THE AVERAGE PERFORMANCE OF THE GREEDY MATCHING ALGORITHM

by

Martin Dyer

School of Computer Studies
University of Leeds
Leeds, U.K.

and

Alan Frieze

Department of Mathematics
Carnegie Mellon University
Pittsburgh, PA U.S.A.

and

Boris Pittel

Department of Mathematics
Ohio State University
Columbus, OH U.S.A.

Research Report No. 91–138 2

October, 1991

randomised version on a tree and they found that the trees with the worst expected ratio of size of matching found to maximum size are *caterpillars* in which case the ratio is about .769... . The average performance of this algorithm when the input is random has been analysed by Tinhofer [9]. He considered its performance on the random graph $G_{n,p}$ (the p-model) in which each edge of the complete graph is independently included with probability p. He only considered the dense case where p is fixed independent of n. In this case it is fairly easy to show that the algorithm produces a near perfect matching with high probability. The algorithm is deceptively simple but it requires a non-trivial analysis to handle the conditioning introduced at each stage. Unfortunately Tinhofer makes an incorrect assertion and the analysis is consequently flawed. (The statement Prob(M | G) = 1/m! on p244 is incorrect). In this paper we consider sparse random graphs. We deal with the random graph $G_{n,m}$ which has vertex set $[n] = \{1,2,\ldots,n\}$ and $m = \frac{1}{2}cn$ random edges where $c > 0$ is a constant. Closely related is the p-model with $p = c/n$. (In both models the average degree is asymptotic to c.) Let $X = X(n,m)$ ($X(n,p)$ resp.) be the random number of edges in the matching produced by GREEDY applied to $G_{n,m}$ ($G_{n,p}$ resp.) when the edge choice in statement A is uniformly random. We will not only compute an asymptotic formula for the mean, but also for the variance and we will establish the asymptotic distribution. Let

$$\phi(c) = \frac{c}{2(c+1)}$$

$$\psi(c) = \frac{c^2(c+3)}{6(c+1)^4}.$$

Theorem 1 *As $n \to \infty$ $(X(n,m) - n\phi(c))/\sqrt{n\psi(c)}$ converges in distribution and with all its moments to the standard normal variable with mean zero and*

variance one.

□

Roughly X(n, *m)* is asymptotically Gaussian with mean *n<f>(c)* and variance *n\l>(c).* As one should expect, ^(oo) = 1/2 which corresponds to a (near) perfect matching.

Using a general rule, which relates $G_{n,m}$ and $G_{n,p}$, (Pittel [8]), we can assert then that X(n,p) is asymptotically Gaussian as well, with the same mean *n<j>{c),*andvariance

$$n(\psi(c) + 2c(\phi'(c))^{2}) = n\frac{c^3 + 3c^2 + 3c}{(c+1)^4}.$$

We also discuss the performance of GREEDY on a randomly chosen labeled tree. So now let $Y = Y_n$ be the random number of edges in the matching produced by GREEDY on a random labelled tree with n vertices. We prove

Theorem 2

$$\mathbf{E}Y = \frac{3}{8}n - \frac{1}{32} + O\left(\frac{1}{n}\right),$$

$$\mathbf{Var} Y = \frac{1}{96}n + O(1).$$

Furthermore* (F — |n)/^/^n *converges in distribution to the standard normal variable with mean zero and variance one.

□

This should be compared with the result of Meir and Moon [7] who showed that Y_n^* « (1 — p)n, in probability and mean, where Y_n^* is the size of the largest matching in a random labelled tree on n vertices and $p = .5671\cdots$ is the unique solution to $xe^x = 1$. The above results mean that with high probability $Y_n/Y_n^* \approx .87$, i.e. GREEDY falls short of the mark by about 13% most of the time.

It is possible to modify the algorithm without considerable complications, so as to improve its likely performance. Perhaps the simplest modification is to first choose a vertex v at random and then to randomly choose an edge incident with v. We refer to this as MODIFIED GREEDY.

MODIFIED GREEDY

```
begin
    M ← ∅;
    while E(G) ≠ ∅ do
    begin
        B:  Choose v ∈ V
        C:  Choose u ∈ Γ(v) and let e = {u, v}
            G ← G\{u,v};
            M ← M ∪ {e}
    end;
Output M
end
```

We have analysed the performance of MODIFIED GREEDY in the same settings as for GREEDY. First of all, let $\hat{X} = \hat{X}(n,m)$, $(\hat{X}(n,p))$ be the random number of edges in the matching produced by MODIFIED GREEDY on $G_{n,m}$ ($G_{n,p}$). Let

$$\hat{\phi}(c) = \tfrac{1}{2} - \frac{\log(2 - e^{-c})}{2c}.$$

Theorem 3 *As $n \to \infty$ $(\hat{X}(n,m) - n\hat{\phi}(c))/\sqrt{n\hat{\psi}(c)}$ converges in distribution, and with all its moments, to the standard normal variable with mean zero and variance one. (Here $\psi(c)$ is the solution of the differential equation (50) below, whose closed form solution, if any, has eluded us, Maple and Mathematica.) Consequently $\hat{X}(n,p)$ is also asymptotically normal with mean $n\hat{\phi}(c)$ and variance $n(\hat{\psi}(c) + 2c\hat{\phi}'(c)^2)$.*

□

MODIFIED GREEDY was also discussed in Tinhofer [9] as well as Goldschmidt and Hochbaum [3] who proved probabilistic lower bounds on the size of the matching produced in $G_{n,p}$. In particular Goldschmidt and Hochbaum prove a probabilistic lower bound of $n(1 - (1+\epsilon)/c)/2$ for any fixed $\epsilon > 0$.

Since $\phi(c) > (1 - c^{-1})/2$, Theorem 1 already provides a better lower bound for the matching number. Since, for any $c > 0$, $\hat{\phi}(c) > \phi(c)$ Theorem 3 yields a further improvement.

We have also analysed the performance of MODIFIED GREEDY on random labelled trees. Let now $\hat{Y} = \hat{Y}_n$ be the random number of edges in the matching produced by MODIFIED GREEDY on a random labelled tree with n vertices. We prove

Theorem 4

$$\begin{aligned}
\mathbf{E}\hat{Y} &= \frac{e-1}{2e-1}n - \frac{3e-2e^2}{2(2e-1)^3} + O\left(\frac{1}{n}\right), \\
\mathbf{Var}\hat{Y} &= \theta n + O(1).
\end{aligned}$$

where

$$\theta = \frac{18e^4 - 42e^3 + 32e^2 - 9e + 2}{(2e-1)^4}$$

.

Furthermore $(\hat{Y} - \frac{e-1}{2e-1}n)/\sqrt{\theta n}$ *converges in distribution to the standard normal variable with mean zero and variance one.*

□

Here $(e-1)/(2e-1) = .3873\cdots$, i.e. MODIFIED GREEDY performs (with high probability) about 3% better than GREEDY.

We note that Karp and Sipser [5] have considered a similar greedy type of algorithm to ours. Their algorithm chooses an edge incident to a vertex of degree 1 while there is one and otherwise chooses a random edge. They show that this algorithm is asymptotically optimal in the sense that with high probability it finds a matching which is within $o(n)$ of the optimum size. What this optimum size actually is remains a mystery.

The main techniques employed in this paper are diffusion type approximations to moment generating functions for establishing normality and the analysis of regular generating functions in the case of trees.

2 Greedy on $G_{n,m}$

In this section we prove Theorem 1 on the behaviour of randomized greedy applied to the random graph $G_{n,m}$. We assume that $m = \lfloor \frac{1}{2}cn \rfloor$, where c is a constant. Thus c is (approximately) the average degree of the graph.

Generically, let X_y^ be the (random) number of matching edges delivered by the algorithm in $G_{„fM}$, and let $f_{y\%ll}$ be its moment generating function. Thus

$$\begin{aligned} f_{\nu,\mu}(z) &= \mathbf{E}(e^{zX_{\nu,\mu}}) \\ &= \mathbf{E}_R(\mathbf{E}(e^{z(1+X_{\nu-2,\mu-1-R})} \mid R)), \\ &= e^z \sum_r f_{\nu-2,\mu-1-r}(z) \Pr(R = r), \qquad (1) \end{aligned}$$

where R is the random number of edges in $Gj_{,tM}$ adjacent to the chosen edge.

(When a chosen edge e and R adjacent edges are deleted, we are left with a random graph G'. Conditioned on e = (fi,!^) and R = r, G' has the same distribution as C_2_{fM}-i-.$_r$ up to a relabelling of vertices.)

