SIAM J. DISCRETE MATH.
Vol. 15, No. 4, pp. 530–537

ROBUST MATCHINGS*

REFAEL HASSIN† AND SHLOMI RUBINSTEIN†

Abstract. We consider complete graphs with nonnegative edge weights. A p-matching is a set of p disjoint edges. We prove the existence of a maximal (with respect to inclusion) matching M that contains for any $p \leq |M|$ p edges whose total weight is at least $\frac{1}{\sqrt{2}}$ of the maximum weight of a p-matching. We use this property to approximate the metric maximum clustering problem with given cluster sizes.

Key words. robust matching, maximum clustering

AMS subject classifications. 05C70, 05C85

PII. S0895480198332156

1. Introduction.

Let $G = (V, E)$ be a complete graph with vertex set V such that $|V| = n$, edge set E, and edge weights $w(u, v) \geq 0$, $(u, v) \in E$. A p-matching is a set of p disjoint edges in a graph. A p-matching with $p = \lfloor \frac{n}{2} \rfloor$ is called perfect. A perfect matching M that for each $p = 1, \ldots, |M|$ contains p edges whose total weight is at least α times the maximum weight of a p-matching is said to be α-robust. We prove that G contains a $\frac{1}{\sqrt{2}}$-robust matching. On the other hand, there are graphs that do not contain an α-robust matching for any $\alpha > \frac{1}{\sqrt{2}}$.

In section 2, we generalize the robustness concept to independence systems. We prove that a greedy algorithm can be applied in some cases to form robust solutions. Our main theorem on robust matchings is proved in section 3, and we use it and a theorem from section 2 to approximate within a factor $\frac{1}{\sqrt{2}}$ the following problem: Given constants $c_1 \geq c_2 \geq \cdots \geq c_p$, find a p-matching M that maximizes $\sum_{i=1}^{p} c_i w_i$, where $w_1 \geq w_2 \geq \cdots \geq w_p$ are the edge weights in M.

In section 4, we use these results to approximate the NP-hard METRIC MAXIMUM CLUSTERING PROBLEM WITH GIVEN CLUSTER SIZES. The input for the problem is a complete graph with edge weights that satisfy the triangle inequality, and a set of cluster sizes. The goal is to find a partition of the vertex set, with part (or "cluster") sizes as required, that maximizes the total edge weight within the same cluster.

For $V' \subseteq V$ we denote by $E(V')$ the edge set of the subgraph induced by V'. For $E' \subset E$ we denote by $W(E')$ the total weight of edges in E'.

2. Robust independent sets.

An *independence system* is a pair $(E, \mathcal{F})$ consisting of a ground set E and a collection of *independent sets*, or, equivalently, *feasible solutions*, such that $F' \subset F \in \mathcal{F}$ implies $F' \in \mathcal{F}$. Let $w_e \geq 0$, $e \in E$, be weights attached to the elements of E. The problem of computing an independent set of maximum weight generalizes many interesting combinatorial optimization problems. Korte and Hausmann [1] analyzed the performance of the greedy algorithm for the above problem. The algorithm sorts the elements by weight and inserts them into the solution, starting with the heaviest one and excluding an element if its addition would generate a set not in $\mathcal{F}$. They proved the following theorem.

*Received by the editors January 5, 1998; accepted for publication (in revised form) July 1, 2002; published electronically September 10, 2002.

http://www.siam.org/journals/sidma/15-4/33215.html

†Department of Statistics and Operations Research, Tel-Aviv University, Tel-Aviv 69978, Israel (hassin@post.tau.ac.il, shlomiru@post.tau.ac.il).

THEOREM 2.1. *For any $E' \subseteq E$ define $l(E')$and $u(E')$ to be the smallest and largest cardinality, respectively, of a maximal (with respect to inclusion) independent set contained in E'. Let $r(E,\mathcal{F}) = min_{E' \subseteq E} \frac{l(E')}{u(E')}$. Then the greedy solution is an $r(E,\mathcal{F})$-approximation; that is, the value of the greedy solution is at least $r(E,\mathcal{F})$ times the optimal value.*

Now consider a game of the following type: You choose a maximal independent set in E. An adversary then selects a bound p on the allowed number of elements. Last, you output the p heaviest elements of your solution (or the solution itself if p is greater or equal to its cardinality). By the definition of an independence system, the output is independent. Your payoff is the ratio between the weight of your output and the maximum weight of an independent set whose cardinality is at most p. A solution is α-robust if it guarantees a payoff of at least α.

