RANDOMIZED GREEDY MATCHING

by

Martin Dyer
School of Computer Studies
University of Leeds
Leeds, U.K.
and
A. M. Frieze
Department of Mathematics
Carnegie Mellon University
Pittsburgh, PA 15213

Research Report No. 90-75 2

February 1990

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.

January 30, 1990

Abstract

We consider a randomized version of the usual greedy algorithm for finding a large matching in a graph. We assume that the next edge is randomly chosen from those remaining at any stage. We analyse the expected 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 prove that randomization does produce an improvement over the worst-case, the ratios to maximum being at least $\frac{6}{11}$ and $0.76\cdots$ respectively.

*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 [2] 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 $G = (E, V)$ 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

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

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 $G \in \mathcal{G}$.

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

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 $0 \leq \kappa(\mathcal{G}) \leq 2$ 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* $v \in V$, $\mu(G) \geq \mu(G \setminus \{v\}) \geq \mu(G) - 1$

Proof The statement clearly holds for $G = (\{v\}, \emptyset)$ and we argue by induction on $|V|$. Let us carry out the *RGA* in G and mimic it in $G \setminus \{v\}$. Owing to the uniform choice mechanism, the simulation will be successful until some random edge $\{u, v\}$ is chosen in G. Suppose k edges have been chosen in $G \setminus \{v\}$, and let H be the remaining subgraph of $G \setminus \{v\}$. The size of the final matchings will thus be, in expectation, $1 + k + \mathbf{E}_u\{\mu(H \setminus \{u\})\}$ in G, and $k + \mu(H)$ in $G \setminus \{v\}$. Let Δ be the difference, so

$$\begin{aligned} \Delta &= 1 + \mathbf{E}_u\{\mu(H \setminus \{u\})\} - \mu(H) \\ &= \mathbf{E}_u\{1 + \mu(H \setminus \{u\}) - \mu(H)\} \end{aligned}$$

By induction, since $|V(H)| < |V(G)|$,

$$0 \leq 1 + \mu(H \setminus \{u\}) - \mu(H) \leq 1.$$

So $0 \leq \Delta \leq 1$. Since $\mu(G) - \mu(G \setminus \{v\}) = \mathbf{E}(\Delta)$, the conclusion follows. □

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 $\rho(H) \leq \rho(G)$. $\square$

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.

Lemma 2 *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

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

However

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

Thus,

$$\sum_{\{u,v\}\in E} m(G \setminus \{u,v\}) \geq m(m-1) + (|E| - m)(m-2)$$

$$= |E|(m-2) + m.$$

Hence, using the inductive hypothesis,

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

completing the induction. □

Corollary 3 $\rho(\text{GRAPHS}) \geq \frac{1}{2}$, $\rho(\text{PLANAR GRAPHS}) \geq \frac{6}{11}$, $\rho(\text{FORESTS}) \geq \frac{2}{3}$. □

5 The class GRAPHS

Lemma 3 $\rho(\text{GRAPHS}) = \frac{1}{2}$.

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,

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

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 $m \to \infty$. $\square$

6 Concentration near the mean

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

Lemma 4 *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*

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

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 $\{u, v\}$. 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,

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

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 [3]) that

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

□

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 Lemma 4. □

Corollary 5 *If $\{G_m\}$ is the graph sequence defined in the proof of Lemma 3, 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 [2]). Putting $\epsilon = \log m/\sqrt{m}$ in Lemma 4, the probability of $\hat{X}(G_m)$ not falling in the required interval is at most $2p(m)e^{-(\log m)^2/2} \to 0$ 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 $\{v, w\}$ $(w \in \Gamma(v) \setminus \{u\})$ with $\{u, w\}$.

Note the restriction that $\{u, v\}$ does not lie in a triangle, to ensure that G' is a simple graph.

Lemma 5 $\quad r(G') \leq r(G)$.