Let us denote the number of edges in the complete graph K_u-i by Ni = ("J^{i}) for i = 0,1,.... We will simply write N for JV_0. Now, since there are $2(v$ - 2) of edges of K_v adjacent to a given edge, iV_2 edges in AV_2> and \JL—1 edges still to be chosen in $G_{y\%ti}$ once any one is fixed, it follows that R has distribution

$$?v(R = r) = \frac{V\ r \quad \dot{\ } \underline{\backslash} i \quad \wedge \underline{r)}}{\binom{}{\mu-1}} = p_r(\nu, \mu),$$

say. If r = O(v^) and ^ = 2^/i/ = 0(1), then

$$\begin{aligned} \frac{\binom{N_2}{\mu-1-r}}{\binom{N-1}{\mu-1}} &= \frac{N_2(N_2-1)\cdots(N_2-\mu+r+2)}{(N-1)(N-2)\cdots(N-\mu+1)}(\mu-1)(\mu-2)\cdots(\mu-r) \\ &= \left(1-\frac{4}{\nu}+O\left(\frac{1}{\nu^2}\right)\right)^{\mu-r-1}\left(\left(1+O\left(\frac{1}{\nu}\right)\right)\frac{1}{N}\right)^r\left(\left(1+O\left(\frac{r}{\nu}\right)\right)\mu\right)^r \\ &= \left(1+O\left(\frac{r^2}{\nu}\right)\right)e^{-2c}N^{-r}\mu^r \end{aligned}$$

Also, we have

$$/2(P\ -\ 2)\backslash \qquad \frac{(2\nu-4)(2\nu-5)\cdots(2i/-3-r)}{r!}$$

$$= \left(1+O\left(\frac{r}{\nu}\right)\right)^r \frac{(2\nu)^r}{r!}$$
$$= \left(1+O\left(\frac{r^2}{\nu}\right)\right) \frac{(2\nu)^r}{r!}$$

Thus

$$p_r(\nu,\mu) = \left(1+O\left(\frac{r^2}{\nu}\right)\right) \frac{e^{-2\xi}}{r!}\left(\frac{2\mu\nu}{N}\right)^r = \left(1+O\left(\frac{r^2}{\nu}\right)\right) \frac{e^{-2\xi}(2\xi)^r}{r!} \tag{2}$$

We will only need an upper bound on $\Pr(R \geq r)$ for large r. Let Δ be the maximum vertex degree of $G_{\nu,\mu}$. A straightforward application of the "first-moment method" now gives

$$\begin{aligned}
\Pr(R \geq 2r) &\leq \Pr(\Delta \geq r) && (3)\\
&\leq \nu\binom{\nu-1}{r}\frac{\mu(\mu-1)\cdots(\mu-r+1)}{N(N-1)\cdots(N-r+1)}\\
&\leq \nu\frac{(\nu-1)^r}{r!}\left(\frac{\mu}{N}\right)^r\\
&\leq \nu\frac{(\nu-1)^r}{r!}\left(\frac{\xi}{\nu-1}\right)^r\\
&= \nu\frac{\xi^r}{r!} && (4)
\end{aligned}$$

It now follows easily from (2) and (4) that

$$\mathbf{E}(R) = 2\xi + O(1/\nu), \quad \mathbf{E}(R^2) = 2\xi + 4\xi^2 + O(1/\nu), \tag{5}$$

and more generally, for any fixed k, $\mathbf{E}(R^k) = O(1)$. We will denote $\mathbf{E}(R^k)$ by $\bar{r^k}$.

In our analysis of the randomized greedy algorithm, we need to assume that the average degree of the graph "remains bounded" as we select sufficiently many edges. Suppose that at a general stage of the algorithm we have ν

vertices and μ edges remaining, so our graph is $G_{\nu,\mu}$. We wish to have a constant c_0 such that $2\mu/\nu \leq c_0$ if ν is "large". To make this precise,

Lemma 1 *Let* $\nu_0 = \lceil\sqrt{n/\ln n}\rceil$ *and* $c_0 = 6\max\{2c, 1\}$. *Then*

$$\Pr(\exists \nu \geq \nu_0 \textit{ such that } 2\mu/\nu > c_0) < \exp(-\sqrt{n \ln n})$$

for all large enough n.

Proof Let $\mathcal{E}$ be the event that there exists *any* induced subgraph $G_{\nu,\mu}$ of $G_{n,m}$ such that $2\mu/\nu > c_0$ and $\nu > \nu_0$. Then clearly

$$\begin{aligned}
\Pr(\mathcal{E}) &\leq \sum_{\nu=\nu_0}^{n} \binom{n}{\nu} \binom{\binom{\nu}{2}}{c_0\nu/2} \left(\frac{cn/2}{N}\right)^{c_0\nu/2} \\
&\leq \sum_{\nu=\nu_0}^{n} \left(\frac{en}{\nu}\left(\frac{e(\nu-1)}{c_0}\right)^{c_0/2}\left(\frac{c}{n-1}\right)^{c_0/2}\right)^{\nu} \\
&\leq \sum_{\nu=\nu_0}^{n} \left((1+o(1))e\left(\frac{ec}{c_0}\right)^{c_0/2}\left(\frac{\nu}{n}\right)^{c_0/2-1}\right)^{\nu} \\
&\leq \sum_{\nu=\nu_0}^{n} \left((1+o(1))e\left(\frac{e}{12}\right)^{3}\left(\frac{\nu}{n}\right)^{2}\right)^{\nu} \\
&\leq \sum_{\nu=\nu_0}^{n} \left(\frac{\nu}{5n}\right)^{2\nu} \\
&\leq n\left(\frac{\nu_0}{5n}\right)^{2\nu_0} \\
&\leq n\left(\frac{1}{25n\ln n}\right)^{\sqrt{n/\ln n}} \\
&\leq e^{-\sqrt{n\ln n}},
\end{aligned}$$

for large enough n. This is clearly stronger than the lemma. □

Remark 1 In fact, we can prove the following stronger result here, using an inductive method similar to the proof of Lemma 2 below. Let *CQ* > *c* and 0 < 7 < 1 be arbitrary constants. Then there exist positive constants A(co, 7) and /?(c, Co, 7) such that the probability of having *2fi/v* > CQ at any stage in the algorithm is at most j4exp(—/?n⁷). We omit the proof, since this stronger result is not true for all choice rules A in GREEDY. It fails, for example, for that used in MODIFIED GREEDY.

We turn now to the functions *<f>(£)* = $\frac{£}{2(1+f)}$ rr?^(O == $\frac{£^2(3+£)}{6(1+f)^4}$ which appear in the mean and variance of $X_{n,m}$ in the statement of Theorem 1. We observe that these satisfy the differential equations

$$1 - 2 \wedge - 2 \wedge (1 + 0 = 0, \quad <£(0) = 0 \qquad (6)$$

$$-tf - tf'(i + 0 + W)^2 = 0, \quad v(o) = o \qquad (7)$$

Note that <^,xj) are "well behaved", i.e. they, and all their derivatives, are uniformly bounded on [0,oo). In particular <£(oo) = |, ^(00) = 0.

We now proceed to analyse the approximation which will enable us to prove Theorem 1. Let $g_{n},m(z)$ = exp(n(z<£(2n/m) + ±z²tj>(2n/m))). We will show that

$$fn.m(z) = {}_{fc|W}W(l + O\wedge-\wedge logn)), \qquad (8)$$

provided $z = O(n''^1\wedge^2)$. This will imply Theorem 1, that X_{ntTn} is asymptotically Gaussian with mean *n<f>(c),* variance *n%l>(c),* by a limit theorem on moment generating functions due to Curtis [1]. As we will see, the functions *<t>,%l>* are chosen to ensure that g_v^ satisfies (1) up to terms quadratic in *z*. Thus, let us choose CQ as in Lemma 1, and consider a stage of the algorithm at which we have $v \geq v_0 = yjn/\overline{\ln n}$ vertices, *fi* edges remaining, with *2fi/v* < Co. Then

Lemma 2 *For z* = **un"$^{1/2}$ (u ^ 0),**

$$e^z \sum_r p_r(\nu, \mu) g_{\nu-2,\mu-1-r}(z) = (1 + O(z/\nu)) g_{\nu,\mu}(z).$$

Proof We will estimate the sum

$$\sum_r p_r(\nu, \mu)(e^z g_{\nu-2,\mu-1-r}/g_{\nu,\mu}).$$

Using the quantities

$$\xi = 2\mu/\nu \qquad\qquad _{\nu-2}(\xi - 1 - r),$$

we may write

$$g_{\nu-2,\mu-1-r}(z) = \exp((\nu - 2)(z\phi(\xi_r) + z^2\psi(\xi_r))).$$

Using Taylor's theorem, we may easily establish

$$\phi(\xi_r) = \phi(\xi) + \phi'(\xi)\frac{2}{\nu}(\xi - 1 - r) + O\left(\frac{r^2}{\nu^2}\right), \tag{9}$$

$$\psi(\xi_r) = \psi(\xi) + \psi'(\xi)\frac{2}{\nu}(\xi - 1 - r) + O\left(\frac{r^2}{\nu^2}\right), \tag{10}$$

uniformly for £ € (0,co). Thus, using (9) and (10),

$$\begin{aligned} \ln(e^z g_{\nu-2,\mu-1-r}/g_{\nu,\mu}) &= z + (\nu - 2)(z\phi(\xi_r) + \tfrac{1}{2}z^2\psi(\xi_r)) \\ &\quad - \nu(z\phi(\xi) + \tfrac{1}{2}z^2\psi(\xi)) \\ &= z(1 - 2\phi(\xi) + 2\phi'(\xi)(\xi - 1 - r)) + \tfrac{1}{2}z^2(-2\psi(\xi) \\ &\quad + 2\psi'(\xi)(\xi - 1 - r)) + O(r^2\wedge)- \end{aligned} \tag{11}$$

