

Rainbow Greedy Matching Algorithms

Patrick Bennett*

Department of Mathematics
Western Michigan University
Kalamazoo MI 49008-5248

Colin Cooper

Department of Informatics
King's College London
London WC2B 4BG

Alan Frieze[†]

Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh PA 15213

July 2, 2023

Abstract

We consider the problem of finding a large rainbow matching in a random graph with randomly colored edges. In particular we analyze the performance of two greedy algorithms for this problem. The algorithms we study are colored versions of algorithms that were previously used to find large matchings in random graphs (i.e. the color-free version of our present problem).

1 Introduction

In this short note, we discuss greedy algorithms for finding rainbow matchings in sparse random graphs. Thus we start with the random graph $G_{n,m}$, $m = cn/2$ where $c > 0$ is a constant and then color each edge uniformly at random (u.a.r.) from a set of colors $Q = [q]$. A set S of edges is said to be *rainbow colored* if every edge in S has a different color. The decision problem for whether a colored graph has a rainbow matching of size k is NP-complete [2]. Here we discuss the efficacy of simple greedy algorithms for finding large rainbow matchings.

*Research supported in part by Simons Foundation Grant #426894.

[†]Research supported in part by NSF grant DMS1952285

The color-free version of this problem has been studied in several previous papers. Dyer, Frieze and Pittel [5] studied two greedy algorithms for finding large matchings. The first algorithm (Greedy Matching) repeatedly chooses an edge $\{x, y\}$ u.a.r., adds it to the current matching and deletes the vertices x, y . This continues until the remaining graph has no edges. They showed that with high probability (i.e. with probability tending to 1 as n grows, henceforth abbreviated w.h.p.) this algorithm produces a matching of size asymptotic to $\frac{n}{2} \left(1 - \frac{1}{c+1}\right)$. They also considered a variation (Modified Greedy Matching) where the algorithm first chooses a vertex uniformly at random, and then chooses a uniform random incident edge and then updates the matching. This algorithm does slightly better than the first, it produces a matching of size asymptotic to $\frac{n}{2} \left(1 - \frac{\log(2-e^{-c})}{2c}\right)$. Further improvements were obtained by Karp and Sipser [7], using KSGreedy, a modification of Greedy Matching. KSGreedy chooses a random vertex of degree one, if there is one, and adds its incident edge to the matching; otherwise it chooses a random edge and adds it to the matching. Karp and Sipser studied this algorithm in $G_{n,p}, p = c/n$ and showed that it produced a matching of asymptotically maximal size. The Karp-Sipser algorithm was further studied for $G_{n,c/n}$ by Aronson, Frieze and Pittel [1], who showed that the algorithm found a matching within $O(n^{1/5} \log^{O(1)} n)$ of the maximum.

We will prove the following theorems which we prefix with a formal statement of the algorithms.:

Greedy Algorithm. Formally the algorithm proceeds as follows:

GREEDY

begin

$M \leftarrow \emptyset;$

while $E(G) \neq \emptyset$ **do**

begin

Choose $e = \{u, v\} \in E$ u.a.r.; $M \leftarrow M \cup \{e\};$

$V \leftarrow V \setminus \{u, v\}; G \leftarrow G[V]$

$F \leftarrow \{f \in E : c(f) = c(e)\}; E \leftarrow E \setminus F;$

end;

Output M

end

Theorem 1. *Suppose that $q = \kappa n$. Let μ denote the size of the matching produced by GREEDY, then following hold w.h.p.*

(a) *If $\kappa = 1/2$ then $\mu \sim \frac{1}{2} \left(1 - \frac{1}{(2c+1)^{1/2}}\right) n$.*

(b) *If $\kappa = \frac{1}{2}(1 + \varepsilon)$, where $|\varepsilon| > 0$ then $\mu \sim \frac{1}{2} \left(1 - \frac{1+O(\varepsilon)}{(2c+1)^{1/2}}\right) n$.*

(c) *If $\kappa < 1/2c$ and $c > 5$ then $\kappa(1 - e^{-c(1/2\kappa-2)})n < \mu < \kappa(1 - e^{-c/2\kappa})n$.*

(d) *If $\kappa \gg 1$ then $\mu \sim \frac{1}{2} \left(1 - \frac{1}{2c+1+\varepsilon_\kappa}\right) n$ where $\varepsilon_\kappa = \frac{\log(c+1)}{2\kappa-1} - \frac{c}{2\kappa} + O\left(\frac{c}{\kappa^2}\right)$.*

Modified greedy algorithm. The modified greedy algorithm is formally described as

MODIFIED GREEDY

```

begin
   $M \leftarrow \emptyset$ ;
  while  $E(G) \neq \emptyset$  do
    begin
      Choose  $v \in V$  u.a.r.;
      If  $N(v) = \emptyset$ ,  $V \leftarrow V \setminus \{v\}$  ;
      Else if  $N(v) \neq \emptyset$  choose  $u \in N(v)$  u.a.r.;
      Let  $e = \{u, v\}$ ;  $M \leftarrow M \cup \{e\}$ ;  $V \leftarrow V \setminus \{u, v\}$  ;
       $F \leftarrow \{f \in E : c(f) = c(e)\}$  ;  $E \leftarrow E \setminus F$ ;
       $G \leftarrow G[V]$ ;
    end;
  Output  $M$ 
end

```

Theorem 2. *Suppose that $q = \kappa n$. Let μ be the size of the matching produced by MODIFIED GREEDY. Then the following hold w.h.p.*

(a) Let $N(\tau)$ be the solution of

$$N(\tau) = 1 - 2\tau + \int_0^\tau \exp \left\{ -\frac{c}{\kappa} N(\sigma)(N(\sigma) + \sigma + \kappa - 1) \right\} d\sigma,$$

and let $\tau_0 \in [0, 1]$ be the solution to $N(\tau) = 0$. Then $\mu \sim (1 - \tau_0)n$.

