AN EXACT PERFORMANCE BOUND FOR AN $\mathbf{O(m+n)}$ TIME GREEDY MATCHING PROCEDURE

Andrew Shapira

ECSE Department
Rensselaer Polytechnic Institute
Troy, New York 12180
`shapiraa@cs.rpi.edu`

Submitted July 20, 1997; accepted October 15, 1997.

AMS Subject Classification (1991). Primary: 05C70. Secondary: 68Q25, 05C35, 05C85, 68R05, 68R10.

Key Words. Analysis of algorithms, extremal problems, greedy procedure, matching algorithms, matching heuristics, maximum matching, performance guarantees.

Abstract. We prove an exact lower bound on $\gamma(G)$, the size of the smallest matching that a certain $\mathrm{O}(m+n)$ time greedy matching procedure may find for a given graph G with n vertices and m edges. The bound is precisely Erdös and Gallai's extremal function that gives the size of the smallest maximum matching, over all graphs with n vertices and m edges. Thus the greedy procedure is optimal in the sense that when only n and m are specified, no algorithm can be guaranteed to find a larger matching than the greedy procedure. The greedy procedure and augmenting path algorithms are seen to be complementary: the greedy procedure finds a large matching for dense graphs, while augmenting path algorithms are fast for sparse graphs. Well known hybrid algorithms consisting of the greedy procedure followed by an augmenting path algorithm are shown to be faster than the augmenting path algorithm alone. The lower bound on $\gamma(G)$ is a stronger version of Erdös and Gallai's result, and so the proof of the lower bound is a new way of proving of Erdös and Gallai's result.

1 Introduction

The following procedure is sometimes recommended for finding a matching that is used as an initial matching by a maximum cardinality matching algorithm [10] . Start with the empty matching, and repeat the following step until the graph has no edges: remove all isolated vertices, select a vertex v of minimum degree, select a neighbor w of v that has minimum degree among v's neighbors, add $\{v, w\}$ to the current matching, and remove v and w from the graph. This procedure is referred to in this paper as "the greedy matching procedure" or "the greedy procedure."

In the worst case, the greedy procedure performs poorly. For all $r \geq 3$, a graph D_r of order $4r+6$ can be constructed such that the greedy procedure finds a matching for D_r that is only about half the size of a maximum matching [13] . This performance is as poor as that of any procedure that finds a maximal matching.

On the other hand, there are classes of graphs for which the greedy procedure always finds a maximum matching [13] . Furthermore, using a straightforward kind of priority queue that has one bucket for each of the n possible vertex degrees, the greedy procedure can be made to run in $\mathrm{O}(m+n)$ time and storage for a given graph with n vertices and m edges [14] . The $\mathrm{O}(m+n)$ running time is asymptotically faster than the fastest known maximum matching algorithm for general graphs or bipartite graphs [1, 3, 5, 6, 7, 8, 9, 11]. The greedy procedure's success on some graphs, $\mathrm{O}(m+n)$ time and storage requirements, low overhead, and simplicity motivate the investigation of its performance.

The matching found by the greedy procedure may depend on how ties are broken. Let $\gamma(G)$ be the size of the smallest matching that can be found for a given graph G by the greedy procedure, i.e., $\gamma(G)$ is the worst case matching size, taken over all possible ways of breaking ties.

We will show that each graph G with n vertices and $m \geq 1$ edges satisfies

$$\gamma(G) \geq \min(\left\lfloor n + \frac{1}{2} - \sqrt{n^2 - n - 2m + \frac{9}{4}} \right\rfloor, \left\lfloor \frac{3}{4} + \sqrt{\frac{m}{2} - \frac{7}{16}} \right\rfloor). \quad (1)$$

It will become clear that this bound is the best possible — when only n and m are given, no algorithm can be guaranteed to find a matching larger than that found by the greedy procedure.

The simpler but looser bound of $\gamma(G) \geq m/n$ is proved in [14] .

The bound in (1) can be considered alone, or in conjunction with augmenting path algorithms — the fastest known algorithms for finding a maximum matching.

All known worst-case time bounds for augmenting path algorithms are $\omega(m+n)$. It is traditional to use a hybrid algorithm: first, use the greedy procedure (or one like it) to find a matching M in $O(m+n)$ time; then, run an augmenting path algorithm with M as the initial matching. We will see that (1) supports the use of such hybrid algorithms. Intuitively, if the input graph is dense, then the greedy procedure finds a large matching, and the augmenting path algorithm needs only a few augmentation phases; if the input graph is sparse, then each augmentation phase is fast.

