RANDOMIZED GREEDY MATCHING

Martin Dyer*
School of Computer Studies, University of Leeds,
Leeds, U.K.

and

Alan Frieze†
Department of Mathematics, Carnegie-Mellon University,
Pittsburgh, U.S.A.

April 14, 2005

Abstract

We consider a randomized version of the greedy algorithm for finding a large matching in a graph. We assume that the next edge is always randomly chosen from those remaining. We analyse the performance of this algorithm when the input graph is fixed. We show that there are graphs for which this Randomized Greedy Algorithm (RGA) usually only obtains a matching close in size to that guaranteed by worst-case analysis (*i.e.* half the size of the maximum). For some classes of sparse graphs (*e.g.* planar graphs and forests) we show that the RGA performs significantly better than the worst-case. Our main theorem concerns forests. We prove that the ratio to maximum here is at least $0.7690\cdots$, and that this bound is tight.

*Supported by NATO grant RG0088/89
†Supported by NSF grant CCR-8900112 and NATO grant RG0088/89

1 Introduction

Perhaps the simplest heuristic for finding a large cardinality matching in a graph $G = (V, E)$ is the "Greedy Heuristic".

GREEDY MATCHING

> **begin**
> $M \leftarrow \emptyset$;
>   **while** $E(G) \neq \emptyset$ **do**
>   **begin**
> $\boldsymbol{A}$: Choose $e = \{u, v\} \in E$
> $G \leftarrow G \setminus \{u, v\}$;
> $M \leftarrow M \cup \{e\}$
>   **end**;
> Output M
> **end**

The choice of e in statement $\boldsymbol{A}$ is unspecified. It is known [3] that, if the worst possible choices are made in $\boldsymbol{A}$, the size of the matching M produced is at least one half of the size of the largest matching, and one half is attainable. (Consider choosing the middle edge of a path of length three.)

Now randomization sometimes improves the performance of algorithms (perhaps the most important example being primality testing). The question we pose here is what effect does randomizing statement $\boldsymbol{A}$ have ? In particular if e is chosen uniformly at random from the remaining edges, what is the expected ratio of the size of M to that of the maximum matching ? We prove that there are graphs for which the average-case is hardly better than the worst-case, but also that there are classes of graphs (*e.g.* planar graphs) for which it is significantly better.

2 Notation

Let $G = (V, E)$ be a (simple) graph with $|V| = n$. For any $v \in V$, $\Gamma(v)$ denotes its neighbours in G. For any $S \subseteq V$, $G \setminus S$ denotes the subgraph induced by the vertex set $V \setminus S$. Let $m(G)$ be the maximum size of a matching in G and $\mu(G)$ be the expected size of the randomized greedy matching. Let

$$\begin{aligned} r(G) &= \mu(G)/m(G) \quad \text{if } m(G) > 0 \\ &= 1 \qquad\qquad\quad\;\; \text{if } m(G) = 0. \end{aligned}$$

If $\mathcal{K}$ is any class of graphs $\rho(\mathcal{K}) = \inf_{G \in \mathcal{K}} r(G)$. Unless otherwise stated, $\mathcal{G}$ will denote any class of graphs closed under vertex deletions and (to avoid trivialities) we suppose $|E| > 0$ for some $G \in \mathcal{G}$.

$$\kappa(\mathcal{G}) = \inf_{G \in \mathcal{G}} \{|V|/|E| : G = (V, E),\ |E| > 0\}$$

Note that since some $G \in \mathcal{G}$ has an edge, and $\mathcal{G}$ is closed under deletions, the graph containing a single edge lies in $\mathcal{G}$. Thus $0 \le \kappa(\mathcal{G}) \le 2$ for any $\mathcal{G}$. In particular $\kappa(\text{GRAPHS}) = 0$, $\kappa(\text{PLANAR GRAPHS}) = \frac{1}{3}$, $\kappa(\text{FORESTS}) = 1$. The abbreviation *RGA* is used for "Randomized Greedy Algorithm".

3 A monotonicity property

Many of our results depend on the following

Lemma 1 *For all* $v \in V$, $\mu(G) \ge \mu(G \setminus \{v\}) \ge \mu(G) - 1$

Proof The statement clearly holds for $G = (\{v\}, \emptyset)$ and we argue by induction on $|V|$. Let $H = G \setminus \{v\}, A = \{xy \in E : v \notin \{x, y\}\}$ and $d = |\Gamma(v)|$. Then

$$\begin{aligned}
\mu(G) &= \frac{1}{|E|} \sum_{xy \in E} (1 + \mu(G \setminus \{x, y\}) \\
&= \frac{1}{|E|} \sum_{xy \in A} (1 + \mu(G \setminus \{x, y\}) + \frac{1}{|E|} \sum_{y \in \Gamma(v)} (1 + \mu(H \setminus \{y\})) \\
&\geq \frac{1}{|E|} \sum_{xy \in A} (1 + \mu(H \setminus \{x, y\}) + \frac{1}{|E|} \sum_{y \in \Gamma(v)} (1 + \mu(H \setminus \{y\})) \\
&\qquad \text{by induction} \\
&= \left(1 - \frac{d}{|E|}\right) \mu(H) + \frac{1}{|E|} \sum_{y \in \Gamma(v)} (1 + \mu(H \setminus \{y\})) \\
&= \mu(H) + \frac{1}{|E|} \sum_{y \in \Gamma(v)} (1 + \mu(H \setminus \{y\}) - \mu(H)) \\
&\geq \mu(H) \qquad \text{by induction.}
\end{aligned}$$