THEOREM 2.2. *The greedy solution is $r(E,\mathcal{F})$-robust.*

Proof. Define $(E,\mathcal{F}_p)$ to be the independence system in which $F \in \mathcal{F}_p$ if and only if $F \in \mathcal{F}$ and $|F| \leq p$. Let l_p and u_p denote the l and u values in the new system, respectively. Then for every $E' \subset E$

$$\frac{l_p(E')}{u_p(E')} = \frac{\min(l(E'),p)}{\min(u(E'),p)} \geq \frac{l(E')}{u(E')}$$

so that $r(E,\mathcal{F}_p) \geq r(E,\mathcal{F})$ and the claim follows from Theorem 2.1. □

The edges and matchings in a graph constitute an independence system for which $r = \frac{1}{2}$ [1]. It follows that the greedy solution is $\frac{1}{2}$-robust. We will reach this conclusion later in a different way but not before we obtain stronger results in the next section.

2.1. Weighted robustness. Let $c_1 \geq c_2 \geq \cdots \geq c_n \geq 0$ be given constants. For an independent set $F = \{e_1, \ldots, e_m\}$ with weights $w_1, w_2, \ldots, w_m$, define $C(F) = \sum_{j=1}^{m} c_j w_j$. Since we are interested in obtaining large values of $C(F)$, we will assume that the elements are numbered so that $w_1 \geq w_2 \geq \cdots \geq w_m$. Thus, $C(F)$ is well defined for any set F without explicitly specifying an order on its elements. We will also denote $F_p = \{e_1, \ldots, e_p\}$, $p = 1, \ldots, m$, and $F_p = F$ for $p > m$.

PROBLEM 2.3. *Compute an independent set $F \in \mathcal{F}$ of cardinality $|F| \leq p$ that maximizes $C(F_p)$.*

The following theorem was proved by Gerhard Woeginger.

THEOREM 2.4. *Problem 2.3 is NP-hard even when $\mathcal{F}$ is the set of matchings in a graph with edge set E (so that $F \subseteq E$ is in $\mathcal{F}$ if it consists of vertex-disjoint edges).*

Proof. The reduction is from the following NP-complete variant of 3-PARTITION.

Input: A positive integer t. $2n$ positive integers $a_1, a_2, \ldots, a_{2n}$ and n positive integers $b_1, b_2, \ldots, b_n$. These integers fulfill the equation $\sum_{i=1}^{2n} a_i + \sum_{i=1}^{n} b_i = nt$. Moreover, $a_i \leq t$ and $b_i \leq t$ holds for all i. These $3n$ integers are encoded in *unary*.

Question: Does there exist a permutation π of $\{1, \ldots, 2n\}$ such that for all $i = 1, \ldots, n$ we have $a_{\pi(2i-1)} + a_{\pi(2i)} + b_i = t$?

Consider an instance I of 3-PARTITION. For $i = 1, \ldots, n$ we define the cost coefficient $c_i = n^{b_i}$. Moreover, we construct from I an edge-weighted complete graph G on $2n$ vertices $v_1, \ldots, v_{2n}$. The weight $w(v_i, v_j)$ of the edge between vertices v_i and v_j equals

$$w(v_i, v_j) \;=\; n^{2t} - n^{a_i + a_j} \geq 0.$$

All these weights and costs are encoded in binary. Then the length of the encoding of every such value is bounded by $2t \log n$, which is polynomial in the size of the instance I (which is encoded in unary). We claim that G has a matching F of n independent edges with

$$C(F) \;\geq\; n^{2t} \sum_{i=1}^{n} n^{b_i} - n^{t+1} \;=:\; T^*$$

if and only if the instance I of 3-PARTITION has answer YES.