We can abstract the following technique for solving maximum cardinality matching problems: use one kind of method (perhaps the greedy procedure) for handling dense graphs, and another kind of method (perhaps an augmenting path algorithm) for handling other graphs. It may be interesting to investigate whether existing matching algorithms can be improved upon by explicitly using this technique.

An outline of the remainder of this paper is as follows. Section 2 contains definitions and notation. Section 3 gives a theorem due to Erdös and Gallai; our results are closely related to this theorem. Section 4 proves (1) and some variants of (1). (This is a new way of proving Erdös and Gallai's theorem.) Section 5 discusses the hybrid approach that uses the greedy procedure followed by an augmenting path algorithm.

2 Definitions and Notation

We consider the problem of finding a maximum matching in finite simple undirected unweighted possibly non-bipartite graphs.

Let $G = (V, E)$ be a graph. We use vw as an abbreviation for an edge $\{v, w\} \in E$. For $v \in V$, the graph $G - v$ is the graph with vertex set $V - v$, and edge set $\{xy \in E : x \neq v \text{ and } y \neq v\}$. The number of vertices and edges in G are respectively $n(G)$ and $m(G)$. The degree of a minimum degree vertex of G is denoted $\delta(G)$. An edge $vw \in E$, $\deg v \leq \deg w$, is called *semi-minimum* if $\deg v = \delta(G)$ and $\deg w$ is minimum over the degrees of v's neighbors. The matching number of G is denoted by $\nu(G)$, i.e., $\nu(G)$ is the size of a maximum matching for G. The *complete graph* on n vertices is K_n; its complement is the *heap* $\overline{K_n}$.

Function arguments are sometimes omitted when the context is clear, e.g., ν may be used instead of $\nu(G)$. The notation $a =^* b$ indicates that some algebraic manipulation showing that $a = b$ has been omitted so as to shorten the presentation.

3 A Related Theorem

This paper's analysis of the greedy matching procedure is closely related to the following theorem.

Theorem 1 (Erdös and Gallai, 1959). *The maximum number of edges in a simple graph of order n with a maximum matching of size k $(2 \leq 2k \leq n)$ is*

$$\begin{cases} \binom{k}{2} + k(n-k) & \text{if } k < \frac{2n-3}{5}, \\ \binom{2k+1}{2} & \text{if } \frac{2n-3}{5} \leq k < \frac{n}{2}, \\ \binom{2k}{2} & \text{if } k = \frac{n}{2}. \end{cases}$$

Erdös and Gallai's theorem can be proved in one direction by considering three graphs: the graph obtained by connecting every vertex of K_k to every vertex of $\overline{K}_{n-k}$; the graph K_{2k+1}; and the graph K_n. The edge counts appearing in the theorem are the number of edges in these graphs. Thus the indicated edge counts can be realized for a given value of k.

Proving the theorem in the other direction is more involved [4] ,[] . The proof of our main result (Theorem 3 in Section 4) is based on the greedy procedure. Since Theorem 3 implies Theorem 1, the proof of Theorem 3 is a new way of proving Erdös and Gallai's result.

Theorem 1 implies that if a graph has more than the indicated number of edges as a function of $k-1$, then the matching number of the graph is at least k. This fact, which is essentially equivalent to Theorem 1, is stated explicitly below.

Corollary 2. *Let G be a graph with n vertices and m edges, and let k be an integer such that*

$$m \geq \begin{cases} \binom{k-1}{2} + (k-1)(n-k+1) + 1 & \text{if } k \leq \frac{2n+2}{5}, \\ \binom{2k-1}{2} + 1 & \text{if } k \geq \frac{2n+2}{5}. \end{cases}$$

Then $\nu(G) \geq k$.

(In Corollary 2, when $k = \frac{2n+2}{5}$, both conditions apply; they are equivalent.)

The edge counting functions that appear in Corollary 2 are prominent in our analysis. In the remainder of this section we explicitly name these functions and establish some basic facts about them. Let

$$\begin{aligned} f(n,r) &= \tfrac{(r-1)(r-2)}{2} + (r-1)(n-r+1) + 1 &&= \big(r-1\big)\big(n-\tfrac{r}{2}\big) + 1, \text{ and} \\ g(n,r) &= \tfrac{(2r-1)(2r-2)}{2} + 1 &&= \big(r-1\big)\big(2r-1\big) + 1. \end{aligned} \tag{2}$$