Conversely,

$$\begin{aligned}
\mu(G) &= \frac{1}{|E|} \sum_{xy \in A} (1 + \mu(G \setminus \{x, y\}) + \frac{1}{|E|} \sum_{y \in \Gamma(v)} (1 + \mu(H \setminus \{y\})) \\
&= 1 + \frac{1}{|E|} \sum_{xy \in A} \mu(G \setminus \{x, y\}) + \frac{1}{|E|} \sum_{y \in \Gamma(v)} \mu(H \setminus \{y\}) \\
&\leq 1 + \frac{1}{|E|} \sum_{xy \in A} (1 + \mu(H \setminus \{x, y\}) + \frac{d}{|E|} \mu(H) \\
&\qquad \text{by induction.} \\
&= 1 + \left(1 - \frac{d}{|E|}\right) \mu(H) + \frac{d}{|E|} \mu(H) \\
&= 1 + \mu(H).
\end{aligned}$$

□

Corollary 1 *Let $v \in V$ be exposed in some maximum matching of G, then*

$$r(G \setminus \{v\}) \leq r(G).$$

Proof Clearly $m(G \setminus \{v\}) = m(G)$, so the result follows from Lemma 1. □

Corollary 2 *Let $\mathcal{H} \subseteq \mathcal{G}$ be the set of $G \in \mathcal{G}$ which are connected and contain a perfect matching. Then $\rho(\mathcal{G}) = \rho(\mathcal{H})$.*

Proof Clearly $\rho(\mathcal{H}) \geq \rho(\mathcal{G})$. By Corollary 1 (applied repeatedly if necessary), any $G \in \mathcal{G}$ can be reduced to a G' which contains a perfect matching and has $r(G') \leq r(G)$. If G' has components G'_i $(i = 1, \ldots, c)$, let $H = G'_j$ where $r(G'_j) = \min_{1 \leq i \leq c} r(G'_i)$. Clearly $H \in \mathcal{H}$ and $r(H) \leq r(G') \leq r(G)$. Thus $\rho(H) \leq \rho(G)$. □

In particular we have the following, which we use below,

$$\rho(\text{FORESTS}) = \rho(\text{TREES WITH A PERFECT MATCHING}).$$

We note in passing that monotonicity under *edge* deletions does not hold. As a simple example, let G be a path of three edges. Then $\mu = \frac{5}{3}$, but, when the middle edge is deleted, $\mu = 2$.

4 A lower bound

We give a weak, but easily proved, lower bound and examine its consequences.

Theorem 1 *Let $\alpha(\mathcal{G}) = 1/(2 - \frac{1}{2}\kappa(\mathcal{G}))$. Then $\rho(\mathcal{G}) \geq \alpha(\mathcal{G})$.*

Proof By induction on $|V|$. Since $0 \leq \kappa(\mathcal{G}) \leq 2$ we have $\frac{1}{2} \leq \alpha(\mathcal{G}) \leq 1$. If $|V| = 0$, $r(G) = 1$ and hence $r(G) \geq \alpha(\mathcal{G})$.

Since (by Corollary 2) we may assume G has a perfect matching we take $|V| = 2m(G) > 0$. Now

$$\mu(G) = 1 + \frac{1}{|E|} \sum_{uv \in E} \mu(G \setminus \{u, v\}).$$

However

$$\begin{aligned} m(G \setminus \{u, v\}) &= m(G) - 1 && \text{if } uv \text{ lies in some perfect matching,} \\ &= m(G) - 2 && \text{otherwise.} \end{aligned}$$

Thus, where $m = m(G)$,

$$\begin{aligned} \sum_{uv \in E} m(G \setminus \{u, v\}) &\geq m(m-1) + (|E| - m)(m-2) \\ &= |E|(m-2) + m. \end{aligned}$$

Hence, using the inductive hypothesis,

$$\begin{aligned} \mu(G) &\geq 1 + \frac{1}{|E|} \sum_{uv \in E} \alpha m(G \setminus \{u, v\}) \\ &\geq 1 + \frac{1}{|E|} \alpha(|E|(m-2) + m) \\ &= \alpha m + 1 - 2\alpha + \frac{\alpha}{2}|V|/|E| \\ &\geq \alpha m + 1 - 2\alpha + \tfrac{1}{2}\alpha\kappa \\ &= \alpha m, \end{aligned}$$

completing the induction. □

Corollary 3 *ρ(GRAPHS) $\geq \frac{1}{2}$, ρ(PLANAR GRAPHS) $\geq \frac{6}{11}$, ρ(FORESTS) $\geq \frac{2}{3}$, ρ(Δ-GRAPHS) $\geq \Delta/(2\Delta - 1)$, where Δ-GRAPHS stands for graphs with maximum degree at most Δ. So, in particular, ρ(CUBIC GRAPHS) $\geq \frac{3}{5}$.* □

5 The class GRAPHS

Theorem 2 ρ(GRAPHS) $= \frac{1}{2}$.

Proof Let G_m be the graph obtained by adding a new vertex and edge adjacent to each vertex of the complete graph K_m.

Clearly $m(G_m) = m$, and write $\mu(G_m) = \mu_m$. Consider the first step of the *RGA* on G_m. There are $\binom{m}{2} + m$ edges. Thus, with probability

$$\frac{m}{\binom{m}{2} + m} = \frac{2}{m+1},$$

we choose an added edge. Its removal leaves G_{m-1}. Otherwise we choose a K_m edge whose removal leaves G_{m-2} (and two isolated vertices). Thus the final matching size will be, in expectation,

$$1 + \mu_{m-1} \quad \text{with probability} \quad \frac{2}{m+1}$$

$$\text{and} \quad 1 + \mu_{m-2} \quad \text{with probability} \quad \frac{m-1}{m+1}.$$

Thus,

$$\mu_m = 1 + \frac{2\mu_{m-1} + (m-1)\mu_{m-2}}{m+1} \qquad (m \geq 2)$$

with $\mu_0 = 0$, $\mu_1 = 1$. Writing this as

$$(\mu_m - \mu_{m-1}) = 1 - \frac{(m-1)}{(m+1)}(\mu_{m-1} - \mu_{m-2}), \tag{1}$$

we make the substitution $u_m = \mu_m - \mu_{m-1}$ and $u_0 = \mu_0$. Thus $u_0 = 0$, $u_1 = 1$, and $\mu_m = \sum_{j=0}^{m} u_j$, and from (1),

$$u_m = 1 - \frac{(m-1)}{(m+1)} u_{m-1} \tag{2}$$

It is easy to show inductively that (2) has solution, for $m \geq 1$:

$$\begin{aligned} u_m &= \tfrac{1}{2} + \tfrac{1}{2m} & (m \text{ odd}) \\ &= \tfrac{1}{2} + \tfrac{1}{2(m+1)} & (m \text{ even}) \end{aligned}$$