Denote the right side of (11) by *r{r)*. Thus we wish to show that

$$\sum_r p_r(\nu, \mu) e^{\tau(r)} = l + O(z/u). \tag{12}$$

We deal first with large r in the sum of (12). Specifically this will mean $r > n^1/^3$, say. Note, since r < 2i/, and z = o(l),

$$r^2*/*/ < 2rz = o(r),$$

and hence r(r) = o(r). Then, using (4),

$$\mathbf{\sum_{r>n^{1/3}} JVeM \lesssim » \sum_{r>\frac{1}{2}n^{1/3}} (ee^{o(1)})7r! = n\text{-}^{fi}<«^{1/3}>,} \qquad \textbf{(13)}$$

say, since the first term dominates the sum and its denominator dominates its numerator. Thus large r do not effectively contribute to the sum in (12). But now, if $r \leq n^{1/3}$,

$$r(r) = O(«) = \wedge(n"^1/^6),$$

since ^ = (^(n""$^1/^2$). Thus we can expand $e^T\wedge$ uniformly over r in this region. Therefore, using (13),

$$\begin{aligned}
\sum_r p_r e^{\tau(r)} &= \sum_{r \leq n»/^3} p_r e^{\tau(r)} + n^{-\Omega(n^{1/3})} \\
&= \sum_{r \leq n^{1/3}} p_r[1 + z(1 - 2\phi + 2\phi'(\xi - 1 - r)) \\
&\quad + \backslash z^2\{-2tf> + 20'U - 1 - r) + (1 - 2\wedge + 2(t>\backslash i - 1 - r))^2) \\
&\quad + 0(rV) + O(r^2 5r/i/)] + n\text{-}"(^{nl/3}) \\
&= 1 + z(l - 2<f> + 2\wedge'U - 1 - f)) + \wedge^2(-2\wedge + 2V\text{-}'(e - 1 - f) \\
&\quad + \sum_r Pr(l - 2\wedge + 2/(f - 1 - r))^2) + O(r"3z^3) + O(\bar{H}*/i/),
\end{aligned} \qquad (14)$$

We will now examine in turn the coefficient of z, of $\backslash z^2$ and the remainder terms in (14). By comparison with (5), we have f = 2£ + $0\{\backslash jv)$. Thus the coefficient of z is

$$1 - 2\phi - 2\phi'(1+\xi) + O(1/\nu) = O(1/\nu),$$

on using (6). Now, the summation in the coefficient of $\frac{1}{2}z^2$ involves

$$(1 - 2\phi + 2\phi'(\xi - 1 - r))^2 = 4(\phi')^2(r - 2\xi)^2,$$

on using (6) again. Thus the summation is

$$4(\phi')^2 \sum_r (r - 2\xi)^2 p_r = 8\xi(\phi')^2 + O(1/\nu),$$

on expanding the bracket and using (5). Thus the coefficient of $\frac{1}{2}z^2$ is, on substituting for $\bar{r}$,

$$-2\psi + 2\psi'(\xi - 1 - 2\xi + O(1/\nu)) + 8\xi(\phi')^2 + O(1/\nu) = O(1/\nu),$$

in view of (7). Finally, for the remainder terms,

$$O(\bar{r^3} z^3) = O(z^3) = O(z/n) = O(z/\nu), \quad O(\bar{r^2} z/\nu) = O(z/\nu),$$

using $z = O(n^{-1/2})$, $\nu \leq n$ and the boundedness of $\bar{r^k}$ for fixed k. Combining these results we see that the right side of (14) is $(1 + O(z/\nu))$, completing the proof of the lemma. □

We are now in a position to prove Theorem 1. Applying Lemma 1 we can assert that the event $2\mu/\nu > c_0$ will rarely be encountered during the algorithm until $\nu \leq \nu_0 = \sqrt{n/\ln n}$. We may therefore use an "approximating" stochastic process $\tilde{X}_{\nu,\mu}$, defined for $0 \leq \nu \leq n, 0 \leq \mu \leq \binom{\mu}{2}$ as follows,

$$\begin{aligned}\tilde{X}_{\nu,\mu} &= 1 + \tilde{X}_{\nu-2,\mu-1-R} \text{ if } 2\mu/\nu \leq c_0, \text{ and } \nu > \nu_0, \\ &= Y_{\nu,\mu} \text{ otherwise,}\end{aligned}$$

where *R* is an independent random variable with distibution p_r(i/, /J), and Y_v^ is a Gaussian random variable with mean *u<f>(2fi/u)^* variance *i/tp(2fi/u).* We can think of this as running GREEDY until $v \leq u_0$ or possibly $2fi/v$ > Co and then adding Y_v^ to the edge count in place of the number of edges found by GREEDY on the remaining G_{Uyil}. Note that Y_v^ is the constant zero if *fi* = 0 in view of the initial conditions in (6) and (7). Clearly, $\tilde{f}_{Vit}i\{z)$ = E($e^{zjfl/\prime\prime i}$) satisfies, for *2\ijv <_CQ* and *v* > z/0>

$$\tilde{f}_{\nu,\mu}(z) = e^z \sum_r p_r(\nu,\mu)\tilde{f}_{\nu-2,\mu-1-r}(z) \qquad (15)$$

Let now *S* refer to the event that in applying GREEDY we reach a point where *v* > *v$* and *2fi/v* > *CQ.* Lemma 1 tells us that

$$\Pr(£) \leq \exp\{-\sqrt{nlnn}\}.$$

Now let

$$x \;—x(^a)\; 4\text{-}\, x\wedge$$

where *X^* counts the edges added before $v \leq v_0$. Define a similar decomposition for $\tilde{X}_{nim}$ except that if *S* occurs then $\tilde{X}_{v>li}$ = $\tilde{X}fy$ and otherwise $\tilde{X}ffy$ = Y_v*,n* where $v \leq v_0$. Then dropping the subscripts *n,m,*

$$\begin{aligned} E(e^{*x}) &= E(e^{zX^{(a)}})\mathbf{E}(e^{zX^{(b)}}) \\ &= (1+o(1))\mathbf{E}(e^{zX^{(a)}}) \end{aligned}$$

since $X^{\wedge(b)} \leq v_0$.

Similarly

$$\mathbf{E}(e^{z\hat{X}}) = (1+o(1))\mathbf{E}(e^{z\hat{X}^{(a)}}).$$

But

$$\mathbf{E}(e^{z\tilde{X}^{(a)}}) = \mathbf{E}(e^{z\tilde{X}^{(a)}} \mid \mathcal{E})\Pr(\mathcal{E}) + \mathbf{E}(e^{z\tilde{X}^{(a)}} \mid \bar{\mathcal{E}})\Pr(\bar{\mathcal{E}})$$

$$= o(e^{-\sqrt{n\ln n}/2}) + \mathbf{E}(e^{z\tilde{X}^{(a)}} \mid \bar{\mathcal{E}})\Pr(\bar{\mathcal{E}}),$$

since $X^{(a)} = \tilde{X}^{(a)}$ if $\mathcal{E}$ does not occur,

$$= (1+o(1))\mathbf{E}(e^{zX^{(a)}}).$$

Therefore $\mathbf{E}(e^{z\tilde{X}}) \sim \mathbf{E}(e^{zX})$. Hence it is sufficient to show that the asymptotic distribution of $\tilde{X}_{\nu,\mu}$ is as described in Theorem 1. But this is now easy. Let us prove by induction that $\tilde{f}_{\nu,\mu}(z) = E(e^{z\tilde{X}_{\nu,\mu}})$ satisfies

$$e^{-C_1 z\ln\nu} g_{\nu,\mu}(z) \le \tilde{f}_{\nu,\mu}(z) \le e^{+C_1 z\ln\nu} g_{\nu,\mu}(z), \tag{16}$$

for some constant $C_1 > 0$. If $2\mu/\nu > c_0$, this is true by definition. Otherwise if $\nu \le 3$, say, then $\mu \le 3$ and hence $g_{\nu,\mu}(z)$, $E(e^{z\tilde{X}_{\nu,\mu}})$are both $e^{O(z)}$. Hence, by suitable choice of C_1, (16) will be true for $\nu \le 3$. Now, for $\nu > 3$, from Lemma 2 there is a constant C_2 such that

$$e^{-C_2 z/\nu} g_{\nu,\mu}(z) \le e^z \sum_r p_r(\nu,\mu) g_{\nu-2,\mu-1-r}(z) \le e^{+C_2 z/\nu} g_{\nu,\mu}(z).$$

We will assume without loss that $C_1 \ge \frac{1}{2}C_2$. Hence, by induction, for $\nu > \nu_0$ and $2\mu/\nu \le c_0$ we have (see (15)):

$$\begin{aligned}
\tilde{f}_{\nu,\mu}(z) &= e^z \sum_r p_r \tilde{f}_{\nu-2,\mu-r-1} \\
&\le e^{+C_1 z\ln(\nu-2)} e^z \sum_r p_r g_{\nu-2,\mu-1-r}(z) \\
&\le e^{+C_1 z\ln(\nu-2)} e^{+C_2 z/\nu} g_{\nu,\mu}(z) \\
&\le \exp(+C_1 z(\ln(\nu-2) + 1/2\nu)) g_{\nu,\mu}(z) \\
&\le e^{+C_1 z\ln\nu} g_{\nu,\mu}(z).
\end{aligned}$$

Similarly $\tilde{f}_{\nu,\mu} \ge e^{-C_1 z\ln\nu} g_{\nu,\mu}(z)$, proving (16).