For fixed n, the functions $f(r) = f(n,r)$ and $g(r) = g(n,r)$ are functions of a real argument r, with $f(r)$ increasing on $[-\infty, n]$, $g(r)$ increasing on $[1, \infty]$, and, if $n \geq 2$,

$$f(r) \geq g(r) \qquad \text{for } 1 \leq r \leq \tfrac{2n+2}{5}, \tag{3}$$

$$f(r) = g(r) \qquad \text{for } r = \tfrac{2n+2}{5}, \tag{4}$$

$$g(r) \geq f(r) \qquad \text{for } r \geq \tfrac{2n+2}{5}. \tag{5}$$

Define $b(n,r)$ by

$$b(n,r) = \begin{cases} f(n,r) & \text{if } r \leq \frac{2n+2}{5}, \\ g(n,r) & \text{if } r \geq \frac{2n+2}{5}. \end{cases} \tag{6}$$

For fixed $n \geq 2$, $b(r) = b(n,r)$ is increasing on $[2, n]$, since $f(r)$ and $g(r)$ are. Also,

$$b(n,r) = \max(f(n,r), g(n,r)) \qquad \text{when } n \geq 2 \text{ and } r \geq 1. \tag{7}$$

4 Performance Guarantees

The following performance guarantee for the greedy matching procedure is the main result of the paper.

Theorem 3. *Let G be a graph with n vertices and m edges, and let k be an integer such that*

$$m \geq \begin{cases} \binom{k-1}{2} + (k-1)(n-k+1) + 1 & \text{if } k \leq \frac{2n+2}{5}, \\ \binom{2k-1}{2} + 1 & \text{if } k \geq \frac{2n+2}{5}. \end{cases}$$

Then $\gamma(G) \geq k$.

Later in this section we will derive (1) from Theorem 3. We now establish two lemmas, and then use them to prove Theorem 3.

Lemma 4. *Let G be a graph with n vertices and m edges, and let $k \leq 2$ be an integer such that*

$$m \geq \begin{cases} \binom{k-1}{2} + (k-1)(n-k+1) + 1 & \text{if } k \leq \frac{2n+2}{5}, \\ \binom{2k-1}{2} + 1 & \text{if } k \geq \frac{2n+2}{5}. \end{cases}$$

Then $\gamma(G) \geq k$.

Proof. Assume the hypotheses of the lemma. The conclusion is trivial for $k \le 0$. The case $k = 1$ is also easy to verify. For $k = 2$, we first want to show that

$$m \ge n \ge 4. \tag{8}$$

If $k = 2 \le \frac{2n+2}{5}$, then $n \ge 4$, and

$$m \ge \binom{2-1}{2} + (2-1)(n-2+1) + 1 = n.$$

If $k = 2 \ge \frac{2n+2}{5}$, then $n \le 4$, and

$$m \ge \binom{2 \cdot 2 - 1}{2} + 1 = 4,$$

implying that $n = 4$. This establishes (8).

Next, let vw be a semi-minimum edge in G, with $\deg v = \delta(G)$, and $\deg w$ minimum over the degrees of v's neighbors. Suppose for the purpose of contradiction that every edge in G is incident on v, w, or both v and w. Since $m \ge n$ and $\deg w \le n-1$, there must exist some edge vx, $x \ne w$. The only possible vertices that x can be adjacent to are v and w, so $\deg x \le 2$. This implies that $\deg w \le 2$, by the minimality of $\deg w$ over the degrees of v's neighbors. It also implies that $\deg v \le 2$, by the minimality of $\deg v$. Therefore, $m \le \deg v + \deg w - 1 \le 3$, contradicting the fact that $m \ge 4$. It follows that G contains some edge that is incident on neither v nor w. Thus $G - v - w$ has at least one edge, and $\gamma(G) \ge 2 = k$. □

Lemma 5. *Let i, $j \ge i$, and k be positive integers. Then $b(j,k) \ge b(i,k)$.*

Proof.
Case I: $k \ge \frac{2j+2}{5}$. Then $k \ge \frac{2i+2}{5}$, and

$$b(j,k) = g(j,k) = (k-1)(2k-1) + 1 = g(i,k) = b(i,k).$$

Case II: $k \le \frac{2j+2}{5}$ and $k \le \frac{2i+2}{5}$. We have

$$b(j,k) = f(j,k) = (k-1)(j-k/2) + 1 \ge (k-1)(i-k/2) + 1 = f(i,k) = b(i,k).$$

Case III: $k \le \frac{2j+2}{5}$ and $k \ge \frac{2i+2}{5}$. Then

$$\begin{aligned} \frac{2j+2}{5} &\ge k, \\ j - \frac{k}{2} &\ge 2k - 1, \\ (k-1)(j-k/2) + 1 &\ge (k-1)(2k-1) + 1, \\ f(j,k) &\ge g(i,k), \\ b(j,k) &\ge b(i,k). \end{aligned}$$

□

Now we can prove Theorem 3.