Let $L_m = 1 + \frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{m}$ for m odd. Thus,

$$\begin{aligned} \mu_m = \textstyle\sum_{j=0}^{m} u_j &= \tfrac{1}{2}m - \tfrac{1}{2} + L_m & (m \text{ odd}) \\ &= \tfrac{1}{2}m - \tfrac{1}{2} + L_{m-1} + \tfrac{1}{2(m+1)} & (m \text{ even}) \end{aligned}$$

Asymptotically $L_m = \frac{1}{2}(\gamma + \log 2m\,)$, where γ is Euler's constant. So

$$\mu_m = \tfrac{1}{2}(m + \log 2m + \gamma - 1) + o(1).$$

Thus $r(G_m) = \frac{1}{2} + O(\log m/m)$ and $r(G_m) \to \frac{1}{2}$ as $m \to \infty$. □

6 Concentration near the mean

We now show that the value of the matching obtained by the *RGA* is "almost always" near its expectation.

Theorem 3 *Let G be a graph with $m = m(G)$, $\mu = \mu(G)$ and let $X = X(G)$ be the random size of the matching obtained by the RGA in G. Then*

$$\Pr(\,|X - \mu| > \epsilon m\,) \leq 2e^{-2\epsilon^2 m}$$

Proof Let Y_i, $(i = 0, 1, \ldots, m)$ be the Doob martingale induced by the first i choices of the RGA on G, *i.e.* $Y_i = \mathbf{E}(X \mid \text{first } i \text{ choices})$. Clearly $Y_i = K + \mu(H)$ for some integer $K \leq i$ and subgraph H of G. In fact $K = i$ unless $H = \emptyset$. Also

$$\begin{aligned} Y_{i+1} &= K + 1 + \mathbf{E}(\mu(H \setminus \{u, v\}) & \text{if } H \text{ contains an edge,} \\ &= K & \text{otherwise,} \end{aligned}$$

where the expectation is over the random choices of the edge uv. Thus,

$$\begin{aligned} Y_{i+1} - Y_i &= 1 + \mathbf{E}(\mu(H \setminus \{u, v\}) - \mu(H)) & \text{if } H \text{ contains an edge,} \\ &= 0 & \text{otherwise.} \end{aligned}$$

Thus if H contains an edge,

$$\begin{aligned} Y_{i+1} - Y_i &= \mathbf{E}(1 + \mu(H \setminus \{u, v\}) - \mu(H)) \\ &\leq 1, \qquad \text{since } \mu(H \setminus \{u, v\}) \leq \mu(H) \end{aligned}$$

Furthermore,

$$\begin{aligned} Y_{i+1} - Y_i &\geq \mathbf{E}(\mu(H \setminus \{u\}) - \mu(H)), \\ & \qquad\qquad \text{since } \mu(H \setminus \{u, v\}) \geq \mu(H \setminus \{u\}) - 1 \\ &\geq -1, \qquad\qquad \text{since } \mu(H \setminus \{u\}) \geq \mu(H) - 1, \end{aligned}$$

where all inequalities follow from Lemma 1.

Thus $|Y_{i+1} - Y_i| \leq 1$ whether or not H has an edge. Hence $\{Y_i\}$ is a bounded difference martingale sequence, and it follows from the Hoeffding-Azuma inequality (see Bollobás [1], McDiarmid [4]) that

$$\Pr(|X - \mu| > \epsilon m) \leq 2e^{-2(\epsilon m)^2/m} = 2e^{-2\epsilon^2 m}$$

□

Corollary 4 *If $\{G_m\}$ is a graph sequence such that $m(G_m)(= m) \to \infty$, and $\omega_m \to \infty$ (arbitrarily slowly), then*

$$\Pr(\mu(G_m) - \omega_m\sqrt{m} \leq X(G_m) \leq \mu(G_m) + \omega_m\sqrt{m}) \to 1$$

Proof Put $\epsilon = \omega_m/\sqrt{m}$ in Theorem 3. □

Corollary 5 *If $\{G_m\}$ is the graph sequence defined in the proof of Theorem 2, let $\hat{X}(G_m)$ be the best solution obtained from any polynomial number $p(m)$ of repetitions of the* RGA *on G_m. Then*

$$\Pr(\tfrac{1}{2}m \leq \hat{X}(G_m) \leq \tfrac{1}{2}m + \log m/\sqrt{m}) \to 1 \text{ as } m \to \infty.$$

Proof $\hat{X}(G_m) \geq \frac{1}{2}m$ follows from the worst-case result (Korte and Haussman [3]). Putting $\epsilon = \log m/\sqrt{m}$ in Lemma 3, the probability of $\hat{X}(G_m)$ not falling in the required interval is at most $2p(m)e^{-2(\log m)^2} \to 0$ as $m \to \infty$. □

7 A monotone transformation

Deletion of exposed vertices does not increase $r(G)$. We consider another transformation with this property. Let $\{u, v\}$ be an edge in a maximum matching of G which does lie in any triangle. Let G' be the graph obtained by substituting all edges vw ($w \in \Gamma(v) \setminus \{u\}$) with uw.

Note the restriction that uv does not lie in a triangle ensures that G' is a simple graph.

Lemma 2 $r(G') \leq r(G)$.