(b) If $\kappa \geq 1/2$ then

$$\mu \leq \frac{c - 1 + e^{-c}}{2c - 1 + e^{-c}} n.$$

Numerical calculations suggest that the matching obtained by MODIFIED GREEDY is significantly larger than by GREEDY. In the following we have fixed $\kappa = 1/2$ so that we can use Theorem 1(a).

c	GREEDY	MODIFIED GREEDY
0.5	0.092	0.148
1.0	0.146	0.216
1.5	0.184	0.257
2.0	0.211	0.285
2.5	0.233	0.316
3.0	0.250	0.322
3.5	0.264	0.334
4.0	0.276	0.345
4.5	0.287	0.355
5.0	0.296	0.361

Conjecture: Given the above table we conjecture that if μ_G and μ_{MG} are the sizes of the matchings produced by GREEDY and MODIFIED GREEDY respectively, then w.h.p. $\mu_G \leq \mu_{MG}$.

2 GREEDY

Let $G(t)$ denote the unmatched graph remaining after t iterations, let $\nu(t)$ be the number of vertices, and let $\mu(t)$ denote the number of edges in $G(t)$. At each step t , we choose a random edge $\{x, y\}$, add it to the matching $M(t)$, delete the vertices x and y from $V(t)$, and delete all edges of the same color as $\{x, y\}$. Let $d_t(\cdot)$ denote degree in G_t . For the sake of our analysis we reveal the random graph and colors as we run the algorithm. More specifically, at each step t we reveal our matching edge e_t by choosing a random pair of distinct vertices. Then for each of the other $\mu(t) - 1$ other edges e' we reveal whether or not e' shares an endpoint with e . Any e' meeting e is deleted. Conditional on the matching edge e and the (say) k deleted edges, the remaining edges comprise a uniform random set of $\mu(t) - 1 - k$ edges on the remaining set of $\nu(t) - 2$ vertices.

A priori we do not know the degrees of any of the vertices. We just know that at step t we have $\nu(t)$ vertices and a uniform random set of $\mu(t)$ edges. We reveal the location of one of these edges, which is equally likely to have any two distinct endpoints among the $\nu(t)$ vertices. We know there is an edge there just because we said we were revealing the location of an edge. Of course, after we reveal the location of that edge, we know that its two endpoints must have degree at least 1. But we only know that because we revealed the edge.

We reveal the color of e_t by choosing a random color from among the unused colors. Finally we reveal any other edges of that same color and delete them. Thus we have

$$\mathbb{E}[\mu(t+1) \mid \mu(t)] = \left(\mu(t) - \mathbb{E}(d_t(x) + d_t(y) - 1 \mid \mu(t)) \right) \left(1 - \frac{1}{q-t} \right). \quad (1)$$

Note that the number of vertices at step t is $\nu(t) = n - 2t$. We will assume (justified later) that

$$\frac{t}{n} \leq \min \left\{ \frac{1}{2}, \kappa \right\} - \Omega(1)$$

so that $\nu(t), q - t \geq \Omega(n)$. Then

$$\begin{aligned} \mathbb{E}(d_t(x) \mid \mu(t)) &= \mathbb{E}(d_t(y) \mid \mu(t)) = 1 + (\nu(t) - 2) \cdot \frac{\mu(t) - 1}{\binom{\nu(t)}{2} - 1} \\ &= 1 + \frac{2(\mu(t) - 1)}{\nu(t) + 1} \\ &= 1 + \frac{2\mu(t)}{\nu(t)} + O(n^{-1}). \end{aligned}$$

So, picking up from (1) we have

$$\begin{aligned}\mathbb{E}[\mu(t+1) \mid \mu(t)] &= \left(\mu(t) - 1 - \frac{4\mu(t)}{\nu(t)} + O(n^{-1}) \right) \left(1 - \frac{1}{q-t} \right) \\ &= \mu(t) - 1 - \frac{4\mu(t)}{\nu(t)} - \frac{\mu(t)}{q-t} + O(n^{-1})\end{aligned}\quad (2)$$

This leads us to consider the differential equation ($t = \tau n$, $M(\tau) = \mu(t)/n$ here) which will simulate the process w.h.p.

$$\frac{dM}{d\tau} = -1 - \frac{4M(\tau)}{1-2\tau} - \frac{M(\tau)}{\kappa-\tau}, \quad M(0) = \frac{c}{2} \text{ where } \kappa = \frac{q}{n}.\quad (3)$$

We let τ_0 be the smallest positive root of $M(\tau) = 0$ where M is the solution to the above initial value problem. We will show that w.h.p. the process ends with a rainbow matching with $\tau_0 n + o(n)$ edges. Unsurprisingly, we will not come close to a perfect matching, nor will we come close to using every color. In particular we will see in Section 2.2 that for fixed c, k we have

$$\tau_0 \leq \min \left\{ \frac{1}{2}, \kappa \right\} - \Omega(1).$$

To do this we will apply the following theorem of Warnke [8].