(If) Consider a solution π of 3-PARTITION. For $i = 1, \ldots, n$ the matching F matches vertex $v_{\pi(2i-1)}$ with vertex $v_{\pi(2i)}$ and assigns the cost coefficient c_i to this edge. Hence this edge contributes $n^{2t} n^{b_i} - n^t$ to the objective value, and the claim follows.

(Only if) Consider a matching F with the desired objective value. For $i = 1, \ldots, n$ let $(v_{\pi(2i-1)}, v_{\pi(2i)})$ denote the edge that is assigned to the cost coefficient $c_i = n^{b_i}$. We claim that π constitutes a solution to the 3-PARTITION instance. Otherwise, there exists an index i with $a_{\pi(2i-1)} + a_{\pi(2i)} + b_i \geq t + 1$. The contribution of this coefficient is then $\leq n^{b_i} n^{2t} - n^{t+1}$. The contribution of every other coefficient c_j is at most $n^{b_j}(n^{2t} - n)$. We can never reach an objective value T^*. □

THEOREM 2.5. *Let F and F' be independent sets. If F' is α-robust, then $C(F'_p) \geq \alpha C(F_p)$ for every $p = 1, 2, \ldots n$ and any constants $c_1 \geq c_2 \geq \cdots \geq c_n \geq 0$.*

Proof. Let $w_1 \geq w_2 \geq \cdots \geq w_p$ and $w'_1 \geq w'_2 \geq \cdots \geq w'_p$ be the weights of the elements of F_p and F'_p, respectively. (If $|F| < p$, then define $w_j = 0$ for $j > |F|$.) Then

$$\begin{aligned} C(F'_p) &= \sum_{i=1}^{p} c_i w'_i \\ &= \sum_{j=1}^{p-1} (c_j - c_{j+1}) \sum_{i=1}^{j} w'_i + c_p \sum_{i=1}^{p} w'_i \\ &= \sum_{j=1}^{p-1} (c_j - c_{j+1}) W(F'_j) + c_p W(F'_p) \\ &\geq \sum_{j=1}^{p-1} (c_j - c_{j+1}) \alpha W(F_j) + c_p \alpha W(F_p) \\ &= \alpha \sum_{i=1}^{p} c_i w_i = \alpha C(F_p). \qquad \square \end{aligned}$$

3. Robust matchings. A *matching* is a set of vertex-disjoint edges. The weight of a matching is the total weight of its edges. A *maximum matching* is a matching with maximum weight. A *p-matching* is a matching with p edges. We denote $m = \lfloor \frac{n}{2} \rfloor$, the maximum number of edges in a matching. An m-matching is said to be *perfect*. Note that this extends the common use of this concept to the case where the graph has an odd number of vertices.

An easy way to produce a maximum p-matching is as follows: Extend G by adding to it $n - 2p$ vertices. Each new vertex is connected to each original vertex by an edge with a "large" weight, say, twice the largest weight in G. Now compute a maximum perfect matching (with $n-p$ edges) in the extended graph. It will be composed of $n-2p$ heavy new edges and a maximum p-matching in G.

For a perfect matching M we define M_p to be the set of its p heaviest edges, $p=1,\ldots,m$. We denote by $M^{(p)}$ a maximum p-matching. A matching is α-robust if

$$W(M_p)\geq \alpha W(M^{(p)}) \qquad p=1,\ldots,m.$$

In this section, we show that for every graph there exists a $\frac{1}{\sqrt{2}}$-robust matching and that it can be constructed by a single application of a maximum matching algorithm. The following example shows that the value of $\frac{1}{\sqrt{2}}$ cannot be increased.

Consider a 4-vertex graph with weights $w(1,2)=w(3,4)=1,\ w(2,3)=\sqrt{2}$, and all other edges have zero weight. For this graph $W(M_1)=\sqrt{2}$ and $W(M_2)=2$. The graph has three perfect matchings and none is α-robust for $\alpha>\frac{1}{\sqrt{2}}$: The matchings $\{(1,2),(3,4)\}$ and $\{(2,3),(1,4)\}$ are $\frac{1}{\sqrt{2}}$-robust, and $\{(1,3),(2,4)\}$ is 0-robust.