Proof Clearly $m(G') = m(G)$ so we only need show $\mu(G') \leq \mu(G)$. We proceed by induction on the number of vertices of G. The base case is when G is a single edge. Then $G' = G$ and the result is trivial. For the induction, let $A = \{xy : \{x, y\} \cap \{u, v\} = \emptyset\}$ and for $w = u, v$ let $\hat{\Gamma}(w) = \Gamma(w) \setminus \{u, v\}$. Note that $\hat{\Gamma}(u)$, $\hat{\Gamma}(v)$ are fixed sets here, defined in G and unaltered in G'. Moreover, they are disjoint by assumption. Thus,

$$\begin{aligned}
\mu(G) &= 1 + \frac{1}{|E|}(\sum_{xy \in A} \mu(G \setminus \{x, y\}) + \mu(G \setminus \{u, v\}) \\
&\quad + \sum_{x \in \hat{\Gamma}(u)} \mu(G \setminus \{u, x\}) + \sum_{x \in \hat{\Gamma}(v)} \mu(G \setminus \{v, x\}) \\
&\geq 1 + \frac{1}{|E|}(\sum_{xy \in A} \mu(G' \setminus \{x, y\}) + \mu(G' \setminus \{u, v\}) \\
&\quad + \sum_{x \in \hat{\Gamma}(u)} \mu(G' \setminus \{u, x\}) + \sum_{x \in \hat{\Gamma}(v)} \mu(G' \setminus \{u, x\}). \\
&= \mu(G').
\end{aligned}$$

where we have used the following:

(i) $(G \setminus \{x, y\})' = G' \setminus \{x, y\}$ if $xy \in A$ and so (by induction) $\mu(G \setminus \{x, y\}) \geq \mu((G \setminus \{x, y\})') = \mu(G' \setminus \{x, y\})$;

(ii) $G \setminus \{u, v\} = G' \setminus \{u, v\}$;

(iii) $G' \setminus \{u, x\} = G \setminus \{u, v, x\}$ for $x \in \hat{\Gamma}(u) \cup \hat{\Gamma}(v)$ (after removing the isolated vertex v in G'). So $\mu(G' \setminus \{u, x\}) \leq \mu(G \setminus \{u, x\})$ and $\mu(G' \setminus \{u, x\}) \leq \mu(G \setminus \{v, x\}$ follow from Lemma 1 for these values of x. □

Let us denote this transformation by σ : GRAPHS $\rightarrow$ GRAPHS, *i.e.* $G' = \sigma(G)$. Let $\mathcal{G}^*$ be any graph-family which is also closed under σ. Let $\mathcal{H}^*$ be the sub-family of $\mathcal{G}^*$ such that any $G \in \mathcal{H}^*$ is connected, has a perfect matching, and such that every edge in any perfect matching either contains a vertex of degree 1 or lies in a triangle. Then,

Corollary 6 $\rho(\mathcal{H}^*) = \rho(\mathcal{G}^*)$. □

This Corollary is useful for FORESTS, since it implies we may assume

(i) The graph is a tree.

(ii) All edges in the maximum matching are leaves.

(iii) Every internal vertex is adjacent to exactly one leaf.

To see this let T be a tree for which (ii) or (iii) does not hold. If vertex v is adjacent to leaves $w_1, w_2, ..., w_d$, $d \geq 2$ then deleting $w_2, ..., w_d$ yields a tree T' for which $m(T') = m(T)$ and $\mu(T') \leq \mu(T)$ (Lemma 1.) If v is a vertex not adjacent to any leaf then we can assume that it lies on an edge uv of some perfect matching. We can then apply σ and if necessary reduce the number of leaf neighbours of u in $\sigma(T)$ to one. After a finite number of iterations of the above procedure we satisfy (i),(ii),(iii) without increasing r.

A tree satisfying (ii) and (iii) will be called an *L-tree.*

8 The class FORESTS

For FORESTS, Corollary 3 gives $\rho \geq \frac{2}{3}$, but this is not tight. In this section we prove

Theorem 4 $\rho(\text{FORESTS}) = \alpha = \frac{2}{3} + 2\sum_{k=0}^{\infty} \frac{(-2)^k}{(2k+5)!!} = 0.7690397\cdots$
(where $n!! = n(n-2)(n-4)\cdots 3 \cdot 1$ for n odd). □

We first establish the upper bound.

Lemma 3 $\rho(\text{FORESTS}) \leq \alpha.$

Proof Let T_m be the graph obtained by adding a leaf to each vertex of an m-vertex path. Let $t_m = \mu(T_m)$. Thus $t_0 = 0$, $t_1 = 1$. Clearly, for $m \geq 2$,

$$\begin{aligned} t_m &= 1 + \frac{1}{2m-1}\left(\sum_{i=1}^{m}(t_{i-1} + t_{m-i}) \;+\; \sum_{i=1}^{m-1}(t_{i-1} + t_{m-i-1})\right) \\ &= 1 + \frac{2}{2m-1}\left(\sum_{i=0}^{m-1} t_i + \sum_{i=0}^{m-2} t_i\right) \end{aligned} \tag{3}$$

From (3), for $m \geq 3$, $(2m-1)t_m = (2m-1) + 2(\sum_{i=0}^{m-1} t_i + \sum_{i=0}^{m-2} t_i)$,

and also $(2m-3)t_{m-1} = (2m-3) + 2(\sum_{i=0}^{m-2} t_i + \sum_{i=0}^{m-3} t_i)$.

Subtracting, $(2m-1)t_m - (2m-3)t_{m-1} = 2 + 2t_{m-1} + 2t_{m-2}$,

or,

$$(2m-1)(t_m - t_{m-1}) = 2(1 + t_{m-2}). \tag{4}$$

In fact, (4) holds also for $m = 2$ since $t_2 = \frac{5}{3}$ from (3).

Let $u_m = t_m - t_{m-1}$, $u_0 = t_0$, so $t_m = \sum_{i=0}^{m} u_i$, and $u_0 = 0$, $u_1 = 1$, $u_2 = \frac{2}{3}$.