Theorem 3 (Restatement and Proof). Let G be a graph with n vertices and m edges, and let k be an integer such that

$$m \geq \begin{cases} \binom{k-1}{2} + (k-1)(n-k+1) + 1 & \text{if } k \leq \frac{2n+2}{5}, \\ \binom{2k-1}{2} + 1 & \text{if } k \geq \frac{2n+2}{5}. \end{cases}$$

Then $\gamma(G) \geq k$.

Proof. We will use induction on k. Lemma 4 is the base of the induction ($k \leq 2$). Fix $k \geq 3$, and suppose that for all $t < k$, all graphs G with $m(G) \geq b(n(G), t)$ have $\gamma(G) \geq t$. Let H be a graph with

$$m(H) \geq b(n(H), k). \tag{9}$$

The graph H has at least one edge, because $b(n(H), k) \geq 1$. Let G be the graph formed from H by removing H's isolated vertices. We have $\gamma(G) = \gamma(H)$, and the proof will be complete if we can show that $\gamma(G) \geq k$.

Since $n(H) \geq n(G)$ and $m(H) = m(G)$, Lemma 5 and (9) imply that $m(G) \geq b(n(G), k)$. Setting $n = n(G)$ and $m = m(G)$, we have

$$m \geq b(n, k). \tag{10}$$

We will now discard H, and consider only G.

The inequalities $m \geq b(n, k)$ and $k \geq 3$ together with (6) imply that

$$3 \leq k \leq \frac{n}{2}. \tag{11}$$

Let vw be an arbitrary semi-minimum edge of G, with $\deg v = \delta = \delta(G)$, and w having minimum degree among v's neighbors. Set $\hat{m} = m(G - v - w) = m - \delta - \deg w + 1$, and $\hat{n} = n(G - v - w) = n - 2$. By induction, it suffices to show that

$$\hat{m} \geq b(n-2, k-1). \tag{12}$$

Let us derive another inequality. The δ vertices adjacent to v have degree $\deg w$ or more, and the $n - \delta$ vertices not adjacent to v have degree δ or more. Therefore,

$$\begin{aligned} m &\geq \frac{1}{2}\Big(\delta \deg w + (n-\delta)\delta\Big) \\ &= \frac{1}{2}\delta\Big(n - \delta + \deg w\Big), \text{ so} \\ \hat{m} &\geq \frac{1}{2}\delta\Big(n - \delta + \deg w\Big) - \delta - \deg w + 1 \\ &= \frac{1}{2}\delta\Big(n - \delta - 2\Big) + \deg w\Big(\frac{\delta}{2} - 1\Big) + 1. \end{aligned} \tag{13}$$

Now, several cases are considered. Some of the cases overlap. Cases where $k-1 \leq \frac{2(n-2)+2}{5}$ are indicated with an "*I*" prefix; the cases where $k-1 > \frac{2(n-2)+2}{5}$ are indicated by "*II*." The goal is to show that $\hat{m} \geq b(n-2, k-1)$ in all cases.

Case I: $k-1 \leq \frac{2(n-2)+2}{5}$. We want to show that $\hat{m} \geq \big(k-2\big)\big(n-2-\frac{k-1}{2}\big)+1$, because $b(n-2,k-1) = f(n-2,k-1) = \big(k-2\big)\big(n-2-\frac{k-1}{2}\big)+1$.

Case I.A: $\deg w \leq n+k-\delta-2$. Then

$$0 \geq (\delta + \deg w - 1) - n - k + 3. \tag{14}$$

Since $b(n,k) \geq f(n,k)$, by (10) we have

$$m \geq \Big(k-1\Big)\Big(n-\frac{k}{2}\Big)+1. \tag{15}$$

Adding (14) and (15) gives

$$\begin{aligned} m &\geq \Big(k-1\Big)\Big(n-\frac{k}{2}\Big)+1+\Big(\delta+\deg w-1\Big)-n-k+3, \text{ so} \\ \hat{m} &\geq \Big(k-1\Big)\Big(n-\frac{k}{2}\Big)-n-k+4 \\ &=^* \Big(k-2\Big)\Big(n-2-\frac{k-1}{2}\Big)+1. \end{aligned}$$

Case I.B: $\delta > n-\frac{k}{2}-1$. In this case, $\delta \geq n-\frac{k}{2}-\frac{1}{2}$. Reorganizing (13), and substituting for δ, we have the following.

