} n}{m_{i-1}^{5/4}}\right)\right) \\
&= \frac{1}{\sqrt{n}}\left(\frac{2m_i}{3}\right)^{3/2}\left(1 + O\left(\sum_{j=0}^{i-1}\frac{n^{3/8}\ln^{1/2} n}{m_j^{5/4}}\right)\right).
\end{aligned}
$$

This completes our proof of (5.3) and hence of (5.2). Next, we note that for $m_i \geq n^{1/2}\ln^3 n$, i is at most $N = \lfloor \frac{3n - 2n^{1/2}\ln^3 n}{2h} \rfloor$ and that $\frac{3n}{2h} - N = \frac{n^{1/2}\ln^3 n}{h} + O(1)$. Hence the error term on the right hand side of (5.1) is at most

$$
\begin{aligned}
n^{3/8}\ln^{1/2} n \sum_{0\leq j\leq N}\frac{1}{(\frac{3}{2}n - jh)^{5/4}} &= \frac{n^{3/8}\ln^{1/2} n}{h^{5/4}}\sum_{0\leq j\leq N}\frac{1}{(\frac{3n}{2h} - j)^{5/4}} \\
&= O(n^{1/16}\ln^{1/2} n)\sum_{\frac{3n}{2h} - N \leq j \leq \frac{3n}{2h}}\frac{1}{j^{5/4}} \\
&= n^{1/16}\ln^{1/2} n\, O\left(\frac{1}{n^{1/16}\ln^{3/4} n}\right) \\
&= o(1).
\end{aligned}
$$

This shows that for any fixed $\epsilon > 0$,

$$
\mathbf{Pr}\left(\exists\, i \text{ such that } m = m_i \geq n^{1/2}\ln^3 n \text{ and } \left| n_3\left(\frac{3}{2}\right)^{3/2}\left(\frac{n^{1/2}}{m^{3/2}}\right) - 1\right| \geq \epsilon\right) = O(1/n^3).
$$

Similar arguments can be used to complete our proof of the lemma. □

6 An Upper Bound for $\mathbf{E}[n_1]$

In this and the following sections, we would like to estimate $\mathbf{E}[n_1]$ when the current G has m edges. Let t_m be the first time when $m(t) \leq m$. Note that $m - m(t_m)$ is bounded by an absolute constant. The big O and small o terms in this and the next sections are uniform in m. We find an upper bound for $\mathbf{E}[n_1(t_m)]$ in this section.

From Table 2, we have the following transition probabilities for $n_1(t)$ (where $\Delta n_1 = \Delta n_1(t) = n_1(t+1) - n_1(t)$ and $\mathbf{Pr}_1$ is the probability conditional on $n_1(t), n_2(t), n_3(t)$). When $n_1 = 0$ and $n_2 \neq 0$,

$$
\mathbf{Pr}_1(\Delta n_1 = i) = \begin{cases} \binom{3}{i}p_3^{4-i}p_2^{i} + \binom{2}{i}p_3^{2-i}p_2^{i+1} + O(1/m), & \text{if } i = 0,1,2,3, \\ 0, & \text{otherwise,} \end{cases}
$$

and when $n_1 \neq 0$,

$$\begin{aligned}
\mathbf{Pr}_1(\Delta n_1 = 1) &= p_2^2 p_3 + O(1/m), \\
\mathbf{Pr}_1(\Delta n_1 = 0) &= 2p_2 p_3^2 + p_2^2 + O(1/m), \\
\mathbf{Pr}_1(\Delta n_1 = -1) &= p_3^3 + 2p_1 p_2 p_3 + p_2 p_3 + O(1/m), \\
\mathbf{Pr}_1(\Delta n_1 = -2) &= p_1 p_2 + 2p_1 p_3^2 + p_1 + O(1/m), \\
\mathbf{Pr}_1(\Delta n_1 = -3) &= p_1^2 p_3 + O(1/m), \\
\mathbf{Pr}_1(\Delta n_1 = i) &= 0, \text{ if } i \neq 1, 0, -1, -2, -3.
\end{aligned}$$

Note that when both $n_1 = 0$ and $n_2 = 0$, we have

$$\begin{aligned}
\mathbf{Pr}_1(\Delta n_1 \neq 0) &= O(1/m), \\
\mathbf{Pr}_1(\Delta n_1 \leq 3) &= 1.
\end{aligned}$$

In particular, we may assume that no vertex of degree 1 is created in the first iteration of MINGREEDY; that is, $n_1(1) = 0$. Observe also that the $O(1/m)$ terms in the transition probabilities depend only on the current state of $\mathcal{M}_1$.