So, from (4), $(2m-1)u_m = 2(1 + \sum_{i=0}^{m-2} u_i) \qquad (m \geq 2).$

Thus, $(2m-3)u_{m-1} = 2(1 + \sum_{i=0}^{m-3} u_i) \qquad (m \geq 3).$

Subtracting, $(2m-1)u_m - (2m-3)u_{m-1} = 2u_{m-2}, \qquad (m \geq 3)$

or, $$(2m-1)(u_m - u_{m-1}) = -2(u_{m-1} - u_{m-2}) \tag{5}$$

Let $v_m = u_m - u_{m-1}$, $v_0 = u_0$, so $u_m = \sum_{i=0}^{m} v_i$ and $v_0 = 0$, $v_1 = 1$, $v_2 = -\frac{1}{3}$.

So, from (5), $$v_m = \frac{-2}{2m-1} v_{m-1} \qquad (m \geq 3). \tag{6}$$

Thus, $$v_m = \frac{-(-2)^{m-2}}{(2m-1)!!} \qquad (m \geq 3).$$

Therefore, $$\begin{aligned} u_m &= 0 + 1 - \tfrac{1}{3} - \sum_{i=3}^{m} (-2)^{i-2}/(2i-1)!! && (m \geq 3), \\ &= \tfrac{2}{3} + 2 \sum_{k=0}^{m-3} (-2)^k/(2k+5)!! && (m \geq 3), \end{aligned}$$

with $u_0 = 0$, $u_1 = 1$, $u_2 = \frac{2}{3}$.

Now $$\begin{aligned} t_m = \sum_{j=0}^{m} u_j &= \sum_{j=0}^{m} \sum_{i=0}^{j} v_i \\ &= \sum_{j=0}^{m} (m+1-j) v_j \\ &= (m+1) \sum_{j=0}^{m} v_j - \sum_{j=0}^{m} j v_j \\ &= (m+1) u_m - \tfrac{1}{2} \sum_{j=0}^{m} ((2j-1)+1) v_j \\ &= (m+\tfrac{1}{2}) u_m - \tfrac{1}{2} \sum_{j=0}^{m} (2j-1) v_j \\ &= (m+\tfrac{1}{2}) u_m - \tfrac{1}{2} \sum_{j=3}^{m} (2j-1) v_j, \end{aligned}$$

$$= (m+\tfrac{1}{2})u_m + \sum_{j=2}^{m-1} v_j, \qquad \text{from (6)},$$

$$= (m+\tfrac{1}{2})u_m + u_m - v_m - 1$$

i.e. $$t_m = mu_m + (\tfrac{3}{2}u_m - v_m - 1), \qquad (m \geq 2).$$

For large m, u_m is close to α and v_m is very small. Therefore let us define ϵ_m by

$$t_m = m\alpha + \beta + \epsilon_m \qquad (m \geq 1), \tag{7}$$

where
$$\beta = \tfrac{3}{2}\alpha - 1 = 0.1535563\cdots \tag{8}$$

and, for $m \geq 2$,
$$\epsilon_m = (m+\tfrac{3}{2})(u_m - \alpha) - v_m.$$

It follows from Lemma 4 below that $\epsilon_m = O(\frac{1}{(m+1)!})$, and thus

$$r(T_m) = \alpha + \beta/m + O(\tfrac{1}{(m+2)!}),$$

so, in particular,
$$r(T_m) \to \alpha \text{ as } m \to \infty. \qquad \square$$

Numerically, the first two terms in (7) are an excellent approximation to t_m. When $m = 10$, for example, the error term ϵ_{10} is less than 10^{-8}.

Our lower bound argument requires knowledge of the behaviour of the sequence $\{\epsilon_m\}$. Its first few terms are (approximately)

$$\begin{array}{lll} \epsilon_1 = +0.0774006, & \epsilon_2 = -0.0249725, & \epsilon_3 = +0.0059878, \\ \epsilon_4 = -0.0011472, & \epsilon_5 = +0.0001834. & \end{array} \tag{9}$$

Lemma 4 $$\epsilon_m = (m+\tfrac{3}{2})(u_{m+2} - \alpha).$$

Proof For $m = 1$, this may be verified directly.

For $m \geq 2$,
$$\epsilon_m = (m+\tfrac{3}{2})(u_m - \alpha) - v_m,$$

and the lemma amounts to the claim that

$$v_m = -(m + \tfrac{3}{2})(v_{m+1} + v_{m+2}).$$

But, from (6), this is the obvious identity

$$1 = -(m + \tfrac{3}{2})\left(\frac{-2}{2m+1} + \frac{4}{(2m+1)(2m+3)}\right). \qquad \square$$

Corollary 7

(a) $\epsilon_{2k-1} > 0$ *and* $\epsilon_{2k} < 0$ $(k = 1, 2, \ldots)$

(b) $|\epsilon_{m+1}| \leq 2|\epsilon_m|/(2m+3)$ *(*$m = 1, 2, \ldots$*), and hence* $\{(2m-1)|\epsilon_m|\}$ *is a decreasing sequence.*

Proof

(a) These inequalities follow from Lemma 4, since $\epsilon_m = -(m+\frac{3}{2})\sum_{i=m+3}^{\infty} v_i$, the v_i alternate in sign and decrease strictly in absolute value.

(b) The case $m = 1$ can be verified by direct calculation. For $m \geq 2$,

$$(m + \tfrac{3}{2})(|v_{m+3}| - |v_{m+4}|) \leq |\epsilon_m| \leq (m + \tfrac{3}{2})|v_{m+3}|,$$

and so, using (6),

$$\frac{|\epsilon_m|}{|\epsilon_{m+1}|} \geq \frac{(2m+3)}{(2m+5)}\left(\frac{|v_{m+3}|}{|v_{m+4}|} - 1\right) = \frac{2m+3}{2} > \frac{2m+1}{2m-1}.$$

$\square$

We now prove the lower bound for FORESTS. We prove that the worst-case examples for forests are the trees T_m of the previous lemma. Let $\mathcal{T} = \{T_m : m = 1, 2, \ldots\}$.