$$\begin{aligned} \hat{m} &\geq \frac{1}{2}\delta\Big(n-\delta+\deg w-2\Big)-\deg w+1 \\ &\geq \frac{1}{2}\Big(n-\frac{k}{2}-\frac{1}{2}\Big)\Big(n+\Big(\deg w-\delta\Big)-2\Big)-\deg w+1 \\ &\geq \frac{1}{2}\Big(n-\frac{k}{2}-\frac{1}{2}\Big)\Big(n-2\Big)-\deg w+1 \\ &=^* \Big(\frac{n}{2}-2\Big)\Big(n-\frac{k}{2}-\frac{3}{2}\Big)+\Big(\frac{n}{2}-\frac{k}{2}-\frac{3}{2}\Big)+\Big(n-\deg w-1\Big)+1 \\ &\geq \Big(\frac{n}{2}-2\Big)\Big(n-2-\frac{k-1}{2}\Big)+0+0+1 \\ &\geq \Big(k-2\Big)\Big(n-2-\frac{k-1}{2}\Big)+1. \end{aligned}$$

Case I.C: $\deg w \geq n+k-\delta-1$ and $\delta \leq n-\frac{k}{2}-1$. Combining the inequalities $\deg w \geq n+k-\delta-1$ and $n-1 \geq \deg w$ yields

$$\delta \geq k. \tag{16}$$

This will be used later. Next, (13) gives

$$\begin{aligned}
\hat{m} &\geq \frac{1}{2}\delta\Big(n-\delta-2\Big)+\Big(n+k-\delta-1\Big)\Big(\frac{\delta}{2}-1\Big)+1 \\
&=^{*} \delta\Big(n-\delta+\frac{k}{2}-\frac{1}{2}\Big)-n-k+2 \\
&= y(\delta),
\end{aligned} \tag{17}$$

where $y(t)$ is the function

$$y(t)=t(n-t+\frac{k}{2}-\frac{1}{2})-n-k+2.$$

Now there are two sub-cases, according to the sign of

$$y\prime(t)=n-2t+\frac{k}{2}-\frac{1}{2}. \tag{18}$$

Case I.C1: $\delta \leq \frac{n}{2}+\frac{k}{4}-\frac{1}{4}$. By (18), $y\prime(t) \geq 0$ for all $t \leq \delta$. Thus, from (16), $y(\delta) \geq y(k)$, so by (17), we have $\hat{m} \geq y(k)$. Thus

$$\begin{aligned}
\hat{m} &\geq y(k) \\
&=^{*} \Big(k-2\Big)\Big(n-2-\frac{k-1}{2}\Big)+1+\Big(n-k-2\Big) \\
&\geq \Big(k-2\Big)\Big(n-2-\frac{k-1}{2}\Big)+1.
\end{aligned}$$

Case I.C2: $\frac{n}{2}+\frac{k}{4}-\frac{1}{4} \leq \delta \leq n-\frac{k}{2}-1$. By (18), $y\prime(t) \leq 0$ for $t \geq \delta$. Thus $y(\delta) \geq y(n-\frac{k}{2}-1)$. From (17),

$$\begin{aligned}
\hat{m} &\geq y(\delta) \\
&\geq y(n-\frac{k}{2}-1) \\
&=^{*} \Big(k-2\Big)\Big(n-2-\frac{k-1}{2}\Big)+1+\Big(\frac{3n}{2}-\frac{7k}{4}-\frac{10}{4}\Big) \\
&\geq \Big(k-2\Big)\Big(n-2-\frac{k-1}{2}\Big)+1.
\end{aligned}$$

Case II: $k-1 > \frac{2(n-2)+2}{5}$. (This is equivalent to $k \geq \frac{2n+4}{5}$.) We want to show that $\hat{m} \geq (k-2)(2k-3)+1$, because $b(n-2,k-1)=g(n-2,k-1)=(k-2)(2k-3)+1$.

Case II.A: $\delta+\deg w \leq 4k-4$. We have $k \geq \frac{2n+4}{5} \geq \frac{2n+2}{5}$, so $b(n,k)=g(n,k)$, and

$$\begin{aligned}
m &\geq g(n,k)\\
&= (k-1)(2k-1)+1\\
&= 2k^2-3k+2\\
&\geq 2k^2-7k+6+\delta+\deg w\\
&= (k-2)(2k-3)+1+(\delta+\deg w-1), \text{ so}\\
\hat{m} &\geq (k-2)(2k-3)+1.
\end{aligned}$$

Case II.B: $\delta+\deg w \geq 4k-3$ and $\delta \geq 2k-2$. As always, $n-1 \geq \deg w \geq \delta$ and $n \geq 2k$. By (13),

$$\begin{aligned}
\hat{m} &\geq \frac{1}{2}\delta(n+(\deg w-\delta)-2)-\deg w+1\\
&\geq \frac{1}{2}(2k-2)(n-2)-\deg w+1\\
&=^* (k-2)(n-2)+(-\deg w+1+n-2)\\
&\geq (k-2)(n-2)\\
&\geq (k-2)(2k-3)+1.
\end{aligned}$$