But, putting $\nu = n$, $\mu = m$ in (16) gives (8) and completes the proof of Theorem 1.

3 Greedy on random trees

We will devote this section to the proof of Theorem 2. Our first task is to prove the claims about the mean and variance of Y_n, the number of edges in the matching produced by GREEDY on a random tree. So for $n \geq 1$ let

$$f_n(z) = \mathbf{E}(z^{Y_n})$$

denote the probability generating function for this random variable and let

$$\Omega(k,d) = \{a \in [k-1]^d : a_1 + a_2 + \dots a_d = k-1\}.$$

The deletion of the edge $e = \{u, v\}$ chosen in statement **A** produces two trees, one containing u and the other v. If the degrees of u and v in those trees are d and δ, respectively, after deleting the edges incident to e in the original tree, two forests of trees will be left, of sizes $a \in \Omega(k,d)$, $b \in \Omega(n-k,\delta)$ for some k, d, δ, where the sub-trees with sizes in a, b are associated with different endpoints of e. Let $\pi(a,b)$ denote the probability of the occurrence of a particular *unordered* pair a, b.

Now let

$$\begin{aligned}
\rho(a,b) &= \frac{1}{(n-1)n^{n-2}}\binom{n}{k}\left(\frac{k!}{d!}\prod_{r=1}^{d}\frac{a_r^{a_r-1}}{a_r!}\right)\left(\frac{(n-k)!}{\delta!}\prod_{s=1}^{\delta}\frac{b_s^{b_s-1}}{b_s!}\right) \\
&= \rho(b,a).
\end{aligned}$$

Then for $n \geq 2$

$$\pi(a,a) = \frac{\rho(a,a)}{2}$$

and

$$\pi(a,b) = \frac{\rho(a,b)+\rho(b,a)}{2}$$

for $a \neq b$.

Explanation: there are $(n-1)n^{n-2}$ (tree,chosen edge) pairs in total; $\binom{n}{k}$ counts the choices for a set K of k vertices; k counts the choices for a vertex $u \in K$; $\frac{(k-1)!}{d!}\prod_{r=1}^{d}\frac{1}{a_r!}$ counts the ways of partitioning $K\setminus\{u\}$ into the subsets of cardinalities $a \in \Omega(k,d)$; $\prod_{r=1}^{d} a_r^{a_r-1}$ counts the number of ways of forming trees of sizes $a_1, a_2, \ldots, a_d$ and designating the vertices (roots) in the trees which will be attached to u; the final factor in $\rho(a,b)$ is constructed similarly. When $a = b$ there is a double counting but there is none when $a \neq b$.

Now, the deletion of the edge e, chosen uniformly at random, produces a forestof random trees, which–conditioned on their vertex sets–are independently uniform. So, in distribution,

$$Y_n = 1 + \sum_{i=1}^{d} Y_{a_i} + \sum_{j=1}^{\delta} Y_{b_j}$$

where the variables on the right side are independent.

It follows then that we have for $n \geq 2$

$$f_n(z) = \frac{z}{2}\sum_{k=1}^{n-1}\sum_{d=0}^{\infty}\sum_{\delta=0}^{\infty}\sum_{a\in\Omega(d,k)}\sum_{b\in\Omega(\delta,n-k)} \rho(a,b)\prod_{r=1}^{d} f_{a_r}(z)\prod_{s=1}^{\delta} f_{b_s}(z). \qquad (17)$$

We can simplify the above expression by introducing two bivariate generating functions

$$F(x,z) = \sum_{n=1}^{\infty}\frac{n^{n-1}f_n(z)}{n!}x^n$$

and

$$G(x,z) = \sum_{n=2}^{\infty}\frac{n^{n-2}f_n(z)}{n!}x^n.$$

Observe that

$$F(x,z) = \frac{\partial}{\partial x}G(x,z).$$

We multiply (17) by n''''' j''''1 >i^B and sum from n = 2 to oo. The left side becomes $F(x, z) - G(x, z)$. The right side becomes

$$\frac{x^2 z}{2}\left(\sum_{d=0}^{\infty}\frac{1}{d!}\sum_{k=1}^{\infty}\sum_{a\in\Omega(d,k)}\prod_{r=1}^{d}\frac{a_r^{a_r-1}f_{a_r}(z)}{a_r!}x^{a_r}\right)\left(\sum_{\delta=0}^{\infty} ht \; \mathrm{E}_{,k)} \prod_{s=1}^{\delta}\frac{b_s^{b_s-1}f_{b_s}(z)}{b_s!}x^{b_s}\right)$$

$$= \frac{x^{*}\pm}{2}\left(\sum_{d=0}^{\infty}\frac{1}{d!}F(x,z)^d\right)\left(\sum_{\delta=0}^{\infty}\frac{1}{\delta!}F(x,z)^\delta\right).$$

So we have the equation

$$F(x,z)-G(x,z) = \wedge \exp\{2F(x,z)\}. \tag{18}$$

Even though the equation appears to be extremely hard to solve in a closed form, it is ideally suited to determine (asymptotically) the mean and variance of y_n.

Let $E_n = E(Y_n)$ and

$$E(x) = \sum^{\infty} E_n \frac{n^{n-1}}{!} x^n.$$

Since $E_n = /\!\ll(!)$ we have

$$E(x) = \frac{\partial F}{\partial z}(x, 1), \tag{19}$$

and if $\tilde{E}(x) = \sum\sum_{2} E_n 2 \wedge x^n = |f(x, 1)$ then

$$E = x\frac{d\tilde{E}}{dx}. \tag{20}$$

Also

$$F(x,1) = \sum^{\infty} \frac{n^{n-1}}{n!}x^n = T(x),$$

the *tree function* ([10]) defined for $|x| \le e^{-1}$. It is well known that

$$T(x) = xe^{T(x)}, \qquad |x| \le e^{-1} \tag{21}$$

and so

$$\frac{dT}{dx} = \frac{T(x)}{x(1-T(x))}, \qquad |x| < e^{-1}. \tag{22}$$

Differentiating (18) with respect to z gives

$$\frac{\partial F}{\partial z} - \frac{\partial G}{\partial z} = e^{2F}\left(\frac{x^2}{2} + x^2 z \frac{\partial F}{\partial z}\right).$$

Putting $z = 1$ and using (19)-(21) we obtain

$$\frac{d\tilde{E}}{dx} - \tilde{E} = \frac{T^2}{x^2}\left(\frac{x^2}{2} + x^3\frac{d\tilde{E}}{dx}\right)$$

or

$$(1 - T^2)\frac{d\tilde{E}}{dx} - \tilde{E} = \frac{T^2}{2},$$

and on using (22)

$$T(1+T)\frac{d\tilde{E}}{dx} - \tilde{E} = \frac{T^2}{2}. \tag{23}$$

Now $\tilde{E}(0) = T(0) = 0$. Solving (23) with this initial condition gives

$$\tilde{E} = \frac{T^2}{2(1+T)}.$$

We then apply (20) and (22) to obtain

$$E = \frac{T^2(T+2)}{2(1+T)^2(1-T)}. \tag{24}$$

Thus

$$E_n = \frac{n!}{2\pi i}\int_{C_0} \frac{T^2(T+2)}{2(1+T)^2(1-T)}\frac{dx}{x^{n+1}}$$

where C_0 is a circle of small radius (less than e^{-1} around the origin in the complex x plane). We now make the substitution $x = te^{-t}$ and work in the complex t plane. (Since C_0 has a small radius the transformation is well behaved.) Cancelling $(1-t)$ in the denominator and in $dx = (1-t)e^{-t}dt$, we arrive at

$$E_n = \frac{n!}{n^{n-1}}\frac{1}{2\pi i}\int_{C_1}\frac{t^2+2t}{2(1+t)}\frac{e^{nt}}{t^n}dt \qquad (25)$$

where now C_1 is a simple contour around the origin in the complex t plane which does not enclose the point $t = -1$. To estimate the integral, we use an identity

$$I_k = \frac{1}{2\pi i}\int_{C_1}\frac{(t-1)^k}{t^n}e^{nt}dt = \sum_{i=0}^{k}\binom{k}{i}(-1)^{k-i}\frac{n^{n-i-1}}{(n-i-1)!}$$

which follows from

$$\frac{1}{2\pi i}\int_{C_1}\frac{e^{nt}}{t^m}dt = \frac{n^{m-1}}{(m-1)!} \text{ for } m \geq 1.$$