Proof $\quad$ Clearly $m(G') = m(G)$ so we only need show $\mu(G') \leq \mu(G)$. We use the "simulation" argument of Lemma 1. The realisations of RGA in G, G' proceed identically until some edge which meets either u or v is chosen. At this stage, suppose the remaining subgraph of G is H. If $\{u, v\}$ is chosen, the remaining graph is $H \setminus \{u, v\}$ in both cases. Otherwise we have, say, $H \setminus \{u, w\}$ and $H \setminus \{u, v, w\}$ in G'. (Note that G' is only re-labelled by changing the roles of u, v in the transformation.) By Lemma 1, the expected value obtained by the remainder of the RGA is at least as large in G as in G' in all cases. Thus, taking expectations, we have $\mu(G') \leq \mu(G)$. $\square$

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 $\quad \rho(\mathcal{H}^*) = \rho(\mathcal{G}^*)$. $\square$

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.

8 The class FORESTS

For FORESTS, Corollary 3 gives $\rho \geq \frac{2}{3}$, but this is not tight. First let us establish an upper bound.

Lemma 6 $\rho(\text{FORESTS}) \leq \frac{2}{3} + 2\sum_{k=0}^{\infty} \frac{(-2)^k}{(2k+5)!!} = 0.7690397\cdots$ *(where n!! = n(n-2)(n-4)...3.1 for n odd).*

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 $t_1 = 0$, $t_2 = 1$. Clearly, for $m \geq 2$,

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

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

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, substitution shows that (4) holds also for $m = 2$.

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 $u_m = \sum_{i=0}^{m} v_m$ 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, $$u_m = 0 + 1 - \tfrac{1}{3} - \sum_{i=3}^{m} (-2)^{i-2}/(2i-1)!! \qquad (m \geq 3),$$

$$= \tfrac{2}{3} + 2\sum_{k=0}^{m-3} (-2)^k/(2k+5)!! \qquad (m \geq 3),$$

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

Now $t_m = \sum_{j=0}^{m} u_j = \sum_{j=0}^{m}\sum_{i=0}^{j} v_i$

$$
\begin{aligned}
&= \sum_{j=0}^{m}(m+1-j)v_j \\
&= (m+1)\sum_{j=0}^{m} v_j - \sum_{j=0}^{m} jv_j \\
&= (m+1)u_m - \tfrac{1}{2}\sum_{j=0}^{m}((2j-1)+1)v_j \\
&= (m+\tfrac{1}{2})u_m - \tfrac{1}{2}\sum_{j=0}^{m}(2j-1)v_j \\
&= (m+\tfrac{1}{2})u_m - \tfrac{1}{2}\sum_{j=3}^{m}(2j-1)v_j, \\
&= (m+\tfrac{1}{2})u_m + \tfrac{1}{2}\sum_{j=2}^{m-1} v_j, \qquad \text{from (6),} \\
&= (m+\tfrac{1}{2})u_m + u_m - v_m - 1
\end{aligned}
$$

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

For large m, $$u_m = \tfrac{2}{3} + 2\sum_{k=0}^{\infty}(-2)^k/(2k+5)!! + O(\tfrac{1}{(m-1)!})$$
$$= 0.76903975\cdots + O(\tfrac{1}{(m-1)!}).$$

So, letting $u = 0.76903975\cdots$,

$$t_m = mu + \tfrac{3}{2}u - 1 + O(\tfrac{1}{(m-2)!}). \tag{7}$$

Thus $$r(T_m) = u + (\tfrac{3}{2}u - 1)/m + O(\tfrac{1}{(m-1)!}),$$

and so $$r(T_m) \to u \text{ as } m \to \infty.$$

$\square$

Numerical results confirm that (7) is indeed an excellent approximation to t_m. When $m = 10$, for example, the implied error term is less than 10^{-8}.

We now consider a lower bound for FORESTS. We will need the following Lemma.