Lemma 5 *Let* T *be an* L*-tree with* $2m$ *vertices. Then*

$$\mu(T) \geq t_m$$

Proof We establish by induction a bound

$$\mu(T) \geq m\alpha + \beta + {\epsilon_m}^+$$

provided $T \neq T_m$ (${\epsilon_m}^+ = \max\{0, \epsilon_m\}$.)

This will, of course, prove the lemma.

We next define

$$\begin{aligned} \hat{\epsilon}_i &= \epsilon_i \quad \text{if } i = 1 \text{ or } i \text{ is even} \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

$$\mu(T) = 1 + \frac{1}{|E|} \sum_{uv \in E} \mu(T \setminus \{u, v\}) \tag{10}$$

Thus suppose $e = uv$ and $T \setminus \{u, v\}$ has components $\{C_i : 1 \leq i \leq k_e\}$ which contain at least one edge. Now define $\delta(C_i)$ $(i = 1, 2, \ldots, k_e)$ by

$$\begin{aligned} \delta(C_i) &= \hat{\epsilon}_{m'} \quad \text{if } C_i = T_{m'} \\ &= 0 \quad \text{otherwise} \end{aligned}$$

where $m' = m(C_i)$, and let $\gamma_e = \sum_{i=1}^{k_e} \delta(C_i)$ for $e \in E$. (The precise form of this definition may seem curious, but will be justified later in our proof.) Now, by induction

$$\begin{aligned} \mu(T \setminus \{u, v\}) = \sum_{i=1}^{k_e} \mu(C_i) &\geq \sum_{i=1}^{k_e} (\alpha m(C_i) + \beta + \delta(C_i)) \\ &= \alpha m(T \setminus \{u, v\}) + \beta k_e + \gamma_e. \end{aligned}$$

Hence

$$\sum_{uv \in E} \mu(T \setminus \{u, v\}) \geq \alpha \sum_{uv \in E} m(T \setminus \{u, v\}) + \beta \sum_{e \in E} k_e + \sum_{e \in E} \gamma_e$$

$$= \alpha(m(m-1) + (m-1)(m-2)) + \beta \sum_{e \in E} k_e + \sum_{e \in E} \gamma_e,$$

using the fact that T is an L-tree.

So

$$\sum_{uv \in E} \mu(T \setminus \{u, v\}) \geq 2(m-1)^2\alpha + \beta \sum_{e \in E} k_e + \sum_{e \in E} \gamma_e. \tag{11}$$

Now let

$$\Delta_0(T) = (\mu(T) - (m\alpha + \beta + {\epsilon_m}^+))(2m-1).$$

We must show that $\Delta_0(T) \geq 0$. But (10) and (11) imply

$$\begin{aligned} \Delta_0(T) &\geq 2m - 1 + 2(m-1)^2\alpha + \beta \sum_{e \in E} k_e + \sum_{e \in E} \gamma_e & (12) \\ &\quad - (2m-1)(m\alpha + \beta + {\epsilon_m}^+) \\ &= 2\alpha + \beta - 1 - (2m-1){\epsilon_m}^+ - 4m\beta + \beta \sum_{e \in E} k_e + \sum_{e \in E} \gamma_e. \\ &\geq 2\alpha + \beta - 1 - 9\epsilon_5 - 4m\beta + \beta \sum_{e \in E} k_e + \sum_{e \in E} \gamma_e. & (13) \end{aligned}$$

by Corollary 7. Now let $\Delta_1(T)$ denote the right hand side of (13). We show that $\Delta_1(T) \geq 0$. We prove this by induction on m. Note that, if $m < 4$ there is nothing to prove. Now our base cases will be, for each $m \geq 4$, the trees $S(a, b, c)$ where a, b, c are positive integers and $a + b + c + 1 = m$. The tree $S(a, b, c)$ consists of a central vertex v, paths $v, x_a, x_{a-1}, \ldots, x_1$, $v, y_b, y_{b-1}, \ldots, y_1$ and $v, z_c, z_{c-1}, \ldots, z_1$ plus leaves w, $x'_1, \ldots, x'_a$, $y'_1, \ldots y'_b$, $z'_1, \ldots z'_c$ where v is adjacent to w, x'_i is adjacent to x_i and so on.

We first evaluate $\sum_{e\in E} k_e$. This is the sum of terms

CONTRIBUTION	EDGES
$2((a-1)+(b-1)+(c-1))$	$x_i x_i'$ $(i \geq 2), \ldots$
$2((a-2)+(b-2)+(c-2))$	$x_i x_{i+1}$ $(i \geq 2), \ldots$
6	$x_1 x_1'$ and $x_1 x_2, \ldots$
9	$v x_a, \ldots$
3	vw

which sum to $4(a+b+c) = 4(m-1)$.

The above analysis only applies if $a, b, c \geq 2$. On the other hand, if say $a = 1$, then the path vx_1 contributes $1 + 2 = 2(a-1) + 2(a-2) + 2 + 3$ and so $\sum_{e\in E} k_e = 4(m-1)$ in this case also.

We must now compute $\sum_{e\in E} \gamma_e$. This comprises terms

CONTRIBUTION	EDGES
$\sum_{i=1}^{a-3} \hat{\epsilon}_i + \sum_{i=1}^{b-3} \hat{\epsilon}_i + \sum_{i=1}^{c-3} \hat{\epsilon}_i$	$x_i x_{i+1}$ $(2 \leq i \leq a-2), \ldots$
$\sum_{i=1}^{a-2} \hat{\epsilon}_i + \sum_{i=1}^{b-2} \hat{\epsilon}_i + \sum_{i=1}^{c-2} \hat{\epsilon}_i$	$x_i x_i'$ $(2 \leq i \leq a-1), \ldots$
$\hat{\epsilon}_{a-2} + \hat{\epsilon}_{b+c+1} + \hat{\epsilon}_{b-2} + \hat{\epsilon}_{a+c+1} + \hat{\epsilon}_{c-2} + \hat{\epsilon}_{a+b+1}$	$x_{a-1} x_a, \ldots$
$\hat{\epsilon}_{a-1} + \hat{\epsilon}_{b+c+1} + \hat{\epsilon}_{b-1} + \hat{\epsilon}_{a+c+1} + \hat{\epsilon}_{c-1} + \hat{\epsilon}_{a+b+1}$	$x_a x_a', \ldots$
$\hat{\epsilon}_{a-1} + \hat{\epsilon}_b + \hat{\epsilon}_c + \hat{\epsilon}_{b-1} + \hat{\epsilon}_a + \hat{\epsilon}_c + \hat{\epsilon}_{c-1} + \hat{\epsilon}_a + \hat{\epsilon}_b$	$x_a v, \ldots$
$\hat{\epsilon}_a + \hat{\epsilon}_b + \hat{\epsilon}_c$	vw