Case II.C: $\delta+\deg w \geq 4k-3$ and $\delta \leq 2k-3$. Substituting into (13) gives

$$\begin{aligned}
\hat{m} &\geq \frac{1}{2}\delta\Big(n-\delta-2\Big)+\Big(4k-3-\delta\Big)\Big(\frac{\delta}{2}-1\Big)+1\\
&=^* \frac{1}{2}\delta\Big(n-2\delta+4k-5\Big)-4k+\delta+4\\
&\geq \frac{1}{2}\delta\Big(2k-2\delta+4k-5\Big)-4k+\delta+4\\
&= \delta\Big(3k-\delta-\frac{5}{2}\Big)-4k+\delta+4\\
&= z(\delta), \qquad (19)
\end{aligned}$$

where $z(t)$ is the function

$$z(t)=t(3k-t-\frac{5}{2})-4k+t+4.$$

The derivative of $z(t)$ is $z\prime(t)=3k-2t-3/2$, so

$$z\prime(t)\leq 0 \text{ for } t\geq \frac{3k}{2}-\frac{3}{4}.$$

Now we will show that $\delta \geq \frac{3k}{2} - \frac{3}{4}$. We have

$$k \geq \frac{2n+4}{5},$$
$$\frac{5k}{2} \geq n+2,$$
$$4k-n-2 \geq \frac{3k}{2} - \frac{3}{4}.$$

By the definition of this case (II.C),

$$\delta \geq 4k-3-\deg w \geq 4k-n-2 \geq \frac{3k}{2} - \frac{3}{4}.$$

Thus $3k/2 - 3/4 \leq \delta \leq 2k-3$, so $z\prime(t) \leq 0$ for all $t \geq \delta$. Therefore

$$\hat{m} \geq z(\delta) \geq z(2k-3) =^{*} (k-3)(2k-3) + 5k - \frac{19}{2} \geq (k-3)(2k-3)+1.$$

□

Fix n and m, and let $k_{\max}$ be the maximum value of k that satisfies Theorem 3 for the given values of n and m. The theorem states that the greedy procedure finds a matching of size at least $k_{\max}$. Given only n and m, no algorithm can be guaranteed to find a matching of size greater than $k_{\max}$, because by Erdös and Gallai's theorem (Theorem 1) there exists a graph on n vertices and m edges having matching number as small as $k_{\max}$. When only n and m are given, then, the greedy procedure is optimal in the following sense: using $O(m+n)$ time and storage, it finds a matching of the maximum possible size that can be guaranteed for the given values of n and m.

We can reformulate Theorem 3 to have fewer variables by explicitly solving for $k_{\max}$. To do this we need two lemmas. These lemmas can be proved using the quadratic formula; the proofs are omitted. The lemmas and the two variable version of Theorem 3 are as follows.

Lemma 6. *Let n and m be integers, $n \geq 1$ and $0 \leq m \leq \binom{n}{2}$, and define k by*

$$k = \max\{i \mid i \textit{ is an integer on } [0,n] \textit{ such that } f(i) \leq m\}.$$

Then the quantity k is well defined, and

$$k = \left\lfloor n + \frac{1}{2} - \sqrt{n^2 - n - 2m + \frac{9}{4}} \right\rfloor.$$

Lemma 7. *Let n and m be integers, $n \geq 1$ and $1 \leq m \leq \binom{n}{2}$, and define k by*

$$k = \max\{i \mid i \textit{ is a positive integer such that } g(i) \leq m\}.$$

Then the quantity k is well defined, and

$$k = \left\lfloor \frac{3}{4} + \sqrt{\frac{m}{2} - \frac{7}{16}} \right\rfloor.$$

Corollary 8. *Let G be a graph with n vertices and m edges. Then*

$$\gamma(G) \geq \begin{cases} \left\lfloor n + \frac{1}{2} - \sqrt{n^2 - n - 2m + \frac{9}{4}} \right\rfloor & \textit{if } m \leq \frac{8n^2-14n+28}{25}, \\ \left\lfloor \frac{3}{4} + \sqrt{\frac{m}{2} - \frac{7}{16}} \right\rfloor & \textit{if } m \geq \frac{8n^2-14n+28}{25}. \end{cases}$$

Proof. The case $m = 0$ can be taken care of by a short analysis that we will omit. So, let G be a graph with $n \geq 2$ vertices and $m \geq 1$ edges. Some arithmetic yields

$$f\left(\frac{2n+2}{5}\right) = g\left(\frac{2n+2}{5}\right) = \frac{8n^2 - 14n + 28}{25}.$$

Now we consider two cases.