Theorem 3 (Warnke Theorem 2 and Lemma 11). *Let $a, n > 1$ be integers. Let $\mathcal{D} \subseteq \mathbb{R}^{a+1}$ be a connected and bounded open set. Let $(F_k)_{1 \leq k \leq a}$ be functions with $F_k : \mathcal{D} \rightarrow \mathbb{R}$. Let $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots$ be σ -fields. Suppose that the random variables $(Y_k(t))_{1 \leq k \leq a}$ are nonnegative and \mathcal{F}_t -measurable for $t > 0$. Furthermore, assume that, for all $t > 0$ and $1 \leq k \leq a$, the following conditions hold whenever $(t/n, Y_1(t)/n, \dots, Y_a(t)/n) \in \mathcal{D}$:*

- (i) $|\mathbb{E}[Y_k(t+1) - Y_k(t) \mid \mathcal{F}_t] - F_k(t/n, Y_1(t)/n, \dots, Y_a(t)/n)| \leq \delta$, where the function F_k is L -Lipschitz-continuous on \mathcal{D} (the ‘Trend hypothesis’ and ‘Lipschitz hypothesis’),
- (ii) $|Y_k(t+1) - Y_k(t)| \leq \theta$ and $\mathbb{P}(|Y_k(t+1) - Y_k(t)| > \phi \mid \mathcal{F}_t) \leq \gamma$ (the ‘Boundedness hypothesis’), and that the following condition holds initially:
- (iii) $\max_{1 \leq k \leq a} |Y_k(0) - \hat{y}_k n| \leq \varepsilon n$ for some $(0, \hat{y}_1, \dots, \hat{y}_a) \in \mathcal{D}$ (the ‘Initial condition’).

Suppose $R \in [1, \infty)$ and $T \in (0, \infty)$ satisfy $t \leq T$ and $|F_k(z)| \leq R$ for all $1 \leq k \leq a$ and $z = (\tau, y_1, \dots, y_a) \in \mathcal{D}$. Then for $\varepsilon > (\delta + \gamma\theta) \min\{T, L^{-1}\} + R/n$, with probability at least

$$1 - 2a \exp \left\{ -\frac{n\varepsilon^2}{8T\phi^2} \right\} - aTn\gamma\quad (4)$$

we have

$$\max_{0 \leq t \leq \sigma n} \max_{1 \leq k \leq a} |Y_k(t) - y_k(t/n)n| < 3 \exp\{LT\} \varepsilon n$$

where $(y_k(\tau))_{1 \leq k \leq a}$ is the unique solution to the system of differential equations $y'_k(\tau) = F_k(\tau, y_1(\tau), \dots, y_a(\tau))$ with $y_k(0) = \hat{y}_k$ for $1 \leq k \leq a$, and $\sigma = \sigma(\hat{y}_1, \dots, \hat{y}_a) \in [0, T]$ is any choice of $\sigma > 0$ with the property that $(\tau, y_1(\tau), \dots, y_a(\tau))$ has L_∞ -distance at least $3 \exp\{LT\}\varepsilon$ from the boundary of \mathcal{D} for all $\tau \in [0, \sigma)$.

The above theorem is a version of Theorem 5.1 of Wormald [9]. We use Warnke's version here because of the condition (ii). The probability bound given by Wormald's theorem is not good enough for us if our variables could see a large one-step change. In particular, for our process our matching edge could have an endpoint with linear degree, causing the loss of a linear number of edges in a single step. However, since the probability of that event is so small, Warnke's version can handle it. Of course, Wormald [9] discusses similar situations and describes how to handle them, but for us it is more convenient to use Warnke's version since it allows us to apply a single theorem as a black box.

We apply Theorem 3 with $a = 1$ to the random variable $Y_1(t) = \mu(t)$. We let

$$\mathcal{D} = \{(\tau, M) : 0 \leq \tau \leq \tau_0, 0 \leq M \leq c\}.$$

Now we make sure to satisfy condition (i). By (2) we can let $\delta = O(n^{-1})$. From (3) we have

$$F_1(\tau, M) = -1 - \frac{4M}{1 - 2\tau} - \frac{M}{\kappa - \tau}$$

which is L -Lipschitz on \mathcal{D} for $L = O(1)$ (here we use the fact that the denominators $1 - 2\tau, \kappa - \tau$ are bounded away from 0). We now move on to condition (ii). We have to take $\theta = O(n)$ since it is possible (though very unlikely) to have a vertex of linear degree. We let $\phi = 2n^{0.1}$, and to find a suitable γ we bound

$$\begin{aligned} \mathbb{P}(|\mu(t+1) - \mu(t)| > 2n^{0.1} \mid \mu(t)) &\leq 2 \frac{\binom{\nu(t)-1}{n^{0.1}} \binom{\binom{\nu(t)}{2} - n^{0.1}}{\mu(t) - n^{0.1}}}{\binom{\binom{\nu(t)}{2}}{\mu(t)}} \\ &= 2 \frac{\binom{\nu(t)-1}{n^{0.1}} (\mu(t))_{n^{0.1}}}{\left(\binom{\nu(t)}{2}\right)_{n^{0.1}}} \\ &\leq 2 \frac{\left(\frac{\nu(t)\varepsilon}{n^{0.1}}\right)^{n^{0.1}} (\mu(t))^{n^{0.1}}}{\left(\binom{\nu(t)}{2} - n^{0.1}\right)^{n^{0.1}}} \\ &= 2 \left[O\left(\frac{\mu(t)}{n^{0.1}\nu(t)}\right) \right]^{n^{0.1}} \\ &= \exp\{-\Omega(n^{0.1})\} = \gamma. \end{aligned} \tag{5}$$

Now for condition (iii), any positive ε will do (we will have to choose ε more carefully later to satisfy future conditions). We can very comfortably choose $T = 1$, and $R = O(1)$. Now we choose

$$\varepsilon = n^{-0.1} > (\delta + \gamma\theta) \min\{T, L^{-1}\} + R/n$$

and so the probability bound in (4) goes to 1. Thus with high probability we have

$$\mu(t) = M(\tau)n + O(n^{0.9}) \quad (6)$$

uniformly for $0 \leq t \leq \tau_0 n + O(n^{0.9})$, where $M(\tau)$ is the solution to (3) and $\tau = t/n$. So w.h.p. the Greedy algorithm will produce a rainbow matching of $\tau_0 n + O(n^{0.9})$.