Putting $u_k = \frac{n!}{n^{n-1}}I_k$ we find in particular that

$$u_0 = n;\ u_1 = -1;\ u_2 = -1 + \frac{2}{n};\ u_3 = \frac{5}{n} - \frac{6}{n^2}. \qquad (26)$$

Suppose now that C is the circle of radius 1 around the origin and k is fixed. On putting $t = e^{i\theta}$

$$\begin{aligned}\left|\frac{1}{2\pi i}\int_C\frac{(t-1)^k}{t^n}e^{nt}dt\right| &= O\left(\int_{-\pi}^{\pi}|t^{-1}e^t|^n|1-t|^k d\theta\right)\\ &= O\left(e^n\int_0^{\pi}e^{-n(1-\cos\theta)}\theta^k d\theta\right)\end{aligned}$$

since $|1-e^{i\theta}| \leq |\theta|$. Now $1-\cos\theta \geq \theta^2/4$ for $\theta \leq 3/2$ and $1-\cos\theta \geq 1-\cos 3/4$ for $3/4 \leq \theta \leq \pi$ and so substituting $u = \theta\sqrt{n}$ we obtain

$$\begin{aligned}\left|\frac{1}{2\pi i}\int_C\frac{(t-1)^k}{t^n}e^{nt}dt\right| &= O\left(e^n n^{-(k+1)/2}\int_0^{\infty}e^{-u^2/4}u^k du\right)\\ &= O(e^n n^{-(k+1)/2}). \qquad (27)\end{aligned}$$

It is easy to see that (27) remains valid if C has a circular dent of a sufficiently small, but fixed, radius at $t = -1$. We return to (25) and expand the function $\frac{t^2+2t}{2(1+t)^2}$ around $t = 1$ to obtain

$$\frac{t^2+2t}{2(1+t)^2} = \frac{3}{8} + \frac{t-1}{8} - \frac{3(t-1)^2}{32} + \frac{(t-1)^3}{16} - \frac{(t-1)^4(2t+3)}{32(t+1)^2}.$$

Integrating over the dented C and using (26), (27) we obtain

$$\begin{aligned} E_n &= \frac{3}{8}\alpha_0 + \frac{1}{8}\alpha_1 - \frac{3}{32}\alpha_2 + \frac{1}{16}\alpha_3 + O(n^{-1}) \\ &= \frac{3}{8}n - \frac{1}{32} + O(n^{-1}). \end{aligned}$$

(Note that $\frac{2t+3}{(t+1)^2}$ is bounded on the integration contour.)

We now estimate the variance of Y_n. So let

$$v_n = E(Y_n(Y_n-1)) = f_n''(1),$$

$$V(x) = \sum_{n=1}^{\infty} V_n \frac{n^{n-1}}{n!} x^n = \frac{\partial^2 F}{\partial z^2}(x,1),$$

and

$$\tilde{V}(x) = \sum_{n=2}^{\infty} V_n \frac{n^{n-2}}{n!} x^n$$

so that $V = x\frac{d\tilde V}{dx}$.

Differentiating (18) twice with respect to z gives

$$\frac{\partial^2 F}{\partial z^2} - \frac{\partial^2 G}{\partial z^2} = e^{2F}\left(2x^2\frac{\partial F}{\partial z} + 2x^2 z\left(\frac{\partial F}{\partial z}\right)^2 + x^2 z^2\frac{\partial^2 F}{\partial z^2}\right).$$

Putting $z = 1$ we obtain

$$x\frac{d\tilde V}{dx} - \tilde V = T^2\left(2E(E+1) + x\frac{d\tilde V}{dx}\right)$$

which becomes on using (22)

$$T(1+T)\frac{d\tilde{V}}{dT} - \tilde{V} = 2T^2E(E+1). \tag{28}$$

The boundary condition for $\tilde{V}$ comes from $V_2 = 0$ ($Y_2 = 0$ always.) Integrating (28) gives

$$\begin{aligned} \tilde{V} &= \frac{2T}{1+T}\int_0^T E(\tau)(E(\tau)+1)d\tau \\ &= \frac{T^4(6T^2+13T+8)}{12(1-T)(T+1)^4}. \end{aligned} \tag{29}$$

Now

$$\begin{aligned} V_n &= \frac{n!}{n^{n-1}}\frac{1}{2\pi i}\int_{C_0}\frac{V}{x^{n+1}}dx \\ &= \frac{n!}{n^{n-1}}\frac{1}{2\pi i}\int_{C_1}\int\frac{1-t}{t}V\frac{e^{nt}}{t^n}dt \\ &= \frac{n!}{n^{n-1}}\frac{1}{2\pi i}\int_{C_1}\frac{d\tilde{V}}{dt}\frac{e^{nt}}{t^n}dt. \end{aligned} \tag{30}$$

It follows from (29) that

$$\frac{d\tilde{V}}{dt} = \frac{9}{64(1-t)^2} - \frac{139}{384} + \frac{11(1-t)}{96} + \frac{(t-1)^2(44t^2+211t^3+333t^2+223t+41)}{384(t+1)^5}.$$

(The absence of a term $(1-t)^{-1}$ is crucially important.) Now

$$\begin{aligned} \frac{1}{2\pi i}\int_{C_1}\frac{1}{(1-t)^2}\frac{e^{nt}}{t^n}dt &= \frac{n}{2\pi i}\int_{C_1}\frac{e^{nt}}{t^{n+1}}dt \\ &= \frac{n^{n+1}}{n!} \end{aligned} \tag{31}$$

where the first equation comes from integrating by parts. Thus using (26), (27) and (30) we obtain

$$V_n = \frac{9}{64}n^2 + \frac{139}{384}n + O(1).$$

Now

$$\mathbf{Var}\ Y_n = V_n - E_n^2 + E_n$$

and the expression for **Var** Y_n in the statement of Theorem 2 follows.

The equation (18) can be used in principle to find the higher moments of Y_n. Fortunately, there is no need for these increasingly arduous computations. Once we have established asymptotic linearity of the mean and variance, the asymptotic normality of Y_n follows from the recurence (17) for the moment generating functions (cf. Mahmoud and Pittel [6].) For $u \in \boldsymbol{R}$ let

$$g_n(u) = \mathbf{E}(e^{uY_n})$$

be the moment generating function (m.g.f.) of Y_n and

$$h_n(u) = \exp\{E_n u + \tfrac{1}{2} W_n u^2\}$$

where

$$W_n = \mathbf{Var}\ \ Y_n.$$

$h_n(u)$ is the m.g.f. of a normal variable with the asymptotic mean and variance of Y_n. In particular for n *fixed* and $u \to \infty$

$$g_n(u) = h_n(u)(1 + O(u^3)). \tag{32}$$

We must show that if $u = O(1/\sqrt{n})$ then $g_n(u) \approx h_n(u)$ as $n \to \infty$.