Lemma 7 *Let $T = (E, V)$ be a tree in which each interior vertex is adjacent to exactly one leaf. If $e = \{u, v\} \in E$, let k_e be the number of components in $T \setminus \{u, v\}$ which contain an edge. Then*

$$\sum_{e \in E} k_e \;\geq\; (2n - 6).$$

Moreover, unless $T = T_m$ $(m = n/2)$ as defined in Lemma 6,

$$\sum_{e \in E} k_e \;\geq\; (2n - 4).$$

Proof Let $I, L \subseteq V$ be the internal vertices and leaves of T. Clearly $|I| = |L| = \frac{1}{2}n$. If $e = \{u, v\}$ then clearly

$$k_e = (d_u - 1) + (d_v - 1) - \delta_u - \delta_v$$

where d_u is the degree of u and $\delta_u = 1$ $(u \in I)$ or $\delta_u = -1$ $(u \in L)$. Thus,

$$\begin{aligned}
\sum_{e \in E} k_e &= \sum_{e \in E} (d_u + d_v) - 2|E| - \sum_{e \in E} (\delta_u + \delta_v) \\
&= \sum_{u \in V} {d_u}^2 - 2(n - 1) - \sum_{u \in V} d_u \delta_u \\
&= \sum_{u \in V} {d_u}^2 - 2(n - 1) - \sum_{u \in I} d_u + |L| \\
&= \sum_{u \in V} {d_u}^2 - 2(n - 1) - \sum_{u \in I} (d_u - 1) \qquad (8)
\end{aligned}$$

By deleting L from T we are left with a tree on the vertex set I with degrees $(d_u - 1)$. Thus,

$$\sum_{u \in I} (d_u - 1) = 2(|I| - 1) = n - 2.$$

Therefore, $$\sum_{e \in E} k_e = \sum_{u \in V} {d_u}^2 - 3n + 4, \qquad \text{from (8)},$$

i.e. $$\sum_{e \in E} k_e = \sum_{u \in I} {d_u}^2 - \tfrac{5}{2} n + 4 \tag{9}$$

But we have $$\sum_{u \in V} d_u = 2(n-1)$$

Thus, $$\sum_{u \in I} d_u = \tfrac{3}{2} n - 2 \tag{10}$$

We must minimize the right side of (9) subject to (10). It is easy to argue by "pairwise improvements" that the optimum will occur when all the d_u are as nearly equal as possible, *i.e.* $d_u = 3$ for all but two $u \in I$, which have $d_u = 2$. Thus the the tree induced by I is a path. Then we clearly have T_m, and

$$\sum_{e \in E} k_e \geq (\tfrac{1}{2} n - 2) \cdot 9 + 2 \cdot 4 - \tfrac{5}{2} n + 4 = 2n - 6$$

If T is not T_m however, it must have value at least 2 more than the minimum in (9), because the d_u are integers with constraint (10). Thus if $T \neq T_m$, $\sum_{e \in E} k_e \geq 2n - 4$.

Lemma 8 $\rho(\text{FORESTS}) \geq \frac{16}{21} = 0.7619047 \cdots$

Proof We establish by induction a bound of the form

$$\mu(F) \geq \alpha m(F) + \beta \tag{11}$$

for forests F which contain at least one edge. We may assume $F \neq T_m$ (of Lemma 6) provided we ensure the resulting bound is also satisfied by all T_m ($m \geq 1$). Now

$$\mu(F) = 1 + \frac{1}{|E|} \sum_{\{u,v\} \in E} \mu(F \setminus \{u,v\}) \tag{12}$$

We may assume F is a tree T in which all interior vertices are adjacent to exactly one leaf, since the operations of Corollary 1 and Lemma 5 both reduce $\mu(F)$ without changing $m(F)$. Thus the right hand side of (11) is unaltered by these operations.