2.1 Solution to the differential equation

When $q = n/2$ (i.e. $\kappa = 1/2$) the solution to (3) is

$$M(\tau) = \frac{(2c+1)(1-2\tau)^3 - (1-2\tau)}{4}.$$

The smallest root of the above is

$$\tau_0 = \frac{1}{2} - \frac{1}{2(2c+1)^{1/2}}. \quad (7)$$

When $\kappa \neq 1/2$ we write (3) as

$$\frac{dm}{d\tau} + M(\tau) \left(\frac{4}{1-2\tau} + \frac{1}{\kappa-\tau} \right) = -1.$$

Let $I = -2 \log(1-2\tau) - \log(\kappa-\tau)$, then

$$M(t) = e^{-I} \left(B - \int e^I d\tau \right),$$

where B solves $M(0) = c/2$. The solution for $\kappa = 1/2$ is given above.

For $k \neq 1/2$, let $A = 1/(2\kappa-1)^2$ then

$$\begin{aligned} \int e^I d\tau &= \int \frac{1}{(\kappa-\tau)(1-2\tau)^2} \\ &= \int A \left(\frac{1}{\kappa-\tau} - \frac{2}{1-2\tau} + \frac{4\kappa-2}{(1-2\tau)^2} \right) \\ &= A \left(-\log(\kappa-\tau) + \log(1-2\tau) + \frac{2\kappa-1}{1-2\tau} \right). \end{aligned}$$

Thus $B = c/2\kappa + A((2\kappa-1) - \log \kappa)$ and for $\kappa \neq 1/2$

$$M(\tau) = \frac{(\kappa-\tau)(1-2\tau)^2}{(2\kappa-1)^2} \left(\frac{c(2\kappa-1)^2}{2\kappa} + \log \frac{\kappa-\tau}{\kappa(1-2\tau)} - \frac{2(2\kappa-1)\tau}{1-2\tau} \right). \quad (8)$$

2.2 Asymptotics for τ_0 from $M(\tau) = 0$

From (8) let

$$f_\kappa(\tau) = \left(\frac{c(2\kappa - 1)^2}{2\kappa} + \log \frac{\kappa - \tau}{\kappa(1 - 2\tau)} - \frac{2(2\kappa - 1)\tau}{1 - 2\tau} \right). \quad (9)$$

We consider the solution τ_0 to $f_\kappa(\tau) = 0$ in three cases.

Case 1 As $\kappa \rightarrow 1/2$, in Lemma 4 we show that τ_0 tends to the value in (7), and give the rate of convergence.

Case 2 For κ large, $M(\tau) = 0$ in (8) satisfies $c = \frac{2\tau}{1-2\tau} + O(1/\kappa)$ which implies that w.h.p. the Greedy algorithm will produce a rainbow matching of size $\sim \left(\frac{1}{2} - \frac{1}{2(c+1)} + O(1/\kappa) \right) n$. In Lemma 5 we give a detailed asymptotic which also bounds τ_0 for finite $\kappa \geq 1$. Thus in the limit as $\kappa \rightarrow \infty$ we obtain a matching of size $nc/2(c+1)$; the value obtained without the coloring constraint.

Case 3 For $\kappa \ll 1/2$, and fixed c the solution to $M(\tau) = 0$ in (8) satisfies $\tau \sim \kappa(1 - e^{-c/2\kappa})$. Lemma 6 gives more detail.

Lemma 4. *If $\kappa = \frac{1}{2}(1 + \varepsilon)$, where $|\varepsilon| > 0$, the solution to $M(\tau) = 0$ in (8) is $\tau_0 = \frac{1}{2} \left(1 - \frac{1+O(\varepsilon)}{\sqrt{2c+1}} \right)$.*

Proof. Put $h = (2\kappa - 1)/(1 - 2\tau)$ so that $\tau = \frac{1}{2}(1 - (2\kappa - 1)/h)$. From (9),

$$f_\kappa(h) = \alpha - h + \log(1 + h), \quad \text{where} \quad \alpha = c \frac{(2\kappa - 1)^2}{2\kappa} + \log \frac{1}{2\kappa} + (2\kappa - 1).$$

Case 1a: $2\kappa < 1$.

We have $\tau \leq \kappa$ and so $2\kappa \geq (1 - (2\kappa - 1)/h)$ or $2\kappa - 1 \geq -(2\kappa - 1)/h$. If $2\kappa - 1 < 0$ then this implies that $0 > h > -1$.

Then

$$h - \frac{h^2}{2} + \frac{h^3}{2} \leq \log(1 + h) \leq h - \frac{h^2}{2}.$$

The lower bound term $h^3/2$ holds for $|h| \leq 1/3$, by comparison with a geometric series. Thus

$$\alpha - \frac{|h|^2}{2} - \frac{|h|^3}{2} \leq 0 \leq \alpha - \frac{|h|^2}{2},$$

so that

$$|h| \leq \sqrt{2\alpha}, \quad \text{and} \quad |h| \geq \sqrt{\frac{2\alpha}{1 + |h|}} \geq \sqrt{\frac{2\alpha}{1 + \sqrt{2\alpha}}}.$$

Case 1b: $2\kappa > 1, 0 < h < 1$.

Then $h > 0$, and now

$$h - \frac{h^2}{2} \leq \log(1 + h) \leq h - \frac{h^2}{2} + \frac{h^3}{2}.$$

Note that these inequalities hold for all $h > 0$.