For $2 \leq \nu \leq n$ let

$$\Omega_\nu = \bigcup_{k+\ell=\nu-2} \bigcup_{d,\delta} [\Omega(k,d) \times \Omega(\ell,\delta)].$$

Then, see (17),

$$g_\nu(u) = e^u \sum_{a,b\in\Omega_\nu} \pi(a,b) \prod_{r=1}^{d} g_{a_r}(u) \prod_{s=1}^{\delta} g_{b_s}(u). \tag{33}$$

Substituting (32) in (33) we obtain

$$g_\nu(u) = \sum_{a,b\in\Omega_\nu} \pi(a,b)\tau(\nu,a,b)(1+O(u^3)) \tag{34}$$

where

$$\begin{aligned}
\tau(\nu,a,b) &= e^u \prod_{r=1}^{d} h_{a_r}(u) \prod_{s=1}^{\delta} h_{b_s}(u) \\
&= h_\nu(u)\exp\{u\Delta_1(\nu,a,b) + \tfrac{1}{2}u^2\Delta_2(\nu,a,b)\}, \qquad (35)\\
|\Delta_1(\nu,a,b)| &= |1 + \sum_{r=1}^{d} E_{a_r} + \sum_{s=1}^{\delta} E_{b_s} - E_\nu| \\
&\leq A_1(d+\delta)
\end{aligned}$$

for some absolute constant A_1,

and similarly,

$$\begin{aligned}
|\Delta_2(\nu,a,b)| &= |\sum_{r=1}^{d} W_{a_r} + \sum_{s=1}^{\delta} W_{b_s} - W_\nu| \\
&\leq A_1(d+\delta)
\end{aligned}$$

Now (34) and (35) imply that as $u \to 0$

$$g_\nu(u) = h_\nu(u) \sum_{a,b\in\Omega_\nu} \pi(a,b)(1 + u\Delta_1 + \tfrac{1}{2}u^2(\Delta_1^2 + \Delta_2))(1+O(u^3)).$$

Comparing with (32) we obtain two identities for $E_m, W_m (m \geq 2)$:

$$\sum_{a,b\in\Omega_\nu} \pi(a,b)\Delta_1(\nu,a,b) = 0, \tag{36}$$

$$\sum_{a,b\in\Omega_\nu} \pi(a,b)(\Delta_1^2(\nu,a,b) + \Delta_2(\nu,a,b)) = 0. \tag{37}$$

From now on let $u = v/\sqrt{n}$ where v is a constant. Then consider

$$\begin{aligned}\sigma_\nu(u) &= \sum_{a,b\in\Omega_\nu} \pi(a,b)\tau(\nu,a,b) \\ &= \tau_1 + \tau_2,\end{aligned}$$

where

$$\tau_1 = \sum_{d,\delta\le\sqrt{n/\log n}} \pi(a,b)\tau(\nu,a,b).$$

The degree sequence of a random tree $d_1, d_2, \ldots, d_\nu$ is such that $d_1 - 1, d_2 - 1, \ldots, d_\nu - 1$ have the joint distribution of the occupancy numbers in the uniform allocation of $\nu - 2$ distinct balls in ν distinct boxes (see e.g. Moon [10]). So in particular d_1 has distribution 1+BINOMIAL$(\nu - 2, 1/\nu)$. Thus,

$$\begin{aligned}\mathbf{Pr}(d_1 \ge \bar{d} = \lceil\sqrt{n/\log n}\rceil) &\le \binom{\nu-2}{d-1}\left(\frac{1}{\nu}\right)^{d-1} \\ &= \exp\{-\Omega(\sqrt{n\log n})\}.\end{aligned}$$

But for $u = O(1/\sqrt{n})$, $\tau(\nu,a,b) = e^{O(\sqrt{n})}$ and so

$$\begin{aligned}\tau_2 &= e^{O(\sqrt{n})}\mathbf{Pr}(d \text{ or } \delta \ge \bar{d}) \\ &= \exp\{-\Omega(\sqrt{n\log n})\}.\end{aligned}$$

Now if $d, \delta \le \sqrt{n/\log n}$ then $u\Delta_i = O(1/\sqrt{\log n}), i = 1, 2$, and so

$$\tau(\nu,a,b) = h_\nu(u)[1 + u\Delta_1(\nu,a,b) + \tfrac{1}{2}u^2(\Delta_1^2(\nu,a,b) + \Delta_2(\nu,a,b)) + R(u,\nu,a,b)]$$

where

$$|R(u,\nu,a,b)| \le A_2|u|^3(d+\delta)^3$$

for some absolute constant A_2. Applying (36) and (37) we obtain

$$\begin{aligned}
\sigma_\nu(u) &= K(u) \sum_{d,\delta \le \sqrt{n/\log n}} \pi(a,b)(1 + u\Delta_1 + \tfrac{1}{2}u^2(\Delta_1^2 + \Delta_2) + R) \\
&\quad + K(u) \sum_{REST} \pi(a,b) \exp\{u\Delta_1 + \tfrac{1}{2}u^2\Delta_2\} \\
&= h_\nu(u) + h_\nu(u) \sum \pi(a,b) R \\
&\quad + K(u) \sum_{REST} \pi(a,b)(\exp\{u\Delta_1 + \tfrac{1}{2}u^2\Delta_2\} - (1 + u\Delta_1 + \tfrac{1}{2}u^2(\Delta_1^2 + \Delta_2)).
\end{aligned}$$

Now the second sum is $h_\nu(u)\exp\{-\Omega(\sqrt{n\log n})\}$.

Also, where d_{max} is the maximum degree of a random tree,

$$\begin{aligned}
\left| \sum_{d,\delta \le \sqrt{n/\log n}} \pi(a,b) R \right| &\le \sum_{d,\delta \le \sqrt{n/\log n}} \pi(a,b)|R| \\
&\le A_2 u^3 \sum_{d,\delta \le \sqrt{n/\log n}} \pi(a,b)(d+\delta)^3 \\
&\le A_2 u^3 \sum_{d,\delta \in \Omega_\nu} \pi(a,b)(d+\delta)^3 \\
&\le 8A_2 u^3 \mathbf{E}(d_{max}^3) \\
&\le A_3 u^3 (\log n)^3
\end{aligned}$$

by an easy computation. Thus for some $A_4 > 0$,

$$\exp\left\{-\frac{A_4(\log n)^3}{n^{3/2}}\right\} \le \frac{\sigma_\nu(u)}{h_\nu(u)} \le \exp\left\{\frac{A_4(\log n)^3}{n^{3/2}}\right\}.$$

It is precisely because of the possibility to bound this ratio without having to take absolute values first, that we work with real u, avoiding the complex valued characteristic function.

We can now easily show that

$$\exp\left\{-\frac{A_4\nu(\log n)^3}{n^{3/2}}\right\} \le \frac{g_\nu(u)}{h_\nu(u)} \le \exp\left\{\frac{A_4\nu(\log n)^3}{n^{3/2}}\right\}.$$

We proceed inductively, starting with g_2 = &2> to get the base case. Then by (33)

$$\begin{aligned} \mathbf{E}(e^{uY_\nu}) &\le e^u \sum_{a,b\in\Omega_\nu} \text{X(«, 6) ft M «) 6 *».(«)} \exp\left\{\frac{A_4(\log n)^3}{n^{3/2}}(d+\delta)\right\} \\ &\le h_\nu(u) \exp\left\{\frac{A_4(\log n)^3}{n^{3/2}}(1+(\nu-2))\right\} \end{aligned}$$

and we have our upper bound for g_ν/h_ν. The lower bound is proved similarly. Setting $\nu = n$, we have finally

$$\begin{aligned} \mathbf{E}(e^{uY_n}) &= \exp\left\{uE_n + \tfrac{1}{2}u^2W_n + O\left(\frac{(\log n)^3}{\sqrt{n}}\right)\right\} \\ &= \exp\left\{\frac{3}{8}n + \frac{1}{192}u^2 n + O\left(\frac{(\log n)^3}{\sqrt{n}}\right)\right\}. \end{aligned}$$

Substituting u = v/v^ we see that

E (v(Yn−(3/8) V ^ = e v²/192+O((logn)³/>A)

for every *real* v and so Y_n Jff $\frac{3}{8}n$ converges to AT(0, ⅛g), together with all its moments, and the proof of Theorem 2 is complete.

To illustrate the power of this result, notice that it leads, for instance,to an asymptotic formula for 22(5^?), exact up to a remainder $O(n^{3/2})$. A direct computation would have required plenty of work, without giving a clear idea of why the final result is so simple.

4 Modified Greedy on $G_{n?m}$

Here we give the proof of Theorem 3. In fact, the method of proof of Theorem 1 carries over with only minimal changes, so we will elaborate only the points of difference. The notation will correspond with that in Section 2. We will use "hats" to indicate quantities which differ from their counterparts in the proof of Theorem 1.