Let

$$\begin{aligned}
\alpha(x) &= x - 2x^2 + x^3 = (1-x)^2 x, \\
\gamma(x) &= x - x^2 + x^3 = x^3 + (1-x)x.
\end{aligned}$$

We next consider a process $X_1(t)$ with initial state $X_1(1) = 0$ and transition probabilities defined below.

$$\begin{aligned}
\mathbf{Pr}_1(\Delta X_1 = 1) &= \alpha(p_3) + O_1(1/m), \\
\mathbf{Pr}_1(\Delta X_1 = 0) &= 1 - (\alpha(p_3) + O_1(1/m)) - (\gamma(p_3) + O_2(1/m))\zeta(X_1), \\
\mathbf{Pr}_1(\Delta X_1 = -1) &= (\gamma(p_3) + O_2(1/m))\zeta(X_1), \\
\mathbf{Pr}_1(\Delta X_1 = i) &= 0, \text{ if } i \neq 1, 0, -1,
\end{aligned}$$

where $\zeta(x) = 1$ if $x > 0$ and zero otherwise. Note that when $n_1(t) = 0$, $\mathbf{Pr}(n_1(t+1) \leq 3) = 1$ and that when $n_1(t) \neq 0$, we have (ignoring the $O(1/m)$ terms)

$$\begin{aligned}
\mathbf{Pr}_1(\Delta n_1 = 1) &\leq \alpha(p_3), \\
\mathbf{Pr}_1(\Delta n_1 \leq -1) &\geq \gamma(p_3).
\end{aligned}$$

Thus, the big O terms $O_1(1/m)$ and $O_2(1/m)$ in the transition probabilities of the process $X_1(t)$ can be properly defined so that when $n_1(t) \neq 0$,

$$\begin{aligned}
\mathbf{Pr}_1(\Delta n_1 = 1) &\leq \mathbf{Pr}_1(\Delta X_1 = 1), \\
\mathbf{Pr}_1(\Delta n_1 \leq -1) &\geq \mathbf{Pr}_1(\Delta X_1 = -1).
\end{aligned}$$

It follows that n_1 and X_1 can be coupled so that n_1 is stochastically at most $X_1 + 3$ for all t. Next, let $X(t)$ be a process, with $X(1) = X_1(1)$, having the same transition probabilities, but without the big O terms, as those of $X_1(t)$. Note that the big O terms in the transition probabilities of $X_1(t)$ can only effect, with probability $1 - O(1/n^A)$ for any $A > 0$, at most $\ln^2 n$ transitions in $O(n)$ transitions of X_1. Now at each time t, conditional on X_1 having taken a transition not effected by the big O terms, we may couple X and X_1 so that $|X(t) - X_1(t)| \leq |X(t-1) - X_1(t-1)|$ (because of the reflecting barrier at the origin). It therefore follows that for all $t = O(n)$,

$$\mathbf{E}[X_1(t)] = \mathbf{E}[X(t)] + O(\ln^2 n).$$

Thus, $\mathbf{E}[n_1] = \mathbf{E}[X] + O(\ln^2 n)$. Since we have a fairly accurate estimate of n_3 as a function of m (and hence an estimate of $p_3 = 3n_3/2m$), we proceed to find a bound for X and hence one for $\mathbf{E}[n_1]$.

Lemma 6.1 *For $m \geq n^{1/2} \ln^3 n$ and for any fixed $\epsilon > 0$,*

$$\mathbf{E}[n_1(t_m)] \leq \frac{(1-p)^2}{p} + O(\ln^2 n),$$

where $p = (1-\epsilon)\,(2m/3n)^{1/2}$.

Proof. Consider a random walk X' with transition probabilities as follows. We write $\alpha' = \alpha(p), \gamma' = \gamma(p)$.

$$\begin{aligned}
\mathbf{Pr}(\Delta X' = 1) &= \alpha', \\
\mathbf{Pr}(\Delta X' = 0) &= 1 - \alpha' - \gamma'\zeta(X'), \\
\mathbf{Pr}(\Delta X' = -1) &= \gamma'\zeta(X').
\end{aligned}$$

Note that the steady state distribution π' of $X'(t)$ satisfies the following equations:

$$\begin{aligned}
\pi'_0 &= (1-\alpha')\pi'_0 + \gamma'\pi'_1, \\
\pi'_i &= \alpha'\pi'_{i-1} + (1-\alpha'-\gamma')\pi'_i + \gamma'\pi'_{i+1}, \quad i \geq 1, \\
\pi'_i &= 0, \quad i < 0.
\end{aligned}$$

These can be solved, giving

$$\begin{aligned}
\pi'_i &= 0, \quad i < 0, \\
\pi'_i &= \frac{\gamma' - \alpha'}{\gamma'} \left(\frac{\alpha'}{\gamma'}\right)^i, \quad i \geq 0,
\end{aligned} \tag{6.1}$$

with expectation $\sum_i i\pi'_i = (1-p)^2/p$. We shall assume that X' starts with its steady state distribution π', and hence the distribution of X' equals π' always.