Case 1. Suppose that $m \leq \frac{8n^2-14n+28}{25} = f\left(\frac{2n+2}{5}\right)$. Define k by

$$k = \max\{i \mid i \text{ is an integer in } [0, n] \text{ such that } f(i) \leq m\}.$$

We have $f(k) \leq m \leq f\left(\frac{2n+2}{5}\right)$. Since $f(i)$ is an increasing function on $[0, n]$, it follows that $k \leq \frac{2n+2}{5}$. Therefore, $\gamma(G) \geq k$, by Theorem 3. Also, by Lemma 6,

$$k = \left\lfloor n + \frac{1}{2} - \sqrt{n^2 - n - 2m + \frac{9}{4}} \right\rfloor.$$

Case 2. Suppose that $m \geq \frac{8n^2-14n+28}{25} = g\left(\frac{2n+2}{5}\right)$. Define k by

$$k = \max\{i \mid i \text{ is a positive integer such that } g(i) \leq m\}. \tag{20}$$

By Lemma 7,

$$k = \left\lfloor \frac{3}{4} + \sqrt{\frac{m}{2} - \frac{7}{16}} \right\rfloor. \tag{21}$$

If $k \geq \frac{2n+2}{5}$, then we have the desired conclusion $\gamma(G) \geq k$, by Theorem 3 and the fact that $m \geq g(k)$. If $k \leq \frac{2n+2}{5}$, then since f is increasing in $[-\infty, n]$, we have

$$m \geq g\left(\frac{2n+2}{5}\right) = f\left(\frac{2n+2}{5}\right) \geq f(k).$$

By the case $k \leq \frac{2n+2}{5}$ in Theorem 3, $\gamma(G) \geq k$. □

Corollary 8 is sometimes more convenient in the following form.

Corollary 9. *Let G be a graph with n vertices and $m \geq 1$ edges. Then*

$$\gamma(G) \geq \min(\left\lfloor n + \frac{1}{2} - \sqrt{n^2 - n - 2m + \frac{9}{4}} \right\rfloor, \left\lfloor \frac{3}{4} + \sqrt{\frac{m}{2} - \frac{7}{16}} \right\rfloor).$$

Proof. Let G, n and m be as above. The following two implications can be shown:

$$n + \frac{1}{2} - \sqrt{n^2 - n - 2m + \frac{9}{4}} \leq \frac{3}{4} + \sqrt{\frac{m}{2} - \frac{7}{16}} \qquad \Rightarrow \quad m \leq \frac{8n^2 - 14n + 28}{25},$$

$$n + \frac{1}{2} - \sqrt{n^2 - n - 2m + \frac{9}{4}} \geq \frac{3}{4} + \sqrt{\frac{m}{2} - \frac{7}{16}} \qquad \Rightarrow \quad m \geq \frac{8n^2 - 14n + 28}{25}.$$

(The details are omitted). The conclusion follows from Corollary 8. □

5 The Hybrid Approach

The $O(m\sqrt{n})$ time general matching algorithms of Micali and Vazirani [9, 12] and Blum [3] operate in phases. Each phase uses $O(m)$ time, and there are at most $2\sqrt{\nu}$ phases. A matching M is maintained; initially, M has size, say, α, $0 \leq \alpha \leq \nu$. Each phase except the last enlarges M, so there are at most $\nu - \alpha + 1$ phases. A bound on the running time T_g of these general matching algorithms, therefore, is

$$T_g = O(m \cdot \min(2\sqrt{\nu}, \nu - \alpha)). \tag{22}$$

(This bound and others in this section are actually too low by a $O(m)$ term. For simplicity this is ignored in the remainder of this section.)

Now consider a hybrid algorithm that finds an initial matching in $O(m+n)$ time using the greedy procedure, and then uses one of the $O(m\sqrt{n})$ general matching algorithms. By Corollary 9, we have

$$\alpha \geq \gamma \geq \min(\left\lfloor n + 1/2 - \sqrt{n^2 - n - 2m + 9/4} \right\rfloor, \left\lfloor 3/4 + \sqrt{m/2 - 7/16} \right\rfloor). \tag{23}$$

Substituting into (22) yields a bound on the running time T_h of a hybrid algorithm:

$$T_h = O(m \cdot \min(2\sqrt{\nu}, \nu - \min(n - \sqrt{n^2 - n - 2m}, \sqrt{m/2}))). \tag{24}$$

This bound is tighter than $O(m\sqrt{\nu})$ for graphs that are dense relative to ν.