Most importantly, of course, we have a different distribution for Ĵ?, the number of edges deleted at each step. Here it is possible that *no* edge is deleted, since the random vertex choice may select an isolated vertex. Thus we must allow $\hat{R} = -1$. (Recall that the number of deleted edges was $R + 1$.) We will determine the distribution of $\hat{R}$. Clearly

$$Pr(\hat{R} = -1) = \frac{\binom{N_1}{\mu}}{\binom{N}{\mu}}. \tag{38}$$

For $R \geq 0$, suppose that the first vertex selected has degree n + 1 > 0, and its chosen neighbour has degree r^ + 1 > 0. There are

$$\binom{\nu - 1}{r_1 + 1}\binom{\nu - 2}{r_2}$$

ways of selecting the (ri +T2 +1) edges attached to the chosen pair of vertices, and then $\binom{N_2}{\mu - r_1 - r_2 - 1}$ ways of selecting the rest. Thus

$$\begin{aligned} Pv(\hat{R} = r) &= \sum_{r_1 + r_2 = r} \frac{\binom{\nu-1}{r_1+1}\binom{\nu-2}{r_2}\binom{N_2}{\mu-r-1}}{\binom{N}{\mu}} \\ &= \frac{\binom{N_2}{\mu-r-1}}{\binom{N}{\mu}} \sum_{r_1+r_2=r} \binom{\nu-1}{r_1+1}\binom{\nu-2}{r_2} \end{aligned} \tag{39}$$

$$= \frac{\binom{N_2}{\mu-r-1}}{\binom{N}{\mu}}\left(\binom{2\nu-3}{r+1}-\binom{\nu-2}{r+1}\right) \quad (r \geq 0), \tag{40}$$

where, in the last equality, we have made use of a simple combinatorial identity which may be proved easily using generating functions and the binomial theorem. Thus, using approximations similar to those leading to (2), we obtain

$$\Pr(\hat{R}=-1) = e^{-\xi}(1+O(1/\nu)) \tag{41}$$

$$\Pr(\hat{R}=r) = \frac{\xi^{r+1}(2^{r+1}-1)e^{-2\xi}}{(r+1)!}(1+O(r^2/\nu)) \quad (r \geq 0) \tag{42}$$

¿From these we may obtain by straightforward computations

$$\mathbf{E}(\hat{R}) = 2\xi-\xi e^{-\xi}+O(1/\nu), \quad \mathbf{E}(\hat{R}^2) = 4\xi^2-\xi^2 e^{-\xi}+2\xi-\xi e^{-\xi}+O(1/\nu). \tag{43}$$

provided that $\xi = O(1)$. We now let $\hat{g}_{\nu,\mu}(z) = \exp(\nu(z\hat{\phi}(2\nu/\mu)+\frac{1}{2}z^2\hat{\psi}(2\nu/\mu)))$, and we will derive the differential equations analogous to (6) and (7) which $\hat{\phi}, \hat{\psi}$ must satisfy. We will show later that (in the complex domain) these functions are analytic on an open region containing the nonnegative real axis. Hence they and their derivatives are uniformly bounded on the interval $[0, c_0]$. This will justify the Taylor expansions (c.f. (9) and (10))

$$\hat{\phi}(\xi_r) = \hat{\phi}(\xi) + \hat{\phi}'(\xi)\frac{2}{\nu}(\xi-1-r) + O\left(\frac{r^2}{\nu^2}\right) \tag{44}$$

$$\hat{\psi}(\xi_r) = \hat{\psi}(\xi) + \hat{\psi}'(\xi)\frac{2}{\nu}(\xi-1-r) + O\left(\frac{r^2}{\nu^2}\right). \tag{45}$$

Hence (c.f. (11))

$$\ln(e^z\hat{g}_{\nu-2,\mu-1-r}/\hat{g}_{\nu,\mu}) = z(1-2\hat{\phi}(\xi)+2\hat{\phi}'(\xi)(\xi-1-r)) + \tfrac{1}{2}z^2(-2\hat{\psi}(\xi) + 2\hat{\psi}'(\xi)(\xi-1-r) + O(r^2 z/\nu). \tag{46}$$

A similar calculation leads to

$$\text{In\&.n./k,) = *(e\#(0 - } id)) + \text{§*}^{2}\text{(£0}\hat{\ }\text{(O - *(0) + } 0(z/v). \qquad (47)$$

Now (c.f. Lemma 2) we must examine

$$\hat{p}_{-1}(\hat{g}_{\nu-1,\mu}/\hat{g}_{\nu,\mu}) + \sum_{r\geq 0} \hat{p}_r(e^z \hat{g}_{\nu-2,\mu-1-r}/\hat{g}_{\nu,\mu}).$$

As in the proof of Lemma 2, we may ignore r > n^1^^{3} in this sum. (Note that the estimate in (4) remains valid here.) Then we expand uniformly for $r \leq n^{1/3}$, giving

$$\hat{p}_{-1}(1 + z(\xi\hat{\phi}' - \hat{\phi}) + \tfrac{1}{2}z^2((\xi\hat{\phi}' - \hat{\phi})^2 + (\xi\hat{\psi}' - \hat{\psi}) + O(z/\nu))$$
$$+ \sum_{r\geq 0} \hat{p}_r(1 + z(1 - 2\hat{\phi} + 2(\xi - 1 - r)\hat{\phi}') + \tfrac{1}{2}z^2((1 - 2\hat{\phi} + 2(\xi - 1 - r)\hat{\phi}')^2$$
$$- 2\text{^} + 2(\text{f - 1 - r})\text{^'}) + 0(r^2 z/\text{i^})) \qquad (48)$$

We need this quantity to be (1 + $0(z/v)$). The conditions for this are clearly that the coefficients of z and z^2 in (48) must be zero. The coefficent of z is

$$e^{-\xi}(\xi\hat{\phi}' - \hat{\phi}) + (1 - e^{-\xi})(1 - 2\hat{\phi} + 2\xi\hat{\phi}') - 2\hat{\phi}'\mathbf{E}(\hat{R}).$$

Equating this to zero, substituting for E(/̂?) using (43) and rearranging gives

$$\xi\hat{\phi}' + \hat{\phi} = \frac{1 - e^{-\xi}}{2 - e^{-\xi}}. \qquad (49)$$

Similarly, the coefficient of z^2 is

$$e^{-\xi}((\xi\hat{\phi}' - \hat{\phi})^2 + (\xi\hat{\psi}' - \hat{\psi})) + (1 - e^{-\xi})((1 - 2\hat{\phi} + 2\xi\hat{\phi}')^2 - 2\hat{\psi} + 2\xi\hat{\psi}')$$
$$- 4(1 - 2\hat{\phi} + 2\xi\hat{\phi}')\hat{\phi}'\mathbf{E}(\hat{R}) + 4\mathbf{E}(\hat{R}^2)(\hat{\phi}')^2 - 2\hat{\psi}'\mathbf{E}(\hat{R})$$

Again equating this to zero, substituting for $\mathbf{E}(\hat{R})$, $\mathbf{E}(\hat{R}^2)$ using (43) and rearranging gives

$$\begin{aligned}\xi\hat{\psi}' + \hat{\psi} &= \frac{1-e^{-\xi}}{2-e^{-\xi}}(1-2\hat{\phi}+2\xi\hat{\phi}')^2 + \frac{e^{-\xi}}{2-e^{-\xi}}(\xi\hat{\phi}'-\hat{\phi})^2 \\ &\quad + \frac{4\xi^2-\xi^2e^{-\xi}+2\xi-\xi e^{-\xi}}{2-e^{-\xi}}(2\hat{\phi}')^2 - 4\xi(1-2\hat{\phi}+2\xi\hat{\phi}')\hat{\phi}'\end{aligned} \tag{50}$$

We have the initial conditions $\hat{\phi}(0) = \hat{\psi}(0) = 0$. Equation (49) clearly has solution

$$\hat{\phi}(\xi) = \frac{1}{\xi}\int_0^{\xi}\frac{1-e^{-x}}{2-e^{-x}}dx = \frac{1}{2\xi}(\xi - \ln(2-e^{-\xi})), \tag{51}$$

as claimed in the statement of Theorem 3. Now consider (50). If its right hand side is denoted by $u(\xi)$, then it is clear that $u(0) = 0$. Then (50) has the solution

$$\hat{\psi}(\xi) = \frac{1}{\xi}\int_0^{\xi} u(z)dz,$$

such that $\hat{\psi}(0) = 0$. Also, $\hat{\psi}(\infty) = 0$ since $\int_0^{\infty} u(z)dz$ converges.

The rest of the proof of Theorem 3 follows closely the lines of that of Theorem 1, as the reader may check.

5 Modified Greedy on random trees

We now consider the proof of Theorem 4. The proof is similar to that of Theorem 2 and so we will give somewhat fewer details. We use the same notation as that in Section 3 except that we will put a ˆ over the corresponding quantity. Then as in equation (17) we have, for $n \geq 2$,

$$\hat{f}_n(z) = z\sum_{k=1}^{n-1}\sum_{d=0}^{\infty}\sum_{\delta=0}^{\infty}\sum_{a\in\Omega(d,k)}\sum_{b\in\Omega(\delta,n-k)}\hat{\pi}(a,b)\prod_{r=1}^{d}\hat{f}_{a_r}(z)\prod_{s=1}^{\delta}\hat{f}_{b_s}(z).$$

where now

$$\hat{\pi}(a,b) = \frac{1}{n^{n-1}}\binom{n-2}{k-1}\sum_{v=1}^{n}\sum_{w\neq v}\left(\frac{(k-1)!}{(d+1)!}\prod_{r=1}^{d}\frac{a_r^{a_r-1}}{a_r!}\right)\left(\frac{(n-k-1)!}{\delta!}\prod_{s=1}^{\delta}\frac{b_s^{b_s-1}}{b_s!}\right).$$

Explanation n^{n-1} counts (tree, chosen vertex) pairs; v denotes the chosen vertex; vw is the chosen edge; having fixed v, w there are $\binom{n-2}{k-1}$ choices for the vertices of the trees attached to v; in the tree in question v has degree $d+1$ and $\frac{1}{d+1}$ is the probability that vw is the chosen edge. The remaining terms count the number of possible forests on the remaining vertices.