THEOREM 3.1. *Let S be a maximum perfect matching with respect to the squared weights $w^2(e)$, $e\in E$. S is $\frac{1}{\sqrt{2}}$-robust.*

The rest of this section is devoted to proving Theorem 3.1. We will prove it by treating the squared edge weights as variables whose sizes are to be determined in order to form a contradiction to the theorem. We will prove that to achieve such a contradiction we may make several assumptions on these variables. Finally, these assumptions will lead to the conclusion that the claim is true.

Consider the set $S\cup M^{(p)}$. It consists of a collection of disjoint paths and cycles. A path may consist of a single edge or it alternates between S and $M^{(p)}$. Since S is perfect, the end edges of the path are from S except possibly one end of one path in the case of odd n (since in this case there is exactly one vertex that is not incident to an edge of S). A cycle alternates between S and $M^{(p)}$. We will construct from the edges of S a p-matching whose weight is at least $\frac{W(M^{(p)})}{\sqrt{2}}$. Since the weight of this matching is at most the weight of the p heaviest edges in S, this construction will prove the theorem.

We choose a p-matching from S as follows: Every edge in $S\cap M^{(p)}$ is chosen. All of the edges of S contained in a cycle of $S\cup M^{(p)}$ are chosen. From every nontrivial path (containing more than a single edge) of $S\cup M^{(p)}$ we choose all the edges that belong to S except for the lightest one. There is one exception to the last rule: If there is a path with only one end edge from S (this happens when n is odd), then we choose all of the S-edges of this path. The total number of edges selected is equal to $|M^{(p)}|=p$. It is sufficient to prove that the claimed bound on the ratio of the edge weights in S and in $M^{(p)}$ holds for every such path and cycle.

Consider a nontrivial path P with squared weights $x_1,y_1,x_2,y_2,\ldots,y_{r-1},x_r$, where the x values correspond to the edges of S and the y values correspond to the edges of $M^{(p)}$ in the order they appear on P.

We denote $x_{[i,j]}=\sum_{l=i}^{j}x_l$, and similarly $y_{[i,j]}=\sum_{l=i}^{j}y_l$. We are interested in subpaths $P_{i,j}$ of P consisting of the edges whose weights are $x_i,y_i,\ldots,y_{j-1},x_j$. Note that $P=P_{1,r}$. Since S is maximum with respect to the squared weights,

$$x_{[i,j]}\geq y_{[i,j-1]} \quad 1\leq i<j\leq r. \tag{3.1}$$

Let $x_{\min}=\min\{x_i\mid i=1,\ldots,r\}$. Our goal is to prove that the ratio of the total weight of the $r-1$ heaviest edges in $P\cap S$ to the weight of $P\cap M_k$ is at least $\frac{1}{\sqrt{2}}$;

that is,

$$Z = \frac{\sum_{i=1}^{r} \sqrt{x_i} - \sqrt{x_{\min}}}{\sum_{i=1}^{r-1} \sqrt{y_i}} \geq \frac{1}{\sqrt{2}}$$

for all x, y that satisfy (3.1).

We will prove that $Z \geq \frac{1}{\sqrt{2}}$ for every nontrivial path by induction on r. Note that the proof and induction hypothesis apply to any nontrivial path P in $S \cup M^{(p)}$, not just to maximal (with respect to inclusion) paths. A subpath is subject to additional constraints arising from longer subpaths that contain it, but these constraints may increase only the lower bound on Z for the subpath in question. We first prove this property for $r = 2$ and $r = 3$.

LEMMA 3.2. $Z \geq \frac{1}{\sqrt{2}}$ *when* $r = 2$.

Proof. Without loss of generality we can assume that $x_1 \geq x_2$; thus $Z = \frac{\sqrt{x_1}}{\sqrt{y_1}}$, and we look for its minimum subject to $x_1 + x_2 \geq y_1$ and $x_1 \geq x_2$. This minimum is obtained when $x_1 = x_2 = \frac{y_1}{2}$ and its value is $\frac{1}{\sqrt{2}}$. □

LEMMA 3.3. $Z \geq \frac{1}{\sqrt{2}}$ *when* $r = 3$.