We shall show later that if for all $t \leq t_m$

$$p_3(t) \geq (1-\epsilon)\left(\frac{2m(t)}{3n}\right)^{1/2} > p, \tag{6.2}$$

then for all $t \le t_m$ and for large n,

$$X(t) \le X'(t) \text{ in distribution.} \tag{6.3}$$

It follows from Lemma 5.1 that (6.2) holds with probability $1 - O(n^{-2})$, and since $n_1 \le n$, we have

$$\begin{aligned}
\mathbf{E}[n_1(t)] &= \mathbf{E}_2[n_1(t)] + O(1/n) \\
&\le \mathbf{E}_2[X(t)] + O(\ln^2 n) \\
&\le \mathbf{E}[X'(t)] + O(\ln^2 n) = (1-p)^2/p + O(\ln^2 n),
\end{aligned}$$

where $\mathbf{E}_2$ is the expectation conditional on p_3 satisfying (6.2). This proves the lemma.

We shall use induction to prove (6.3). This is trivial when $t = 1$. Assuming it is true for t, we shall bound the distribution of $X(t+1)$. Define a random variable $Y = X'(t) + \Delta Y$, where the distribution of ΔY is as follows. Writing $p_3 = p_3(t)$, define

$$\begin{aligned}
\mathbf{Pr}(\Delta Y = 1) &= \alpha(p_3), \\
\mathbf{Pr}(\Delta Y = 0) &= 1 - \alpha(p_3) - \gamma(p_3)\zeta(X'), \\
\mathbf{Pr}(\Delta Y = -1) &= \gamma(p_3)\zeta(X').
\end{aligned}$$

We shall show that

$$X(t+1) \le Y \le X'(t+1) \tag{6.4}$$

holds in distribution. The first inequatlity comes from the fact that we can couple $X(t)$ and $X'(t)$ so that $X(t) \le X'(t)$ (this uses the induction hypothesis). For the case where $X(t) = X'(t)$, we see that $\Delta X(t) = \Delta Y$ in distribution; for the case $X(t) < X'(t)$, we use the fact that $\Delta X(t)$ and ΔY can be coupled so that $\Delta X(t) \le \Delta Y + 1$. The first inequality thus holds for both cases. To show the second inequality, we consider the distribution π of Y. Writing $\alpha = \alpha(p_3(t)), \gamma = \gamma(p_3(t)), \alpha' = \alpha(p), \gamma' = \gamma(p)$, we have the following equations

$$\begin{aligned}
\pi_0 &= (1-\alpha)\pi'_0 + \gamma\pi'_1, \\
\pi_i &= \alpha\pi'_{i-1} + (1-\alpha-\gamma)\pi'_i + \gamma\pi'_{i+1}, \quad i \ge 1, \\
\pi_i &= 0, \quad i < 0.
\end{aligned}$$

Hence for $i \ge 1$,

$$\begin{aligned}
\pi_i &= \pi'_i\left(\frac{\gamma'\alpha}{\alpha'} + 1 - \gamma - \alpha + \frac{\gamma\alpha'}{\gamma'}\right) \\
&= \pi'_i\left(1 + (\gamma' - \alpha')\frac{\gamma}{\alpha'}\left(\frac{\alpha}{\gamma} - \frac{\alpha'}{\gamma'}\right)\right).
\end{aligned} \tag{6.5}$$

Note that the functions $\alpha(x)$ and $\gamma(x)$ are such that for all $x \in [0,1]$,

$$\gamma(x) - \alpha(x) \ge 0,$$

and

$$\frac{d}{dx}\frac{\alpha(x)}{\gamma(x)} = \frac{-1+x^2}{(1-x+x^2)^2} \leq 0.$$

It therefore follows from (6.2) and (6.5) that for $i \geq 1$,

$$\pi_i \leq \pi'_i,$$

giving that for $i \geq 1$,

$$\sum_{j\geq i} \pi_i \leq \sum_{j\geq i} \pi'_i.$$

As $\pi_i = \pi'_i = 0$ for $i < 0$, the distribution π is smaller than π'. This shows the second inequality in (6.4), and so our proof of the lemma is complete. □

Lemma 6.1 gives that for $m \geq n^{1/2}\ln^3 n$,

$$\mathbf{E}[n_1(t_m)] = O((n/m)^{1/2}).$$

Hence, if $m \geq m_0 = n^{3/5}$, then $\mathbf{E}[n_1(t_m)] = O(n^{1/5})$. The next lemma shows that $\mathbf{E}[n_1(t_m)] = O(n^{1/5})$ holds when $m < m_0$ too.

Lemma 6.2 $\mathbf{E}[n_1(t_m)] = O(n^{1/5})$ *for all* $m \leq m_0$.