So

$$\alpha - \frac{h^2}{2} \leq 0 \leq \alpha - \frac{h^2}{2} + \frac{h^3}{2},$$

leading to

$$\sqrt{2\alpha} \leq h \leq \sqrt{\frac{2\alpha}{1-h}} \leq \sqrt{\frac{2\alpha}{1-\sqrt{2\alpha}}}. \quad (10)$$

Finally

$$\begin{aligned} \alpha &= \varepsilon^2(c + \frac{1}{2}) - \varepsilon^3(c + 1/3) + O(c\varepsilon^4), \\ h &= (2\alpha)^{1/2}(1 + O(\alpha^{1/2})) = \varepsilon\sqrt{2c+1}(1 + O(\varepsilon)), \\ \tau_0 &= \frac{1}{2} \left(1 - \frac{\varepsilon}{h}\right) = \frac{1}{2} \left(1 - \frac{1 + O(\varepsilon)}{\sqrt{2c+1}}\right). \end{aligned}$$

□

Lemma 5. Assume $\kappa \geq 1$ and $c \geq c^* = (e-1) \left(\frac{2\kappa}{2\kappa-1}\right)^2$. The solution to $M(\tau) = 0$ in (8) is

$$\tau_0 = \frac{1}{2} \left(1 - \frac{1}{(c+1) - \frac{c}{2\kappa} + \frac{\log(c+1)}{2\kappa-1} + O\left(\frac{c}{\kappa^2}\right)}\right).$$

Proof. Put $z = (\kappa - \tau)/(\kappa(1 - 2\tau))$ so that $\tau = \kappa(z - 1)/(2\kappa z - 1)$. Provided $\kappa \geq 1/2$, $z \geq 1$ and $z \rightarrow 1/(1 - 2\tau)$ as $\kappa \rightarrow \infty$. Then (9) becomes

$$f_\kappa(z) = c \frac{(2\kappa - 1)^2}{2\kappa} + 2\kappa + \log z - 2\kappa z, \quad (11)$$

and $f_\kappa(z) = 0$ iff

$$z = \beta + \frac{\log z}{2\kappa} \quad \text{where} \quad \beta = c \left(\frac{2\kappa - 1}{2\kappa}\right)^2 + 1. \quad (12)$$

For $\delta < 1$ put

$$z = \beta + \frac{(1 + \delta)}{2\kappa} \log \beta,$$

and equate. At equality (12) becomes

$$\begin{aligned} \beta + \frac{(1 + \delta)}{2\kappa} \log \beta &= \beta + \frac{1}{2\kappa} \log \left(\beta + \frac{(1 + \delta)}{2\kappa} \log \beta \right) \\ &= \beta + \frac{1}{2\kappa} \log \beta + \frac{1}{2\kappa} \log \left(1 + \frac{(1 + \delta)}{2\kappa\beta} \log \beta \right). \end{aligned}$$

Thus

$$\delta \log \beta = \log \left(1 + \frac{(1 + \delta)}{2\kappa\beta} \log \beta \right).$$

The conditions $\delta < 1$, $\kappa \geq 1$ and $c \geq c^*$ imply that $\beta \geq e$ so that $(\log \beta)/\beta \leq 1/e$. Thus $\frac{(1+\delta)}{2\kappa\beta} \log \beta < 1$. If $x < 1$, then

$$\frac{x}{1+x} \leq \log(1+x) \leq x,$$

which (after canceling a $\log \beta$ term) means δ must satisfy

$$\frac{1+\delta}{2\kappa\beta + (1+\delta)\log \beta} \leq \delta \leq \frac{1+\delta}{2\kappa\beta}.$$

The RHS inequality is true if $\delta \leq 1/(2\kappa\beta - 1)$.

For $c \geq c^*$, $\log \beta \geq 1$, and we can strengthen the LHS to

$$\frac{1+\delta}{2\kappa\beta + (1+\delta)} \leq \delta.$$

This reduces to $1 \leq 2\delta\kappa\beta + \delta^2$, which is satisfied for $\delta \geq 1/2\kappa\beta$.

We conclude that

$$z_0 = \beta + \frac{(1+\delta)}{2\kappa} \log \beta, \quad \text{where} \quad \frac{1}{2\kappa\beta} \leq \delta \leq \frac{1}{2\kappa\beta - 1}.$$

Thus

$$\begin{aligned} \tau_0 &= \frac{\kappa(z_0 - 1)}{2\kappa z_0 - 1} = \frac{1}{2} \left(1 - \frac{2\kappa - 1}{2\kappa z_0 - 1} \right) \\ &= \frac{1}{2} \left(1 - \frac{2\kappa - 1}{2\kappa\beta + (1+\delta)\log \beta - 1} \right) \\ &= \frac{1}{2} \left(1 - \frac{2\kappa - 1}{c^{\frac{(2\kappa-1)^2}{2\kappa}} + 2\kappa - 1 + \frac{1+\delta}{\delta} \log \left(1 + \frac{1+\delta}{2\kappa\beta} \log \left(c + 1 - \frac{1}{2\kappa} + \frac{1}{\kappa^2} \right) \right)} \right) \\ &= \frac{1}{2} \left(1 - \frac{1}{c + 1 - \frac{c}{2\kappa} + \frac{\log(c+1)}{2\kappa-1} + O\left(\frac{c}{\kappa^2}\right)} \right). \end{aligned}$$

□

Lemma 6. Let $\kappa < 1/2c$ where $c > c_0 = 5$. Let τ_0 be the smallest positive root of (9). Then

$$\kappa(1 - e^{-c/2\kappa+2c}) < \tau_0 < \kappa(1 - e^{-c/2\kappa}).$$

Proof. Let $\tau = \kappa(1 - e^{-c/2\kappa+\varepsilon})$, then

$$\begin{aligned} f_\kappa(\tau) &= \frac{c}{2\kappa} - 2c + 2c\kappa + \log(e^{-c/2\kappa+\varepsilon}) + \log \frac{1}{1-2\tau} + \frac{2\tau(1-2\kappa)}{1-2\tau} \\ &= \varepsilon - 2c + 2c\kappa + \log \frac{1}{1-2\tau} + \frac{2\tau(1-2\kappa)}{1-2\tau}, \end{aligned}$$

where the last two terms on the RHS are positive.