Proof. In this case x_2 appears in every constraint of (3.1), and thus Z can be minimized with $x_2 \geq x_1, x_3$. Without loss of generality we assume that $x_3 = x_{\min}$ so that $Z = \frac{\sqrt{x_1}+\sqrt{x_2}}{\sqrt{y_1}+\sqrt{y_2}}$. We minimize Z subject to (3.1) and $x_2 \geq x_1 \geq x_3$. If $x_1 > x_{\min}$, then, by concavity of the square root function, Z can be reduced by decreasing x_1 and increasing x_2. Thus we assume $x_1 = x_3 = x_{\min}$. Again by concavity, Z can be decreased by increasing x_2 and simultaneously decreasing x_1 and x_3. This change is feasible if $x_1 + x_2 + x_3 > y_1 + y_2$. Thus we assume equality in this constraint, that is, $2x_1 + x_2 = y_1 + y_2$. We now show that the claimed bound holds even for the relaxed problem of minimizing Z subject to only $2x_1 + x_2 = y_1 + y_2$ and $x_2 \geq x_1 \geq 0$. Suppose first that y_1 and y_2 are fixed and we minimize over x_1 and x_2. The feasible set of solutions is a convex polyhedron and the objective function, $\sqrt{x_1} + \sqrt{x_2}$, is concave. Thus the minimum value is attained at an extreme point of the feasible set. There are two such points. In one, $x_1 = x_2 = \frac{y_1+y_2}{3}$ and $Z = \frac{2\sqrt{\frac{y_1+y_2}{3}}}{\sqrt{y_1}+\sqrt{y_2}}$. In the other one, $x_1 = 0$ and $x_2 = y_1 + y_2$, giving $Z = \frac{\sqrt{y_1+y_2}}{\sqrt{y_1}+\sqrt{y_2}}$, which is clearly smaller than the former case and attains its minimum over y_1 and y_2 when $y_1 = y_2$ and $Z = \frac{1}{\sqrt{2}}$. □

We now proceed to proving the general step of the induction for $r > 3$. Thus, we assume that the claim holds for smaller r values.

LEMMA 3.4. *We can assume that* $x_j > x_{min}$ $j = 2, \ldots, r-1$.

Proof. Suppose that $x_j = x_{\min}$ for some $j \in \{2, \ldots, r-1\}$. Then

$$Z = \frac{(\sum_{i=1}^{j} \sqrt{x_i} - \sqrt{x_{\min}}) + (\sum_{i=j}^{r} \sqrt{x_i} - \sqrt{x_{\min}})}{\sum_{i=1}^{j-1} \sqrt{y_i} + \sum_{i=j}^{r-1} \sqrt{y_i}}$$
$$\geq \min\left\{ \frac{\sum_{i=1}^{j} \sqrt{x_i} - \sqrt{x_{\min}}}{\sum_{i=1}^{j-1} \sqrt{y_i}}, \frac{\sum_{i=j}^{r} \sqrt{x_i} - \sqrt{x_{\min}}}{\sum_{i=j}^{r-1} \sqrt{y_i}} \right\}.$$

Since $x_j = \min\{x_i \mid i = 1, \ldots, j\} = \min\{x_i \mid i = j, \ldots, r\}$, it follows from the induction hypothesis that $Z \geq \frac{1}{\sqrt{2}}$. □

We call a subpath $P_{i,j}$ for which $x_{[i,j]} = y_{[i,j-1]}$ *tight.*

LEMMA 3.5. (i) *Let $i \le k \le j \le l$ such that $i < j$ and $k < l$. If $k < j$ and both $P_{i,j}$ and $P_{k,l}$ are tight, then so is $P_{k,j}$.* (ii) *Let $i < j < l$. If P_{ij} is tight, then $P_{j,l}$ is not.*

Proof. (i) By assumption, $x_{[i,j]} = y_{[i,j-1]}$ and $x_{[k,l]} = y_{[k,l-1]}$. Summing these equations we get

$$x_{[i,l]} + x_{[k,j]} = x_{[i,j]} + x_{[k,l]} = y_{[i,j-1]} + y_{[k,l-1]} = y_{[i,l-1]} + y_{[k,j-1]}.$$

Since $x_{[i,l]} \ge y_{[i,l-1]}$ and $x_{[k,j]} \ge y_{[k,j-1]}$ it follows that both of the latter relations satisfy equality and the respective subpaths are tight.