which sum to

$$2\sum_{i=1}^{a} \hat{\epsilon}_i + 2\sum_{i=1}^{b} \hat{\epsilon}_i + 2\sum_{i=1}^{c} \hat{\epsilon}_i + \hat{\epsilon}_a + \hat{\epsilon}_b + \hat{\epsilon}_c + 2\hat{\epsilon}_{a+b+1} + 2\hat{\epsilon}_{a+c+1} + 2\hat{\epsilon}_{b+c+1}.$$

Now, for any $p \geq 1$, by Corollary 7,

$$\sum_{i=1}^{p} \hat{\epsilon}_i \;\; \geq \;\; \epsilon_1 + \epsilon_2 + \epsilon_4 \left(1 + \frac{4}{11 \cdot 13} + \frac{16}{11 \cdot 13 \cdot 15 \cdot 17} + \ldots\right)$$

$$
\begin{aligned}
&> \epsilon_1 + \epsilon_2 + \epsilon_4 \left(1 + \frac{4}{143} + \left(\frac{4}{143}\right)^2 + \dots\right) \\
&= \epsilon_1 + \epsilon_2 + 143\epsilon_4/139 \\
&\geq 0.05, \quad \text{by direct calculation.}
\end{aligned} \tag{14}
$$

Thus

$$
\begin{aligned}
\sum_{e\in E} \gamma_e &\geq 0.3 + \hat{\epsilon}_a + \hat{\epsilon}_b + \hat{\epsilon}_c + 2\hat{\epsilon}_{a+b+1} + 2\hat{\epsilon}_{a+c+1} + 2\hat{\epsilon}_{b+c+1} \\
&\geq 3\epsilon_2 + 6\epsilon_4 + .3 \\
&> 0.2,
\end{aligned}
$$

and hence

$$
\begin{aligned}
\Delta_1(T) &\geq 2\alpha + \beta - 1 - 9\epsilon_5 - 4m\beta + 4(m-1)\beta + 3(\epsilon_2 + \epsilon_4) + .3 \\
&> 0, \quad \text{by direct calculation.}
\end{aligned}
$$

(See Theorem 4, (8) and (9) for numerical estimates of the relevant quantities.)

We now have a basis for our induction. So suppose that $T \neq T_m$ and T is not of the form $S(a,b,c)$. Let $\hat{T}$ be obtained from T by deleting all leaves. Let x be a leaf of $\hat{T}$. Consider the path from x in $\hat{T}$, $P = \{x_1(=x), x_2, \dots x_k, y\}$, where the degree of x_i (in $\hat{T}$) is 2, ($i = 2, 3, \dots, k$) and the degree of y is $d+1$ ($d \geq 2$). Now, in T, let $x_i', i = 1, 2, \dots, k$ be the leaf neighbours of x_i, y' be the leaf neighbour of y, $z_1, z_2, \dots, z_d$ be the non-leaf neighbours of y and finally, let $z_i', i = 1, 2, \dots, d$ be the leaf neighbours of z_i. Let $T' = T \setminus P$ (after deleting all isolated vertices). Then $T' \neq T_{m-k}$ as T is not of the form $S(a,b,c)$. Let now $D = \Delta_1(T) - \Delta_1(T')$. Then

$$
D = (D_1 - 4k)\beta + D_2 \tag{15}
$$

where

$$D_1 = \sum_{e \in E} k_e - \sum_{e \in E'} k'_e$$

and

$$D_2 = \sum_{e \in E} \gamma_e - \sum_{e \in E'} \gamma'_e$$

and where E' is the edge set of T' and k'_e, γ'_e stand for k_e, γ_e in T'.

We will now justify our definition of γ_e. For each component C_i associated with edge e in T, there is a (possibly edgeless) component $C'_i = C_i \cap E'$ in T'. Now $(\gamma_e - \gamma'_e)$ is the sum of terms $(\delta(C_i) - \delta(C'_i))$ for $C_i \neq C'_i$. Now we certainly have $(\delta(C_i) - \delta(C'_i)) \geq \delta(C_i)$ unless $C'_i = T_1$. But clearly $C'_i = T_1$ implies $C_i = T_1$ so the inequality holds regardless. Therefore we may justifiably use the bound

$$\gamma_e - \gamma'_e \geq \sum_{C'_i \neq C_i \in \mathcal{T}} \delta(C_i)$$

The consequence is that we do not need to consider the effect of the "destruction" of members of $\mathcal{T}$ in T', only their "creation" in T. This allows combination of what might otherwise be distinct cases in the induction. We use these ideas below without further comment.

Now, when $k \geq 2$,

$$\begin{aligned} D_1 &= 1 + 1 + 2(k-1) + 2(k-2) + (d+1) + 1 + d \\ &= 4k + 2d - 2 \end{aligned}$$

The successive terms in the expression for D_1 are the contributions due, respectively, to $x_1x'_1$, x_1x_2, $x_ix'_i$ $(2 \leq i \leq k)$, x_ix_{i+1} $(2 \leq i \leq k-1)$, x_ky, yy' and yz_i $(1 \leq i \leq d)$. The last two terms are increases $(k_e - k'_e)$. For all other edges $e \in E'$, it is clear that $k_e = k'_e$.

The expression $4k + 2d - 2$ is also valid for $k = 1$, as the reader can easily check.