Thus suppose $T \setminus \{u,v\}$ has components $\{C_i : 1 \leq i \leq k_e\}$ which contain at least one edge. Then, by induction

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

Hence, $$\begin{aligned} \sum_{e \in E} \mu(T \setminus \{u,v\}) &\geq \sum_{e \in E} m(T \setminus \{u,v\}) + \beta \sum_{e \in E} k_e \\ &= \alpha(m(m-1) + (m-1)(m-2)) + \beta \sum_{e \in E} k_e, \end{aligned}$$

using the assumed structure of T.

So, $$\begin{aligned} \sum_{e \in E} \mu(T \setminus \{u,v\}) &= 2(m-1)^2\alpha + \beta \sum_{e \in E} k_e, \\ &\geq 2(m-1)^2\alpha + \beta(4m-4), \end{aligned}$$

from Lemma 7, since $T \neq T_m$ and $n = 2m$,

$$= 2(m-1)((m-1)\alpha + 2\beta).$$

Thus, from (12),

$$
\begin{aligned}
\mu(F) &\geq 1+\frac{1}{2m-1}\cdot 2(m-1)((m-1)\alpha+2\beta) \\
&= \alpha m+\beta+\frac{m(2-3\alpha+2\beta)+(2\alpha-3\beta-1)}{2m-1} \\
&\geq \alpha m+\beta
\end{aligned}
$$

provided $2-3\alpha+2\beta \geq 0$ and $2\alpha-3\beta-1 \geq 0$. (13)

Thus $\beta \geq \frac{3}{2}\alpha - 1$ (*c.f.* (7)) and (13) has a solution for any $\alpha \leq \frac{4}{5}$.

We must now consider T_m. Suppose $\mu(T_m) = t_m \geq \alpha m + \beta$ for all $m < k$ with $k \geq 3$. From (4),

$$
\begin{aligned}
t_k &\geq t_{k-1}+\frac{2}{2k-1}(1+t_{k-2}) \\
&\geq \alpha(k-1)+\beta+\frac{2}{2k-1}(1+\alpha(k-2)+\beta) \\
&= \alpha k+\beta+\frac{2-3\alpha+2\beta}{2k-1} \\
&\geq \alpha k+\beta,
\end{aligned}
$$

provided $2-3\alpha+2\beta \geq 0$ (*c.f.* (13)). Thus, if we take $\beta = \frac{3}{2}\alpha - 1$, we need only check

$$
\begin{aligned}
t_1 &= 1 \geq \alpha\cdot 1+\tfrac{3}{2}\alpha-1, \\
t_2 &= \tfrac{5}{3} \geq \alpha\cdot 2+\tfrac{3}{2}\alpha-1.
\end{aligned}
$$

These give $\alpha \leq \frac{4}{5}$, $\alpha \leq \frac{16}{21}$, respectively. Thus taking $\alpha = \frac{16}{21}$ (and hence $\beta = \frac{1}{7}$) we have proved $\mu(F) \geq \frac{16}{21}m(F) + \frac{1}{7}$ for all forests F containing an edge. Thus $r(F) \geq \frac{16}{21}$ and hence $\rho(\text{FORESTS}) \geq \frac{16}{21}$. □

9 Concluding remarks

From Lemmas 6 and 8, we have $0.761\cdots < \rho(\text{FORESTS}) < 0.769\cdots$ (the difference in the bounds is less than 1%.) Therefore it seems reasonable to make the following

Conjecture If F is a forest with n vertices then $r(F) \geq t_{\lfloor n/2 \rfloor}$.

While we believe that the trees T_m of Lemma 6 are worst case examples for FORESTS, 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 Lemma 3 which gives $\rho(\text{PLANAR GRAPHS}) \leq \frac{11}{15}$. This leaves a very large gap.

Finally, we note that there are other classes of graphs which fall within the scope of our analysis, for example graphs with degrees bounded by a given positive integer.

References

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

[2] 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.

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