(ii) From the same equation with $j = k$ it follows that $x_j = 0$ and $1 < j < r$, in contrast to Lemma 3.4. $\square$

Suppose that $r \ge 3$. Let $1 < j < r$. We can assume that there exists a tight interval containing the edge e_j whose weight is x_j; otherwise, we reduce x_j until some subinterval containing e_j becomes tight, and this change reduces Z. Consider the intersection of all tight intervals containing $e_j \in S$. It follows from Lemma 3.5 that the intersection is a nontrivial tight subpath. Again by this lemma, the x values in this subpath share the same set of tight subpaths, and therefore we can assume that the sum of their squared roots is minimized subject to a single constraint on their sum. By concavity of the square root function, this objective is attained by setting all of these values to 0 except for a single one, say $x_k > 0$. From Lemma 3.4 and since $x_{\min} \ge 0$, it follows that $k \le 4$. For $r \le 3$, the claim has already been proved in Lemmas 3.2 and 3.3. Suppose that $r = 4$; then it must be that P_{12} and P_{34} are tight, and thus $x_1 = x_4 = y_2 = 0$, $y_1 = x_2$, and $y_3 = x_3$. In this case $Z = 1$, and this completes the proof for paths with two ends from S.

For a path with only one end edge from S we can assume that a fictitious S-edge of zero weight is added at that end. The set of constraints (3.1) then extends in a natural way, and the same proof holds.

Now suppose that there is a cycle C that contradicts the claim. We will show how to construct an instance consisting of a path that contradicts the claim. Since we have already proved that this is impossible, it will follow that such a cycle cannot exist. Specifically, let the cycle's edges have weights $x_1, y_1, \ldots, x_r, y_r$, in this order, with the x-weights corresponding to edges of S. Form a path by concatenating many repetitions of this sequence of weights. Last, add an x-edge at the end where it is missing, with a sufficiently large weight, such as $W(C \cap M^{(p)})$, so that (3.1) is satisfied. The path obtained this way will have (asymptotically, as the number of pasted copies increases) the same Z-value as C. This concludes the proof of Theorem 3.1.

3.1. More robustness results. Most of the proof of Theorem 3.1 is valid for any concave function, not just the square root function. Using this observation, the theorem can be generalized as follows. Let S_b be a maximum perfect matching with respect to the weights $w^b(e)$, $e \in E$, where $b \ge 1$ is a constant. Let $\beta = \frac{1}{b}$. The following lemmas and theorem follow easily by adapting the proofs of the respective results obtained in the previous subsection for $\beta = \frac{1}{2}$.

LEMMA 3.6. *$Z \ge \frac{1}{2^\beta}$ when $r = 2$.*

LEMMA 3.7. *$Z \ge \frac{1}{2^{1-\beta}}$ when $r = 3$.*

THEOREM 3.8. *S_b is $\min\{\frac{1}{2^\beta}, \frac{1}{2^{1-\beta}}\}$-robust.*

Maximum robustness is obtained when $\beta = \frac{1}{2}$, the case to which Theorem 3.1 applies. We note two interesting extreme cases. When $b = \beta = 1$, S_b is just a matching of maximum weight. When $b \to \infty$ and thus $\beta \to 0$, we get a *greedy* matching. Such a matching is obtained by sorting the edge weights in nonincreasing order, and then scanning the list and adding an edge to the matching if it is disjoint to the previously selected edges. In both cases the resulting bound is $\frac{1}{2}$, giving the following corollary.