We now turn to D_2. If y is deleted from T then T breaks up into T_k plus d subtrees $H_1, H_2, \ldots H_d$, say. Let us assume, without loss, that $H_i \in \mathcal{T}$ for $1 \leq i \leq s$, for some $0 \leq s \leq d$. Note that, if $d = 2$, then $s < 2$ since T is not an $S(a, b, c)$.

We consider two cases.

Case 1: $d \geq 3$ or $d = 2$, $s = 0$.

For $k \geq 2$,

$$D_2 = 2\hat{\epsilon}_1 + 2\hat{\epsilon}_2 + \ldots + 2\hat{\epsilon}_{k-2} + \hat{\epsilon}_{k-1} + (\hat{\epsilon}_{k-1} + \hat{\epsilon}_{i_1} + \ldots + \hat{\epsilon}_{i_s}) + \hat{\epsilon}_k + d\hat{\epsilon}_k.$$

The terms here arise as follows. The terms $2\hat{\epsilon}_i$, $(1 \leq i \leq k-2)$ come from $x_{i+1}x'_{i+1}$ and $x_{i+1}x_{i+2}$; $\hat{\epsilon}_{k-1}$ comes from $x_k x'_k$; $(\hat{\epsilon}_{k-1} + \hat{\epsilon}_{i_1} + \cdots + \hat{\epsilon}_{i_s})$ comes from $x_k y_1$; $\hat{\epsilon}_k$ comes from yy' and, finally, $d\hat{\epsilon}_k$ comes from yz_i, $(1 \leq i \leq d)$.

So

$$D_2 = 2\sum_{i=1}^{k} \hat{\epsilon}_i + (d-1)\hat{\epsilon}_k + \hat{\epsilon}_{i_1} + \ldots + \hat{\epsilon}_{i_s}.$$

The same expression is also valid for $k = 1$.

Thus, from (15),

$$\begin{aligned} D &\geq (2d-2)\beta + 2\sum_{i=1}^{k} \hat{\epsilon}_i + (d-1)\hat{\epsilon}_k + \hat{\epsilon}_{i_1} + \ldots + \hat{\epsilon}_{i_s} \\ &\geq 0.1 + (d-1)(2\beta + \hat{\epsilon}_k + d\epsilon_2/(d-1)) \\ &\geq 0.1 + (d-1)(2\beta + 3\epsilon_2) \\ &> (d-1)/5 \\ &> 0. \end{aligned}$$

Now, by the induction hypothesis, $\Delta_1(T') \geq 0$ and so $\Delta_1(T) \geq 0$ is immediate.

Case 2: $d = 2$, $s = 1$.

Let the degree of z_2 be $d' + 2$, where $d' \geq 1$ (since T is not of the form $S(a, b, c)$). Let $w_1, \ldots, w_{d'}$ be the neighbours of z_2 other than y and z_2'. Assume first that $d' \geq 2$. Now the only difference from Case 1 is that $\gamma_{z_2 z_2'}$ and $\gamma_{z_2 w_i}$, $(i = 1, 2, \ldots, d')$ contain new contributions $\hat{\epsilon}_{k+p+1} \geq \epsilon_4$, and so

$$D \geq 0.1 + 2\beta + 3\epsilon_2 + (d' + 1)\epsilon_4. \tag{16}$$

If $d' \leq 10$ then $D > 0$, by direct calculation, and so $\Delta_1(T) \geq \Delta_1(T') \geq 0$ as in the previous case. For $d' > 10$ we could have $D < 0$ and so we let $T'' = T' \setminus (H_1 \cup \{y\})$, the tree obtained by deleting vertex y and H_1 from T'. By the analysis of the previous case $\Delta_1(T') \geq \Delta_1(T'') + (d'-1)/5 \geq (d'-1)/5$, since T'' is clearly not in $\mathcal{T}$ and is not an $S(a, b, c)$. Hence, from (16),

$$\begin{aligned} \Delta_1(T) &\geq 0.1 + 2\beta + 3\epsilon_2 + 2\epsilon_4 + (d' - 1)(0.2 + \epsilon_4) \\ &\geq 0.1 + 2\beta + 3\epsilon_2 + 2\epsilon_4 + 10(0.2 + \epsilon_4) \\ &> 0 \quad \text{by direct calculation.} \end{aligned}$$

This completes the induction, and the proof of the lemma. □

9 Concluding remarks

We have established the worst case examples for FORESTS, but we have little idea for PLANAR GRAPHS. The lower bound $\frac{6}{11}$ is almost certainly not tight, but currently our best upper bound is G_4 from Theorem 2 which gives $\rho(\text{PLANAR GRAPHS}) \leq \frac{11}{15}$. This leaves a large gap.

A randomised greedy algorithm could also be applied to the problem of finding a large H-matching in a graph G. We consider H to be some fixed graph like a triangle, and an H-matching of G is a collection of vertex-disjoint copies of H in G. Finding the largest H-matching is NP-hard for all graphs H with at least three vertices, Kirkpatrick and Hell [2]. One difficulty with extending our analysis to this case is that the analogue of Lemma 1 fails to hold in general. Consider, for example, H to be a path of length two, and compare the performance of the RGA on two paths of length two with its performance on a path of length six.

Another possible generalisation is to weighted problems. For example, in the weighted matching problem, we might consider choosing the next edge with probability proportional to its edge weight.

References

[1] B. Bollobás, *Martingales, isoperimetric inequalities and random graphs*, Coll. Math. Soc. J. Bolyai **52**, 113-139 (1987).

[2] D. G. Kirkpatrick and P. Hell, *On the complexity of a generalised matching problem*, Proceedings of the 10th Annual ACM Symposium on Theory of Computing, ACM, New York, 1978, pp. 240-245.

[3] 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.

[4] C. J. H. McDiarmid, *On the method of bounded differences*, in Surveys in Combinatorics, Proceedings of the Twelfth British Combinatorial Conference (J. Siemons, Ed.), 1989, pp. 148-188.