Let us see what happens when (24) is used to obtain a bound that is in terms of only n and m. Substituting $\nu \leq n/2$ yields

$$T_h = O(m \cdot \min(2\sqrt{n/2}, n/2 - \min(n - \sqrt{n^2 - n - 2m}, \sqrt{m/2}))).$$

The $n - \sqrt{n^2 - n - 2m}$ term turns out to be redundant; eliminating it gives

$$T_h = \mathrm{O}(m \cdot \min(2\sqrt{n/2}, n/2 - \sqrt{m/2})). \tag{25}$$

The right side of (25) reduces to $\mathrm{O}(m\sqrt{n})$ unless m is $\Theta(\binom{n}{2})$; thus (25) is almost no improvement over $\mathrm{O}(m\sqrt{n})$. In practice, however, it might be useful to bound the number of phases by using the non-asymptotic version of (25).

The bounds (24) and (25) imply a complementary relationship between the greedy procedure and general matching algorithms that use repeated $\mathrm{O}(m)$ time augmentation phases. For dense graphs, the greedy procedure finds a large matching, and few augmentation phases are needed; for sparse graphs, each augmentation phase is fast. Although hybrid algorithms has long been considered to give better performance than, say, using Micali and Vazirani's algorithm alone [10] , this specific complementary relationship seems not to have been generally known.

Since the $\mathrm{O}(m\sqrt{n})$ general matching algorithms are complicated [12] , a less complicated but possibly slower algorithm is sometimes preferred. For example, one might do just one augmentation per phase [10] . This can require as many as $n/2$ augmenting phases, as opposed to $\mathrm{O}(\sqrt{n})$ phases. In this case the greedy procedure's performance bounds take on a larger role. An analysis similar to the one earlier in this section shows that the running time for the resulting hybrid algorithm is

$$\mathrm{O}(m \cdot \max(\sqrt{n^2 - n - 2m} - n/2, n/2 - \sqrt{m/2})). \tag{26}$$

For dense graphs this is a significant improvement over $\mathrm{O}(mn)$.

6 Acknowledgement

The author thanks M. Krishnamoorthy and the anonymous referees for valuable comments.

References

[1] H. Alt, N. Blum, K. Mehlhorn, and M. Paul. Computing a maximum cardinality matching in a bipartite graph in time $\mathrm{O}(n^{1.5}\sqrt{m/\log n})$. *Information Processing Letters*, 37:237–240, February 1991.

[2] Claude Berge. *Graphs and Hypergraphs*. North-Holland Publishing Company, 1973.

[3] Norbert Blum. A new approach to maximum matching in general graphs. In *ICALP 90 Automata, Languages and Programming*, pages 586–597, Berlin, July 1990. Springer.

[4] Paul Erdös and T. Gallai. On maximal paths and circuits of graphs. *Acta Math. Acad. Sc. Hungar.*, 10:337–356, 1959.

[5] Tomás Feder and Rajeev Motwani. Clique partitions, graph compression, and speeding-up algorithms. In Baruch Awerbuch, editor, *Proceedings of the 23rd Annual ACM Symposium on the Theory of Computing*, pages 123–133, New Orleans, LA, May 1991. ACM Press.

[6] Zvi Galil. Efficient algorithms for finding maximum matchings in graphs. *Computing Surveys*, 18:23–38, 1986.

[7] John E. Hopcroft and Richard M. Karp. An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. *SIAM Journal on Computing*, 2(4), December 1973.

[8] L. Lovász and M. D. Plummer. *Matching Theory*, volume 121 of *North-Holland mathematics studies*. North-Holland, Amsterdam, 1986.

[9] S. Micali and V. V. Vazirani. An $O(\sqrt{|V|} \cdot |E|)$ algorithm for finding maximal matching in general graphs. In *Proceedings of the 21st Annual IEEE Symposium on Foundations of Computer Science*, pages 17–27. IEEE Computer Society Press, 1980.

[10] B. M. E. Moret and H. D. Shapiro. *Algorithms from P to NP*. The Benjamin/Cummings Publishing Company, Inc., 1991.

[11] Christos H. Papadimitriou and Kenneth Steiglitz. *Combinatorial optimization: algorithms and complexity*. Prentice Hall, 1982.

[12] Paul A. Peterson and Michael C. Loui. The general maximum matching algorithm of Micali and Vazirani. *Algorithmica*, 3:511–533, 1988.

[13] Andrew Shapira. Classes of graphs for which an $O(m+n)$ time greedy matching procedure does and does not find a maximum matching, in preparation.

[14] Andrew Shapira. Matchings, degrees, and $O(m+n)$ time procedures, in preparation.