Referring to (9) we see that $f_\kappa(0) = c(1 - 2\kappa)^2/2\kappa > 0$ for any $\kappa < 1/2$. If $\varepsilon = 0$ then assuming $\tau < \kappa < 1/2c$ and $c > c_0$ then $f_\kappa(\tau) < 0$ so $\tau > \tau_0$, since f_κ is monotone increasing in τ . On the other hand, if $\varepsilon \geq 2c$ then $f_\kappa(t) > 0$ so $\tau < \tau_0$. \square

3 MODIFIED GREEDY

Here we choose a random vertex x and then a random neighbor y , add $\{x, y\}$, to the matching M and delete the vertices of the edge from $V(t)$ and all edges of the same color as the edge. If $d_t(x) = 0$ then we just delete x . Let $\nu(t) = |V(t)|$. We let $q(t)$ denote the number of unused colors. Let t_0 be the number of steps when a single vertex x of degree zero is deleted, and t_1 the number of steps when a pair of vertices x, y are deleted. As $t = t_0 + t_1$, $\nu(t) = n - t_0 - 2t_1$, and $q(t) = q - t_1$ it follows that

$$q(t) = t + \nu(t) + q - n. \quad (13)$$

If we condition on $\mu(t), \nu(t)$, the remaining unrevealed random graph is uniform over all graphs with $\mu(t)$ edges and $\nu(t)$ vertices. Let

$$\lambda = \lambda(t) := \frac{2\mu(t)}{\nu(t)}$$

be the average degree of the unrevealed graph. Recall (see, for example, [4]) that if $b = o(a^{2/3})$ then

$$\binom{a}{b} \sim \frac{a^b}{b!} \exp\left\{-\frac{b^2}{2a}\right\}.$$

Thus if μ, ν are at least $n/\log \log n$ (and observing they are at most $O(n)$) then the probability the first chosen vertex has degree k for $k \leq \log^2 n$ is

$$\begin{aligned} \frac{\binom{\nu-1}{k} \binom{\binom{\nu-1}{2}}{\mu-k}}{\binom{\binom{\nu}{2}}{\mu}} &= \frac{(\nu-1)^k}{k!} \frac{\binom{\nu-1}{2}^{\mu-k}}{(\mu-k)!} \frac{\mu!}{\binom{\nu}{2}^\mu} \exp\left\{-\frac{k^2}{2(\nu-1)} - \frac{(\mu-k)^2}{2\binom{\nu-1}{2}} + \frac{\mu^2}{2\binom{\nu}{2}}\right\} \\ &= \binom{\mu}{k} \left(\frac{\nu-1}{\binom{\nu}{2}}\right)^k \left(\frac{\binom{\nu-1}{2}}{\binom{\nu}{2}}\right)^\mu \exp\left\{\tilde{O}\left(\frac{1}{n}\right) + \frac{\mu^2}{\nu(\nu-1)} \left(1 - \frac{(1-k/\mu)^2}{1-2/\nu}\right)\right\} \\ &= \frac{\mu^k}{k!} \cdot \left(\frac{2}{\nu}\right)^k \cdot \left(1 - \frac{2}{\nu}\right)^\mu \exp\left\{\tilde{O}\left(\frac{1}{n} + \frac{k\mu}{\nu^2} + \frac{\mu^2}{\nu^3}\right)\right\} \\ &= \frac{\lambda^k}{k!} \exp\left\{\left(-\frac{2}{\nu} + O\left(\frac{1}{\nu^2}\right)\right)\mu + \tilde{O}\left(\frac{1}{n}\right)\right\} \\ &= \frac{\lambda^k}{k!} e^{-\lambda} + \tilde{O}\left(\frac{1}{n}\right). \end{aligned} \quad (14)$$

Conditional on the first chosen vertex having degree k for $1 \leq k \leq \log^2 n$, the expected number of additional neighbors of the second vertex is

$$\frac{2(\mu - k)}{\nu - 1} = \lambda + \tilde{O}\left(\frac{1}{n}\right). \quad (15)$$

Conditional on a total of $\leq 2 \log^2 n$ edges being adjacent to the two chosen vertices, the expected number of edges deleted due to being the same color as the matching edge is

$$\frac{\mu - O(\log^2 n)}{q} = \frac{\mu}{q} + \tilde{O}\left(\frac{1}{n}\right) \quad (16)$$

where the big-O term on the last line follows since (as we will see later) we have $q(t) = \Omega(n)$.

Let $\xi(t) = (\mu(t), \nu(t), q(t))$. Then (explanation follows)

$$\mathbb{E}(\nu(t+1) \mid \xi(t)) = \nu(t) - 2 + e^{-\lambda} + \tilde{O}\left(\frac{1}{n}\right) \quad (17)$$

$$\mathbb{E}(\mu(t+1) \mid \xi(t)) = \mu(t) - \lambda - (1 - e^{-\lambda}) \left(\lambda + \frac{\mu(t)}{q(t)} \right) + \tilde{O}\left(\frac{1}{n}\right). \quad (18)$$

Indeed, for the first line note that we lose 2 vertices unless the first vertex has degree 0 (which happens with probability about $e^{-\lambda}$ by (14)) in which case we lose only 1. For the second line, note that we expect to lose λ edges adjacent to the first vertex. Then, if the first vertex has positive degree (probability about $1 - e^{-\lambda}$ by (14)), we expect to lose about λ additional edges adjacent to the second vertex by (15), as well as about $\frac{\mu(t)}{q(t)}$ edges of the same color as the matching edge by (16). The big-O term comes from the approximations (14), (15) and (16).