Proof. As argued in Lemma 6.1, the assertion in this lemma follows if we can show that for all $m < m_0$,

$$\mathbf{E}[X(t_m)] = O(n^{1/5}). \tag{6.6}$$

Given p_3, consider a random walk Z with transition probability defined as below.

$$\begin{aligned}
\mathbf{Pr}(\Delta Z = 1) &= (\alpha(p_3)+\gamma(p_3))(1-\zeta(Z)/2),\\
\mathbf{Pr}(\Delta Z = 0) &= 1-\alpha(p_3)-\gamma(p_3),\\
\mathbf{Pr}(\Delta Z = -1) &= (\alpha(p_3)+\gamma(p_3))\zeta(Z)/2.
\end{aligned}$$

Write $t_0 = t_{m_0}$. We consider the processes X and Z after time t_0. Suppose $Z(t_0) = X(t_0)$. As $\alpha(p_3) \leq \gamma(p_3)$ for all $p_3 \in [0,1]$, we can couple the processes $X(t)$ and $Z(t)$ for all $t \geq t_0$ so that Z is stochastically at least X. Hence (6.6) follows if

$$\mathbf{E}[Z(t)] = O(n^{1/5}), \qquad \forall t \geq t_0. \tag{6.7}$$

Since Z is a symmetric random walk with a reflecting barrier at $Z = 0$, the reflection principle shows that

$$\mathbf{E}[Z(t)] = \mathbf{E}[|\,Z'(t)\,|]$$

where Z' is the same symmetric random walk as Z but without the reflecting barrier. Next, let Z'' be a random walk with the same transition probabilities as those of Z' but with $Z''(t_0) = 0$. Then it is clear that $Z'(t) = Z''(t) + Z'(t_0)$ and so

$$\begin{aligned}\mathbf{E}[|\,Z'(t)\,|] &\leq \mathbf{E}[|\,Z''(t)\,|] + \mathbf{E}[Z'(t_0)] \\ &= \mathbf{E}[|\,Z''(t)\,|] + \mathbf{E}[X(t_0)].\end{aligned}$$

Let M be the number of moves that Z'' makes after t_0 until the end of algorithm MIN-GREEDY. Then we have

$$\mathbf{E}[|\,Z''(t)\,|] = O(\mathbf{E}[\sqrt{M}]),$$

giving that

$$\mathbf{E}[Z(t)] = O(\mathbf{E}[\sqrt{M}] + \mathbf{E}[X(t_0)]). \tag{6.8}$$

We next estimate M. Let $m_1 = n^{1/2} \ln^3 n$ and $m_2 = n^{2/5}$. We write

$$M = M_1 + M_2 + M_3,$$

where M_1 (respectively M_2) is the number of moves that Z'' makes when m is between m_0 and m_1 (respectively between m_1 and m_2), and M_3 is the number of moves when $m \leq m_2$. Then clearly

$$M_3 \leq n^{2/5}.$$

To estimate M_1, note that with error probability $O(1/n^2)$, we can assume that for fixed $\epsilon > 0$, $p_3 \leq (1+\epsilon)(2m/3n)^{1/2} = O((2m_0/3n)^{1/2}) = O(n^{-1/5})$ for m between m_0 and m_1. Thus

$$\begin{aligned}\mathbf{E}[M_1] &= \sum(\alpha(p_3(t)) + \gamma(p_3(t))) \\ &= O(n^{3/5} n^{-1/5}) = O(n^{2/5}).\end{aligned}$$

For M_2, we note that according to Lemma 5.1, when $m = m_1$, n_3 is of order $n^{1/4} \ln^5 n$ and n_3 decreases as m decreases. Hence when m is between m_1 and m_2, we have

$$p_3 = O(n^{1/4} \ln^5 n / n^{2/5}) = O(n^{-0.14}),$$

and so

$$\begin{aligned}\mathbf{E}[M_2] &= \sum(\alpha(p_3(t)) + \gamma(p_3(t))) \\ &= O(n^{1/2} n^{-0.14}) = O(n^{2/5}).\end{aligned}$$

As M_1 and M_2 are sums of independent Bernoulli variables with probability of success $\alpha(p_3(t)) + \gamma(p_3(t))$ (when the sequence p_3 is known), they are concentrated near their means. Hence, we see that

$$M = O(\mathbf{E}[M_1]) + O(\mathbf{E}[M_2]) + M_3 = O(n^{2/5}),$$

almost surely. This gives that

$$\mathbf{E}[\sqrt{M}] = O(n^{1/5}).$$

Equation (6.7) now follows from (6.8) and the fact that $\mathbf{E}[X(t_0)] = O(n^{1/5})$. This completes our proof of the lemma. □

We are now in a position to prove the upper bound in the theorem. Using (4.5), we have

$$\begin{aligned}
\mathbf{E}[\lambda(G_n)] &\leq 2\sum_{t\geq 0} \mathbf{E}[n_1(t)/m(t)] + O(\ln n) \\
&\leq 2\sum_{m=1}^{3n/2} \mathbf{E}[n_1(t_m)/m] + O(\ln n) \\
&= 2\sum_{m=1}^{3n/2} \mathbf{E}[n_1(t_m)]/m + O(\ln n) \\
&= O(n^{1/5})\sum_{m=1}^{3n/2} 1/m + O(\ln n) \\
&= O(n^{1/5}\ln n).
\end{aligned}$$

7 A Lower Bound for $\mathbf{E}[n_1]$

Let $B \geq 3000$ be a constant, and let $m_B = Bn^{3/5}$. Write t_B be the first time at which $m(t) \leq m_B$. Fix a constant $A > B$ and a small constant $\epsilon > 0$. (The numbers A, ϵ will be chosen later.) Let $m_A = An^{3/5}$. Define t_A as the time immediately after $m \leq m_A$. Hence $t_A = t_{m_A}$ and $t_B = t_{m_B}$. Let $N = \lfloor \epsilon n^{1/5} \rfloor$ and $r = (\epsilon n^{1/5} + 3)/m_B \approx \epsilon n^{-2/5}/B$. We shall find a lower bound for $\mathbf{E}[n_1(t_B)]$ by considering a random walk with reflecting barriers at 0 and N as follows.