COROLLARY 3.9. *Maximum and greedy matchings are* $\frac{1}{2}$*-robust.*

4. Clustering. In the METRIC MAXIMUM CLUSTERING PROBLEM WITH GIVEN CLUSTER SIZES, the goal is to partition the vertex set V into sets ("clusters") of given sizes so that the total weight of edges inside the clusters is maximized. Specifically, the input for the problem consists of a complete graph with edge weights satisfying the triangle inequality and cluster sizes $c_1 \geq c_2 \geq \cdots \geq c_p \geq 1$ such that $c_1 + \cdots + c_p = n$. We want to partition V into clusters of these sizes maximizing their total weight.

Let $d_j = \lfloor \frac{c_j}{2} \rfloor$, $D_j = d_1 + \cdots + d_j$ $j = 1, \ldots, p$, and $D_0 = 0$. We propose the following algorithm.

ALGORITHM 4.1.

1. *Compute a maximum matching S with respect to the squared weights. Let $S = \{(u_j, v_j)\ j = 1, \ldots, m\}$, where $w(u_j, v_j) \geq w(u_{j+1}, v_{j+1})$ $j = 1, \ldots, m-1$.*
2. *Set $V_i = \{u_j, v_j \mid j = D_{i-1} + 1, \ldots, D_i\}$ $i = 1, \ldots, p$.*
3. *For each i such that c_i is odd, add to V_i an arbitrary yet unassigned vertex.*

THEOREM 4.2. *Let opt and apx denote the solution values of the optimal and approximate solutions, respectively. Then*

$$apx \geq \frac{1}{2\sqrt{2}} opt.$$

Proof. Consider an optimal partition $O_1, \ldots, O_p$. Let M_i be a maximum matching in the subgraph induced by O_i, $i = 1, \ldots, p$. Denote the edge weights in M_i by $w_1^i \geq \cdots \geq w_{d_i}^i$.

Let $b_i = c_i - 1$ if c_i is even and $b_i = c_i$ if c_i is odd. The edges $E(O_i)$ can be covered by a set of $b_i \leq c_i$ disjoint matchings. Since M_i is a maximum matching in G_i it follows that $b_i W(M_i) \geq W(E(O_i))$ and therefore

$$opt \leq \sum_{i=1}^{p} c_i W(M_i).$$

Let $V_1, \ldots, V_p$ be the partition produced by Algorithm 4.1. Let $S_i = S \cap E(V_i)$. Consider a cluster V_i with vertices $u, v, q \in V_i$ such that $(u, v) \in S_i$. By the triangle inequality, $w(u, q) + w(v, q) \geq w(u, v)$.

Suppose that c_i is even. Sum this inequality over all $q \neq u, v \in V_i$; then sum again over $(u, v) \in S_i$. Note that every edge in $E(V_i) \setminus S_i$ is summed twice. Thus, every edge $(u, v) \in S_i$ contributes to the total weight of $E(V_i)$ in addition to its own weight also at least $\frac{1}{2}(c_i - 2)$ times its weight through the edges incident to it. Thus, $W(E(V_i)) \geq \frac{1}{2} c_i W(S_i)$.

Suppose now that c_i is odd. In this case V_i contains a vertex, say v_i, that was added to V_i in step 3 of the algorithm. In the summation, the weight of edges incident to v_i is used just once. Thus, each edge $(u, v) \in S_i$ contributes its weight $\frac{1}{2}(c_i - 3)$ times when summed over $V_i \backslash \{u, v, v_i\}$, once more through $w(u, v_i) + w(v, v_i)$, and once when it contributes its own weight. Thus, also in this case, $W(E(V_i)) \geq \frac{1}{2} c_i W(S_i)$.

By Theorem 2.5 and the assumption $c_1 \geq \cdots \geq c_p$,

$$\begin{aligned} apx &\geq \frac{1}{2} \sum_{i=1}^{p} c_i W(S_i) \\ &\geq \frac{1}{2\sqrt{2}} \sum_{i=1}^{p} c_i W(M_i) \\ &\geq \frac{1}{2\sqrt{2}} opt. \qquad \square \end{aligned}$$

Acknowledgment. We thank Gerhard Woeginger for permitting us to include his proof of Theorem 2.4.

REFERENCES

[1] B. Korte and D. Hausmann, *An analysis of the greedy heuristic for independence systems*, Ann. Discrete Math., 2 (1978), pp. 65–74.