Substituting $\nu/n = N$, $\mu/n = M$ and using (13) with $\kappa = q(0)/n$, $\tau = t/n$, this leads to the equations

$$\frac{dM}{d\tau} = \lambda(e^{-\lambda} - 2) - \frac{M(1 - e^{-\lambda})}{N + \tau + \kappa - 1}. \quad (19)$$

$$\frac{dN}{d\tau} = e^{-\lambda} - 2, \quad (20)$$

where $\lambda = 2M/N$.

If we substitute (21) into (19) it gives the following.

$$\begin{aligned} M' &= \frac{2M}{N} N' + M \frac{N' + 1}{N + \tau + \kappa - 1} \\ \frac{M'}{M} &= 2 \frac{N'}{N} + \frac{(N + \tau + \kappa - 1)'}{N + \tau + \kappa - 1} \quad \text{on division by } M \\ \implies \log M &= 2 \log N + \log(N + \tau + \kappa - 1) + \log B \\ M &= B N^2 (N + \tau + \kappa - 1). \end{aligned}$$

Assuming $M(0) = c/2$ and $N(0) = 1$ this gives

$$M = \frac{c}{2\kappa}N^2(N + \tau + \kappa - 1), \quad \lambda = \frac{c}{\kappa}N(N + \tau + \kappa - 1).$$

Interestingly, the above expression for M is intuitive in a sense. In particular, it essentially says that the number of edges is approximately

$$\mu(t) \approx \binom{\nu(t)}{2} \cdot \frac{c}{n} \cdot \frac{q(t)}{q},$$

i.e. the edge density in the remaining graph is about the original edge density c/n times the probability that a given color is unused, $q(t)/q$.

Substituting this expression for λ into (21) gives

$$N' = e^{-\frac{c}{\kappa}N(N+\tau+\kappa-1)} - 2, \quad N(0) = 1. \quad (21)$$

Let τ_0 be the smallest positive solution to $M(\tau) = 0$, so either $N(\tau_0) = 0$ or $N(\tau_0) + \tau_0 + \kappa - 1 = 0$. We will show that $N(\tau_0) = 0$, i.e. we run out of vertices before we run out of colors. This should be unsurprising since we ought to never run out of colors (e.g. a positive proportion of colors simply never appear).

Indeed, letting $Q(\tau) = N(\tau) + \tau + \kappa - 1$, we have

$$Q' = N' + 1 = e^{-\frac{c}{\kappa}NQ} - 1 \geq -\frac{c}{\kappa}NQ$$

and $Q(0) = \kappa > 0$ and so

$$Q(\tau) \geq \kappa e^{-\frac{c}{\kappa} \int_0^\tau N(x) dx} > 0.$$

Therefore $N(\tau_0) = 0$ and $Q(\tau_0) > 0$.

We will apply Theorem 3 again. This time $a = 2$ and our two random variables are $\nu(t)$ and $\mu(t)$. We let

$$\mathcal{D} = \left\{ (\tau, N, M) : 0 \leq \tau \leq \frac{1}{2}, \frac{1}{\log \log n} \leq N \leq 2, \frac{1}{\log \log n} \leq M \leq c \right\}.$$

Now we make sure to satisfy condition (i). By (17) and (18) we can let $\delta = \tilde{O}(n^{-1})$. The functions on the right hand sides of (19) and (21) are L -Lipschitz on \mathcal{D} for $L = O(\log \log n)$. We now move on to condition (ii). We have to take $\theta = O(n)$ since it is possible (though very unlikely) to have a vertex of linear degree. We let $\phi = 2n^{0.1}$, and then the same calculation from (5) indicates we can choose $\gamma = \exp\{-\Omega(n^{0.1})\}$. Now for condition (iii), any positive ε will do (we will have to choose ε more carefully later to satisfy future conditions). We choose $T = 1$, $R = O(\log \log n)$, $\varepsilon = n^{-0.1}$. Thus with high probability we have

$$\mu(t) = M(\tau)n + \tilde{O}(n^{0.9}), \quad \nu(t) = N(\tau)n + \tilde{O}(n^{0.9}) \quad (22)$$

uniformly for $0 \leq t \leq \tau_0 n + \tilde{O}(n^{0.9})$. This implies that w.h.p. our process ends when it runs out of vertices, which happens after $\tau_0 n + \tilde{O}(n^{0.9})$ steps. Now observe that at any step t the number of remaining vertices $\nu(t)$ is equal to $n - t - x$ where x is the number of edges in the matching. So when the process terminates we have $0 = n - \tau_0 n + \tilde{O}(n^{0.9}) - x$ and so the number of edges in the final matching μ is

$$\mu = (1 - \tau_0)n + \tilde{O}(n^{0.9}). \quad (23)$$

3.1 Upper bound on the value of μ

Integrating (21) subject to $N(0) = 1$, we have

$$N(\tau) = 1 - 2\tau + \int_0^\tau e^{-\frac{c}{\kappa}N(N+\sigma+k-1)} d\sigma. \quad (24)$$

Let τ_0 be the smallest positive solution to $N(\tau) = 0$. Assume the result of the previous section that the algorithm terminates when this condition is met (asymptotically).