Suppose for now that the sequence $p_3(t), t \in [t_A, t_B]$ is known. Consider a random walk $W(t), t \in [t_A, t_B]$ with initial state $W(t_A) = 0$ and transition probabilities given below. With functions α, γ and ζ as defined before and writing $\alpha = \alpha(p_3)$ and $\gamma = \gamma(p_3)$,

$$\begin{aligned}
\mathbf{Pr}(\Delta W = 1) &= \alpha\zeta_N(W) \\
\mathbf{Pr}(\Delta W = -1) &= \gamma\zeta(W), \\
\mathbf{Pr}(\Delta W = 0) &= 1 - \alpha\zeta_N(W) - \gamma\zeta(W),
\end{aligned}$$

where $\zeta_N(x) = 1$ if $x < N$ and zero otherwise. Associated with W is a process U where $U(t_A) = 0$ and U is decreased by 3 with probability $9r$ at each step. We claim that if the processes n_1, W and U are "suitably" defined in a probability space, then

$$n_1(t) \geq W(t) + U(t), \qquad t \in [t_A, t_B]. \tag{7.1}$$

Equation (7.1) holds for $t = t_A$ trivially. Assume that (7.1) holds for t. We want to show that it holds for $t+1$ also. Now if $n_1(t) \geq N + 3$, then (7.1) holds for $t+1$ as $W \leq N$ and $\Delta n_1 \leq 3$ always. Thus, we assume that $n_1(t) < N + 3$ and like to show next that if W and

U are properly defined, then $\Delta n_1 \geq \Delta W + \Delta U$. Note first that if $n_1(t) < N+3$, then $p_1 \leq r$ from our definition of r. From the transition probabilities of n_1 , we see that (ignoring the $O(1/m)$ terms)
Case 1 $\Delta n_1 = 1$ with probability at least $2p_2^2 p_3 \geq \alpha(p_3) - 2p_1 \geq \alpha(p_3) - 2r$;
Case 2 $\Delta n_1 = -1$ with probability at most $\gamma(p_3) + 2p_1 \leq \gamma(p_3) + 2r$;
Case 3 $\Delta n_1 = -2$ with probability at most $4p_1 \leq 4r$ and $\Delta n_1 = -3$ with probability at most $p_1 \leq r$.
Thus, very crudely, we see that n_1 and W can be coupled so that $\Delta n_1 \geq \Delta W$ with probability at least $1-9r$ and $\Delta n_1 \geq \Delta W - 3$ always. (The big O terms in the transition probabilities of n_1 equal $O(n^{-3/5})$ and are negligible when compared with $r = O(n^{-2/5})$.) This establishes (7.1). Note that

$$\mathbf{E}[U(t_B)] = -3(t_B - t_A)(9r)(1+o(1)) \geq -27r(A-B)n^{3/5}(1+o(1)) \approx -\frac{27(A-B)}{B}\epsilon n^{1/5},$$

and so for $t \in [t_A, t_B]$,

$$\mathbf{E}[n_1(t_B)] \geq \mathbf{E}[W(t_B)] - \frac{27(A-B)}{B}\epsilon n^{1/5}(1+o(1)). \tag{7.2}$$

We next bound $\mathbf{E}[W(t_B)]$. Let D be a probability distribution with support $[N]$, and let $W_D(t)$ be a random walk with the same transition probabilities as those of $W(t)$, but $W_D(t_A)$ has distribution D. Since $W(t_A) = 0$, W and W_D can be coupled so that $W_D(t) \geq W(t)$ for all $t \geq t_A$ and that $W_D(t) - W(t)$ is non-increasing in t. Let T be the first time that $W(t) = W_D(t)$. Then

$$W(t_B) \geq W_D(t_B)\chi(\{T \leq t_B\}),$$

where $\chi(\{T \leq t_B\})$ is the indicator function for the event $\{T \leq t_B\}$. Since $W \leq N$, we have

$$\mathbf{E}[W(t_B)] \geq \mathbf{E}[W_D(t_B)] - N\mathbf{Pr}(T > t_B).$$

Let T' be the first time that $W_D = 0$. Then clearly $T \leq T'$ in distribution as $W_D(t) \geq W(t) \geq 0$. Thus,

$$\mathbf{E}[W(t_B)] \geq \mathbf{E}[W_D(t_B)] - N\mathbf{Pr}(T' > t_B). \tag{7.3}$$