Putting

$$\begin{aligned}
\hat{F}(x,z) &= \sum_{n=1}^{\infty}\frac{n^{n-1}\hat{f}_n(z)}{n!}x^n \\
&= x + x^2 z\left(\sum_{d=0}^{\infty}\frac{1}{(d+1)!}\sum_{k=1}^{\infty}\sum_{a\in\Omega(d,k)}\prod_{r=1}^{d}\frac{a_r^{a_r-1}f_{a_r}(z)}{a_r!}x^{a_r}\right)\times \\
&\quad \left(\sum_{\delta=0}^{\infty}\frac{1}{\delta!}\sum_{k=1}^{\infty}\sum_{b\in\Omega(\delta,k)}\prod_{s=1}^{\delta}\frac{b_s^{b_s-1}f_{b_s}(z)}{b_s!}x^{b_s}\right)
\end{aligned}$$

we obtain the equation

$$\hat{F} = x + x^2 z\left(\frac{e^{\hat{F}}-1}{\hat{F}}\right)e^{\hat{F}}$$

or

$$\hat{F}^2 = x\hat{F} + x^2 z(e^{2\hat{F}} - e^{\hat{F}}). \tag{52}$$

Unlike GREEDY, this is not a differential equation! As a partial check, set $z = 1$ and notice that $\hat{F}(x,1) = T(x)$, so that (53) becomes

$$T^2 = xT + x^2(e^{2T} - e^T).$$

This is certainly correct since $T = xe^T$, see (21). Differentiating (52) with respect to z gives

$$2\hat{F}\frac{\partial \hat{F}}{\partial z} = x\frac{\partial \hat{F}}{\partial z} + x^2(e^{2\hat{F}} - e^{\hat{F}}) + x^2 z\frac{\partial \hat{F}}{\partial z}(2e^{2\hat{F}} - e^{\hat{F}})$$

or

$$\frac{\partial \hat{F}}{\partial z} = \frac{x^2(e^{2\hat{F}} - e^{\hat{F}})}{2\hat{F} - x - x^2 z(2e^{2\hat{F}} - e^{\hat{F}})}. \tag{53}$$

Now

$$\hat{F}(x,1) = T(x)$$

and

$$\frac{\partial \hat{F}}{\partial z}(x,1) = \hat{E}(x) = \sum_{n=1}^{\infty} \hat{E}_n \frac{n^{n-1}}{n!} x^n$$

where

$$\hat{E}_n = \mathbf{E}(\hat{Y}_n).$$

So putting $z = 1$ in (53) we obtain, using (21), that

$$\begin{aligned} \hat{E} &= \frac{x^2(e^{2T} - e^T)}{2T - x - x^2(2e^{2T} - e^T)} \\ &= \frac{T - Te^{-T}}{(2 - e^{-T})(1 - T)}. \end{aligned} \tag{54}$$

Thus

$$\begin{aligned} \hat{E}_n &= \frac{n!}{n^{n-1}} \frac{1}{2\pi i} \int_{C_0} \frac{T - Te^{-T}}{(2 - e^{-T})(1 - T)} \frac{dx}{x^{n+1}} \\ &= \frac{n!}{n^{n-1}} \frac{1}{2\pi i} \int_{C_1} \frac{1 - e^{-t}}{2 - e^{-t}} \frac{e^{nt}}{t^n} dt. \end{aligned}$$

Here C_0 is a circular contour in the x plane, of radius less than e^{-1}.(Notice that $|2 - e^{-T(x)}|$ is bounded away from zero in the closed disc $|x| \le e^{-1}$.) As for C_1, it consists of two circular arcs L_1 and L_2 , where $L_1 = \{t = e^{i\theta} :$

$-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$, and L_2 passes through the points $e^{\pm i\pi/2}$ and the point t_0 on the negative real line, such that $-\ln 2 < t_0 < 0$. The contribution of L_2 to the value of $\hat{E}_n$ is of order $O(n^{3/2}(e|t_0|)^{-n})$, which is exponentially small, provided that $|t_0|$ is sufficiently close to $\ln 2$ (because $e\ln 2 > 1$.) We now expand $\frac{1-e^{-t}}{2-e^{-t}}$ around $t = 1$ to obtain

$$\frac{1-e^{-t}}{2-e^{-t}} = \frac{e-1}{2e-1} + \frac{e}{(2e-1)^2}(t-1) - \frac{e(2e+1)}{2(2e-1)^3}(t-1)^2 + \frac{e(4e^2+8e+1)}{6(2e-1)^4}(t-1)^3 + (t-1)^4\phi(t)$$

where ϕ is bounded on L_1.

While integrating the first four summands we can, and do, extend the integral over the whole unit circle making an exponentially small error. So using (26) and (27) we now obtain

$$\begin{aligned} \hat{E}_n &= \frac{e-1}{2e-1}n + \frac{e}{(2e-1)^2} - \frac{e(2e+1)}{2(2e-1)^3}\frac{2}{n} + \frac{e(4e^2+8e+1)}{6(2e-1)^4}\frac{3}{n^2} + O\left(\frac{1}{n}\right) \\ &= \frac{e-1}{2e-1}n - \frac{3e-2e^2}{2(2e-1)^3} + O\left(\frac{1}{n}\right) \end{aligned} \quad (55)$$

as claimed.

We continue by estimating $\hat{V}_n = E(K(K-1))$. Differentiating (52) twice with respect to z gives

$$\begin{aligned} & 2\left(\frac{\partial\hat{F}}{\partial z}\right)^2 + \hat{F}\frac{\partial^2\hat{F}}{\partial z^2} \\ &= x\frac{\partial^2\hat{F}}{\partial z^2} + x^2\left(2e^{2\hat{F}} - e^{\hat{F}}\right)\left(2\frac{\partial\hat{F}}{\partial z} + z\frac{\partial^2\hat{F}}{\partial z^2}\right) + x^2 z\left(\frac{\partial\hat{F}}{\partial z}\right)^2\left(4e^{2\hat{F}} - e^{\hat{F}}\right). \end{aligned}$$

Putting $z = 1$ and using (21) we obtain

$$2\hat{E}^2 + T\hat{V} = x\hat{V} + x^2(2e^{2T} - e^{T})(2\hat{E} + \hat{V}) + x^2\hat{E}^2(4e^{2T} - e^{T})$$

$$= Te^{-T}\hat{V} + T^2(2 - e^{-T})(2\hat{E} + \hat{V}) + \hat{E}^2T^2(4 - e^{-T})$$

where $\hat{V} = \hat{V}(x) = \sum_{n=1}^{\infty} \hat{V}_n x^n$.

Thus

$$\hat{V} = \frac{T\hat{E}^2(2 - e^{-T}) + 2T^2\hat{E}(2 - e^{-T}) - 2\hat{E}^2}{T(1-T)(2-e^{-T})} \tag{56}$$

Also, as in (30),

$$P_n = \frac{1}{2\pi i} \int_{C_1} \frac{1-t}{t} \cdot \frac{e^{nt}}{t} \tag{57}$$

It follows from (54) and (56) that in terms of t

$$\frac{1-t}{t}\hat{V} = \left(\frac{e-1}{2e-1}\right)^2 \frac{1}{(1-t)^2} + \frac{1-5e+17e^2-20e^3+8e^4}{(2e-1)^4} + \alpha(1-t) + (t-1)^2\beta(t)$$

where α is an absolute constant and β is bounded on C_1. Thus, using (26), (27) and (57)

$$\hat{V}_n = \left(\frac{e-1}{2e-1}\right)^2 n^2 + \frac{1-5e+17e^2-20e^3+8e^4}{(2e-1)^4} n + O(1).$$

Using this and (55) we obtain the variance estimate given in the theorem.

We finally consider asymptotic normality. Fortunately, no work is needed. If we examine the proof of asymptotic normality in Theorem 2 we see that all we need do is replace $\pi(a,b)$ by $\hat{\pi}(a,b)$ throughout to obtain the result for MODIFIED GREEDY.

References

[1] **I. M. Curtis,** ***A note on the theory of moment generating functions,*** Annals of mathematical Statistics 13 (1972) 430-433.

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

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

[4] 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.

[5] 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.

[6] H. M. Mahmoud and B. Pittel, ***Analysis of the space of search trees under the random insertion algorithm,*** **Journal of Algorithms 10 (1989)** 52-75.

[7] A. Meir and J. W. Moon, ***The expected node-independence number of*** *random trees* INCOMPLETE REF

[8] B. Pittel, ***On tree census and the giant component in sparse random*** *graphs,* Random Structures and Algorithms 1 (1990) 311-332.

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

[10] J. W. Moon, *Counting labelled trees,* Canadian Mathematical Congress, Montreal, 1970.