Lemma 7. *Assuming $\kappa \geq 1/2$, the function $F(\tau) = N(N + \tau + \kappa - 1)$ is convex in $[0, \tau_0]$.*

Proof. We first prove that $N''(\tau) > 0$ for $\tau \in [0, \tau_0)$. From (21) it is clear that $N' + 1 < 0$ in this interval. Thus

$$N'' = e^{-\frac{c}{\kappa}N(N+\tau+\kappa-1)} - \frac{c}{\kappa} \cdot [N(N + \tau + \kappa - 1)]',$$

where

$$[N(N + \tau + \kappa - 1)]' = N'(N + \tau + \kappa - 1) + N(N' + 1).$$

We show below that this derivative is negative for $\kappa \geq 1/2$. The result that $N'' > 0$ follows from this. As $N > 0$ and $N' + 1 < 0$ we only need to show that $N + \tau + \kappa - 1 \geq 0$.

(i) As $\nu(t) = n - 2t_1 - t_0 = n - 2t + t_0$, $\nu(t) \geq n - 2t$. So if $\tau \leq 1/2$ and $\kappa \geq 1/2$,

$$N(\tau) + \tau + \kappa - 1 \geq 1 - 2\tau + \tau + \kappa - 1 = \kappa - \tau \geq 0.$$

(ii) Also if $1/2 \leq \tau < \tau_0$, then as $N(\tau) > 0$,

$$N + \tau + \kappa - 1 \geq \tau + \kappa - 1 \geq \tau - \frac{1}{2} \geq 0.$$

Thus if

$$\begin{aligned} F &= N(N + \tau + \kappa - 1), \\ F' &= N'(N + \tau + \kappa - 1) + N(N' + 1) \leq 0, \\ F'' &= N''(N + \tau + \kappa - 1) + 2N'(N' + 1) + NN'' \geq 0, \end{aligned} \quad (25)$$

as all the terms on the RHS of (25) are non-negative. □

Noting that F is convex in $[0, \tau_0]$ and that $F(0) = \kappa$, $F(\tau_0) = 0$, let $C(\tau) = \kappa - \kappa\tau/\tau_0$ be the chordal line of F such that $F(\tau) \leq C(\tau)$ in that interval. This implies that

$$e^{-\frac{c}{\kappa}F(\tau)} \geq e^{-\frac{c}{\kappa}C(\tau)} = e^{-c+c\tau/\tau_0}.$$

Inserting this into (24) we have

$$N(\tau) \geq 1 - 2\tau + \int_0^\tau e^{-c+c\sigma/\tau_0} d\sigma.$$

Put $\tau = \tau_0$ so that $N(\tau) = 0$ to obtain

$$0 \geq 1 - 2\tau_0 + \frac{\tau_0}{c}(1 - e^{-c}) \quad \implies \quad \tau_0 \geq \frac{1}{2 - \frac{1}{c}(1 - e^{-c})}.$$

From (23) it follows asymptotically that

$$\mu \leq \frac{1 - \frac{1}{c}(1 - e^{-c})}{2 - \frac{1}{c}(1 - e^{-c})} n.$$

4 Final thoughts

We have made progress in understanding the performance of simple greedy algorithms for finding large rainbow matchings in sparse random graphs. Our result is precise for GREEDY in the case $\kappa = 1/2$. This is not an easy question, given that it is related to finding large matchings in sparse random 3-uniform hypergraphs. Here an edge $\{u, v\}$ of color c can be thought of as a triple $\{u, v, c\}$. Of course, the decision problem of whether a 3-uniform hypergraph has a matching of size k is on Karp's famous list of 21 NP-complete problems [6].

We have reduced the analysis of algorithms GREEDY and MODIFIED GREEDY to the analysis of some differential equations. While these differential equations are not terribly complex, in general they lack an explicit elementary solution. We hope that further study will lead to a better understanding. We briefly looked at a colored version of the Karp-Sipser algorithm and constructed the appropriate differential equations. We will not pursue this line here, but instead wait until we can better understand these equations.

It would be nice to prove the conjecture, mentioned in Section 1, that MODIFIED GREEDY performs better than GREEDY. In the color-free setting, Dyer, Frieze and Pittel [5] proved that the Modified Greedy Matching algorithm performs better than Greedy Matching on a random graph. However the rainbow version is complicated by the lack of explicit elementary solutions to the differential equations.

References

- [1] J. Aronson, A.M. Frieze and B. Pittel, Maximum matchings in sparse random graphs: Karp-Sipser revisited, *Random Structures and Algorithms* 12 (1998) 111-178.
- [2] V. Bang Le and F. Pfender, Complexity results for rainbow matchings, *Theoretical Computer Science* 524 (2014) 27-33,
- [3] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled graphs, *European Journal on Combinatorics* 1 (1980) 311-316

- [4] A. Frieze and M. Karoński, Introduction to random graphs, *Cambridge University Press* (2016)
- [5] M. Dyer, A.M. Frieze and B. Pittel, The average performance of the greedy matching algorithm, *Annals of Applied Probability* 3 (1993) 526-552.
- [6] R. M. Karp. Reducibility among Combinatorial Problems, in: Complexity of Computer Computations. Miller, R.E., Thatcher, J.W., Bohlinger, J.D. (eds). The IBM Research Symposia Series. Springer (1972).
- [7] R. M. Karp and M. Sipser. Maximum matchings in sparse random graphs, *Proceedings of the 22nd IEEE Annual Symposium on Foundations of Computer Science* (1981) 364–375.
- [8] L. Warnke, On Wormald’s differential equation method, *Combinatorics, Probability and Computing*, to appear.
- [9] N. C. Wormald, The differential equation method for random graph processes and greedy algorithms, Lectures on approximation and randomized algorithms (M. Karoński and H.J. Prömel, eds.), Advanced Topics in Mathematics, Polish Scientific Publishers PWN, Warsaw, 1999, pp. 73-155.