According to Lemma 5.1, with probability $1 - O(n^{-2})$, we have that for t between t_A and t_B and fixed $\epsilon_1 > 0$,

$$(1-\epsilon_1)\left(\frac{2m_B}{3n}\right)^{1/2} \leq p_3(t) \leq (1+\epsilon_1)\left(\frac{2m_A}{3n}\right)^{1/2}. \tag{7.4}$$

The assumption of (7.4) only gives an extra factor of $1 - O(n^{-2})$ to our estimate of $\mathbf{E}[n_1]$. Let $q_0 = (1+\epsilon_1)\left(\frac{2m_A}{3n}\right)^{1/2}$ and $q_1 = (1-\epsilon_1)\left(\frac{2m_B}{3n}\right)^{1/2}$. We shall assume that (7.4) is satisfied.

We next define the distribution D and bound W_D suitably. Write $\alpha' = \alpha(q_0)$ and $\gamma' = \gamma(q_0)$. Consider a random walk W' with transition probabilities given by

$$\begin{aligned}
\mathbf{Pr}(\Delta W' = 1) &= \alpha'\zeta_N(W'), \\
\mathbf{Pr}(\Delta W' = -1) &= \gamma'\zeta(W'), \\
\mathbf{Pr}(\Delta W' = 0) &= 1 - \alpha'\zeta_N(W') - \gamma'\zeta(W').
\end{aligned}$$

Note that here α' and γ' no longer depend on t and that the steady state distribution λ' of W' satisfies

$$\begin{aligned}
\lambda'_0 &= (1-\alpha')\lambda'_0 + \gamma'\lambda'_1 \\
\lambda'_i &= \alpha'\lambda'_{i-1} + (1-\alpha'-\gamma')\lambda'_i + \gamma'\lambda'_{i+1}, i = 1, 2, \ldots, N-1 \\
\lambda'_N &= (1-\gamma')\lambda'_N + \alpha'\lambda'_{N-1} \\
\lambda'_i &= 0, \quad \forall i < 0,
\end{aligned}$$

from which we obtain

$$\begin{aligned}
\lambda'_i &= \frac{1-(\alpha'/\gamma')}{1-(\alpha'/\gamma')^{N+1}}\left(\frac{\alpha'}{\gamma'}\right)^i, \quad i = 0, 1, \ldots, N \\
\lambda'_i &= 0, \quad \forall i < 0.
\end{aligned} \tag{7.5}$$

We assume that $W'(t_A)$ has distribution λ' and so the distribution of $W'(t)$ equals λ' always.

Suppose now that $W_D(t)$ starts with distribution λ' at time t_A (that is, $D = \lambda'$). We shall show by induction that $W_D(t)$ is at least $W'(t)$ in distribution. For this, we follow the method used in showing $X \leq X'$ in distribution in our proof of Lemma 6.1. Assume p_3 satisfies (7.4) and that $W_D(t) \geq W'(t)$ as our induction hypothesis. Let Y be the state of a process which starts with distribution λ' (which equals the distribution of $W'(t)$) followed by a one-step transition with transition probabilities as those of W_D. We claim that

$$W_D(t+1) \geq Y \geq W'(t+1). \tag{7.6}$$

The justification for the first inequality here follows similar arguments as those used in showing the first inequality in (6.4). To show the second inequality in (7.6), use λ to denote the distribution of Y. Writing $\alpha = \alpha(p_3(t))$ and $\gamma = \gamma(p_3(t))$, we find that $\lambda_i = 0$ for all $i < 0$ and for i between 1 and $N-1$ we have

$$\begin{aligned}
\lambda_0 &= (1-\alpha)\lambda'_0 + \gamma\lambda'_1, \\
\lambda_i &= \alpha\lambda'_{i-1} + (1-\alpha-\gamma)\lambda'_i + \gamma\lambda'_{i+1} \\
\lambda_N &= (1-\gamma)\lambda'_N + \alpha\lambda'_{N-1}.
\end{aligned}$$

Note that, since $q_0 \geq p_3(t)$, we have $\alpha'/\gamma' \leq \alpha/\gamma$. Also

$$\lambda_N = \lambda'_N + \alpha\lambda'_{N-1} - \gamma\lambda'_N = \lambda'_N + \gamma\left(\frac{\alpha}{\gamma} - \frac{\alpha'}{\gamma'}\right)\lambda'_{N-1} \geq \lambda'_N,$$

and by following (6.5), we have that for $i = 1, 2, \ldots, N-1$,

$$\lambda_i = \lambda'_i\left(1 + (\gamma'-\alpha')\frac{\gamma}{\alpha'}\left(\frac{\alpha}{\gamma} - \frac{\alpha'}{\gamma'}\right)\right) \geq \lambda'_i.$$

As $\lambda'_i = \lambda_i = 0$ for $i < 0$, the distribution λ is at least λ'. This completes the induction step, and so we conclude that the distribution of $W_D(t)$ is at least λ'. It therefore follows that for $t \in [t_A, t_B]$,

$$\mathbf{E}[W_D(t)] \geq \sum_i i\lambda'_i = \frac{1}{1-(\alpha'/\gamma')} - \frac{N+1}{1-(\alpha'/\gamma')^{N+1}} + N.$$

Since $q_0 \approx (1+\epsilon_1)(2A/3)^{1/2} n^{-1/5}$ and $(\alpha'/\gamma')^N \approx \exp(-q_0 N)$, we have

$$\mathbf{E}[W_D(t)] \geq \frac{n^{1/5}}{1+\epsilon_1}\sqrt{\frac{3}{2A}} - \epsilon n^{1/5} \frac{\exp(-(1+\epsilon_1)\epsilon(2A/3)^{1/2})}{1-\exp(-(1+\epsilon_1)\epsilon(2A/3)^{1/2})}.$$

By choosing $\epsilon_1 > 0$ such that $(3/2)^{1/2}/(1+\epsilon_1) = 1.22$, we have

$$\mathbf{E}[W_D(t_B)] \geq \frac{1.22 n^{1/5}}{A^{1/2}} - \epsilon n^{1/5} \frac{\exp(-\epsilon A^{1/2}/1.22)}{1-\exp(-\epsilon A^{1/2}/1.22)}. \tag{7.7}$$

To estimate T', we write $\alpha'' = \alpha(q_1)$ and $\gamma'' = \gamma(q_1)$, and let λ'' be the distribution given in (7.5) with α', γ' replaced with α'', γ'' respectively. Let W'' be the same random walk as W' but with α', γ' replaced with α'', γ'' and let $W''(t_A)$ have distribution λ'' instead of λ'. Note that as $q_0 \geq q_1$, we have $\alpha'/\gamma' \leq \alpha''/\gamma''$, and it is not difficult to check that distribution λ' is at most λ''. Hence, our previous argument for showing $W' \leq W_D$ in distribution can be used to show $W_D \leq W''$ in distribution. Next, let W_1 be the walk whose transition probabilities are as those of W'' but $W_1(t_A) = N$ and without the reflecting barrier at $W_1 = N$. A simple coupling argument gives that $W'' \leq W_1$ in distribution. Let T_1 be the first time that $W_1 = 0$. Then we have

$$T' \leq T_1 \quad \text{in distribution.}$$

We next estimate T_1. Let L be the time elapsed before W_1 first gets to $N-1$. Then we have

$$T_1 = L_1 + \ldots + L_N + t_A,$$

where $L_1, \ldots, L_N$ are independent variables identically distributed as L. Note that
(I) with probability γ'', $L = 1$;
(II) with probability α'', $L = 1 + L' + L''$;
(III) with probability $1 - \alpha'' - \gamma''$, $L = 1 + L'''$,
where L', L'', L''' are independent and identically distributed as L. Hence, writing $M(z) = \mathbf{E}[z^L]$ as the probability generating function of L, we have

$$M(z) = \gamma'' z + \alpha'' z M(z)^2 + (1 - \alpha'' - \gamma'') z M(z),$$

giving

$$M(z) = \frac{-(z - \alpha'' z - \gamma'' z - 1) - \sqrt{(z - \alpha'' z - \gamma'' z - 1)^2 - 4\alpha''\gamma'' z^2}}{2\alpha'' z}.$$

Put $z = \exp(\kappa q_1^3)$, for some $\kappa < 1/4$ to be chosen later. Then

$$M(z) = 1 + (1 - \sqrt{1-4\kappa}) q_1/2 + O(q_1^2).$$

Since each step of MINGREEDY destroys at most 5 edges,

$$\begin{aligned} \mathbf{Pr}(T' \geq t_B) &\leq \mathbf{Pr}(T_1 - t_A \geq t_B - t_A) \\ &\leq \mathbf{Pr}(L_1 + \ldots + L_N \geq (A-B) n^{3/5}/5) \end{aligned}$$

$$
\begin{aligned}
&\leq \mathbf{E}[\exp((L_1+\ldots+L_N)\ln z-(A-B)n^{3/5}\ln z/5)]\\
&\leq M(z)^N\exp(-(A-B)(\ln z)n^{3/5}/5)\\
&\approx \exp((1-\sqrt{1-4\kappa})Nq_1/2-(A-B)\kappa q_1^3 n^{3/5}/5)\\
&\approx \exp((1-\sqrt{1-4\kappa})\epsilon(1-\epsilon_1)(B/6)^{1/2}-(A-B)\kappa(1-\epsilon_1)^3(2B/3)^{3/2}/5).
\end{aligned}
$$

Put $\kappa=0.249$ and $\epsilon_1=0.00001$, we obtain

$$\mathbf{Pr}(T'\geq t_B)\leq \exp(0.39\epsilon B^{1/2}-0.027(A-B)B^{3/2}). \tag{7.8}$$

Hence, assuming (7.4), we have from (7.2), (7.3) and (7.7) that

$$
\begin{aligned}
\mathbf{E}[n_1(t_B)] \;\geq\; & \frac{1.22n^{1/5}}{A^{1/2}}-\epsilon n^{1/5}\frac{\exp(-\epsilon A^{1/2}/1.22)}{1-\exp(-\epsilon A^{1/2}/1.22)}-\\
& \epsilon n^{1/5}\exp(0.39\epsilon B^{1/2}-0.027(A-B)B^{3/2})-\frac{27(A-B)}{B}\epsilon n^{1/5}(1+o(1)).
\end{aligned}
$$

We now set $A=B+B^{1/2}$, $\epsilon=1.22x/A^{1/2}$, where $x=0.04$. Then

$$
\begin{aligned}
\mathbf{E}[n_1(t_B)] &\geq \frac{1.22n^{1/5}}{(B+\sqrt{B})^{1/2}}\left(1-\frac{x\mathrm{e}^{-x}}{1-\mathrm{e}^{-x}}-x\exp(0.5x-0.027B^2)-\frac{27x(1+o(1))}{\sqrt{B}}\right)\\
&\geq \frac{10^{-4}n^{1/5}}{(B+\sqrt{B})^{1/2}},
\end{aligned} \tag{7.9}
$$

since $B\geq 3000$. We have therefore showed the following lemma.

Lemma 7.1 *For any constant $B\geq 3000$, if $m=Bn^{3/5}$, then*

$$\mathbf{E}[n_1(t_m)]\geq \frac{10^{-4}n^{1/5}}{(B+\sqrt{B})^{1/2}}(1-o(1)).$$

Proof. Let $\mathcal{E}$ be the event that (7.4) occurs. Then

$$\mathbf{E}[n_1(t_m)]\geq \mathbf{E}_2[n_1(t_m)](1-O(1/n^2)),$$

where $\mathbf{E}_2$ is the expectation conditional on $\mathcal{E}$. It follows from (7.9) that

$$\mathbf{E}[n_1(t_m)]\geq \frac{10^{-4}n^{1/5}}{(B+\sqrt{B})^{1/2}}(1-o(1)).$$

□

To prove the lower bound in the theorem, we let $B_0=3600$ and $B_1=3000$. Write $m_0=B_0n^{3/5}$, $m_1=B_1n^{3/5}$, $t_0=t_{m_0}$ and $t_1=t_{m_1}$ where t_m is (as before) the greatest t

such that $m(t) \leq m$. Note that when t is between t_0 and t_1, we have $n_3 = \Omega(n^{2/5})$ and $n_1 = O(n^{1/5})$, which gives $p_3 > p_1$ almost surely. Then using (4.6), we have

$$\begin{aligned}
\mathbf{E}[\lambda(G_n)] &\geq \sum_{t \geq 0} \mathbf{E}[p_1(t)](1 - o(1)) - O(\ln n) \\
&\geq \sum_{t \geq 0} \mathbf{E}[n_1(t)/m(t)](1 - o(1)) - O(\ln n) \\
&\geq \sum_{t=t_0}^{t_1} \mathbf{E}[n_1(t)/m(t)](1 - o(1)) - O(\ln n) \\
&\geq \sum_{t=t_0}^{t_1} \mathbf{E}[n_1(t)]/m_0(1 - o(1)) - O(\ln n) \\
&\geq \frac{t_1 - t_0}{m_0} \frac{10^{-4} n^{1/5}}{(B_0 + \sqrt{B_0})^{1/2}} - O(\ln n) \\
&\geq \frac{m_0 - m_1}{5m_0} \frac{10^{-4} n^{1/5}}{\sqrt{3660}} (1 - o(1)) - O(\ln n) \\
&= \Omega(n^{1/5}).
\end{aligned}$$

Acknowledgement. We would like to thank the referee for a careful reading of the manuscript.

References

[1] E. A. Bender and E. R. Canfield, The asymptotic number of labelled graphs with given degree sequences, *Journal of Combinatorial Theory, Series A*, 24 (1978) 296-307.

[2] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *European Journal on Combinatorics*, 1 (1980) 311-316.

[3] M. E. Dyer and A. M. Frieze, Randomized greedy matching, *Random Structures and Algorithms*, 2 (1991) 29-45.

[4] M. E. Dyer, A. M .Frieze and B. Pittel, On the average performance of the greedy matching algorithm, to appear.

[5] O. Goldschmidt and D. S. Hochbaum, A fast perfect-matching algorithm in random graphs, *SIAM Journal on Discrete mathematics*, 3 (1990) 48-57.

[6] B. Korte and D. Hausmann, An analysis of the greedy algorithm for independence systems, in Algorithmic Aspects of Combinatorics, B. Alspach, P. Hall and D. J. Miller Eds., *Annals of Discrete Mathematics*, 2 (1978) 65-74.

[7] R. M. Karp and M. Sipser, Maximum matchings in sparse random graphs, *Proceedings of the 22nd Annual IEEE Symposium on Foundations of Computing*, (1981) 364-375.

[8] G. Tinhofer, A probabilistic analysis of some greedy cardinality matching algorithms, *Annals of Operations Research*, 1 (1984) 239-254.