UNIVERSITAS COMENIANA
ACTA MATHEMATICA UNIVERSITATIS COMENIANAE
L—LI—1987

WORST-CASE ANALYSIS OF A GREEDY HEURISTIC FOR OPTIMUM MATCHINGS IN GRAPHS

JÁN PLESNÍK, Bratislava

Our graph terminology is based on [1]. Given a graph G, its vertex and edge sets and their cardinalities are denoted $V(G)$, $E(G)$, n and m, respectively. It is supposed that every edge ij of G has been assigned a cost or weight or length c_{ij}, which is a real number. A *matching* of G is a subset M of the edges of G such that any two edges of M have no vertex in common. Clearly, $|M| \leqslant n/2$. If $|M| = n/2$, then the matching M is said to be *perfect*. The *cost* of a matching M is the sum of the costs of the edges in M. The *maximum matching problem* is to find a matching M of G with the maximum cost. Edmonds [3] has discovered a polynomial time algorithm for this problem. The most efficient implementations require $O(n^3)$ operations [4], [5]. Related problems are to find a *maximum perfect matching*, a *minimum matching* and a *minimum perfect matching*. All these problems are easily reducible each to other and thus algorithms of complexity $O(n^3)$ are available. Since the Edmonds algorithm is rather complicated, simpler heuristic methods are often suggested providing at least an approximation. In [6] two such heuristics are analyzed for the minimum perfect matching problem on graphs with positive costs satisfying the triangle inequality ($c_{ik} \leqslant c_{ij} + c_{jk}$).

In this paper, we study the known greedy method for some matching problems. As reductions among those problems do not conserve a degree of approximation, we consider each matching problem separately. In fact, different results are achieved. If c^* and c^H denote the optimum cost and the cost of an output of a heuristic H, respectively, then the ratio c^H/c^* is taken as a measure of approximation for a given example ($c^* \neq 0$). Thus a heuristic H can be evaluated according to the worst-case ratio. We show that a natural greedy heuristic provides a matching with cost at least 1/2 of the maximum cost. On the other hand, no small constant exists for the minimum perfect matching problem with positive costs fulfilling the triangle inequality.

The symbols c^* or c_* will denote the cost of an optimal solution in a maximization or minimization matching problem, respectively.

1. The maximum matching problem

At first no restrictions on G or c_{ij} are given. The problem is to find a matching M of G with the maximum cost. We suggest the following greedy heuristic.

GREEDY-1

Step 1: Delete all edges of G with negative or zero costs.
Order the remaining edges into a non-increasing sequence $S := (e^1, e^2, \ldots, e^q)$ according to their costs, , i. e. $c(e^1) \geqslant c(e^2) \geqslant \ldots \geqslant c(e^q)$.
Put $k := 1$, $M := \emptyset$, $t := 0$. No vertex of G is labelled (covered by an edge of M)

Step 2: If $k > q$ or $t \geqslant n - 1$, then go to Step 3.
If neither endvertex of the edge e^k is labelled, then $M := M \cup \{e^k\}$ and $t := t + 2$.
Put $k := k + 1$ and go to Step 1.

Step 3: Output M and stop.

The complexity of Step 1 is detemined by the complexity of a sorting procedure yielding the required sequence S. It is well known that this can be done in time $O(m \log m)$. As Step 2 can be performed in time $O(m)$, the overall complexity of GREEDY-1 is $O(m \log m)$ operations. If it is expressed as a function of n only, we obtain the complexity $O(n^2 \log n)$.

The maximum perfect matching problem is defined only for complete graphs with an even number of vertices. The costs of edges are any real numbers. Here GREEDY-1 is slightly modified such that in Step 1 no edges are deleted. This heuristic will be referred to as GREEDY-2.

Theorem 1. For maximum matching problems, the following assertions hold.

(i) The maximum matching problem in arbitrary graphs with arbitrary costs: For any output M of GREEDY-1 we have

$$c(M) \geqslant \frac{1}{2} c^*.$$

Moreover, coefficient $\frac{1}{2}$ in this estimation is best possible.

(ii) The maximum perfect matching problem with arbitrary costs: For any small real number $r > 0$ there is an example where GREEDY-2 outputs a perfect matching M with

$$c(M) < rc^*.$$

(iii) The maximum perfect matching problem with nonnegative costs: For any complete graph with an even number of vertices and with nonnegative edge costs, GREEDY-2 provides a perfect matching M with

$$c(M) \geqslant \frac{1}{2} c^* .$$

Moreover, coefficient $\frac{1}{2}$ in this inequality is best possible.

Proof. (i) Let M^* denote a maximum matching of G. We may suppose that all edges of M^* have positive costs. The proof is trivial for graphs with at most 2 vertices. We proceed by induction on n. Clearly, it suffices to deal with edges of positive costs.

Let us consider the first edge e^1 of the sequence S (e^1 has the maximum cost). Let the endvertices of e^1 be u and v. Since the graph $\hat{G}: = G - u - v$ has $n - 2$ vertices, the induction hypothesis can be applied. One can easily see that GREEDY-1, when applied to $\hat{G}$, can output the matching $\hat{M}: = M - \{e^1\}$. Thus, denoting by $\hat{M}^*$ a maximum matching of $\hat{G}$, we have

$$c(\hat{M}) \geqslant \frac{1}{2} c(\hat{M}^*) . \tag{1}$$

Trivially,

$$c(\hat{M}) = c(M) = c(e^1) . \tag{2}$$

To estimate $c(\hat{M}^*)$ by $c(M^*)$, we distinguish three cases.

Case 1: The edge $e^1 \in M^*$. As $M^* - \{e^1\}$ is a matching of $\hat{G}$, we have

$$c(\hat{M}^*) \geqslant c(M^*) - c(e^1) . \tag{3}$$

Thus using (2) and (3) in (1), we receive

$$c(M) - c(e^1) \geqslant \frac{1}{2} c(M^*) - \frac{1}{2} c(e^1) ,$$

or

$$c(M) \geqslant \frac{1}{2} \geqslant c(M^*) + \frac{1}{2} c(e^1) \geqslant \frac{1}{2} c(M^*) ,$$

as desired.

Case 2: The edge e^1 is adjacent to exactly one edge, say, e^j of M^* (clearly, $e^1 \notin M^*$). According to the choice of e^1, we have $c(e^1 \geqslant c(e^j)$ and thus

$$c(\hat{M}^*) \geqslant c(M^*) - c(e^j) \geqslant c(M^*) - c(e^1) . \tag{4}$$

Therefore using (2) and (4) in (1), we obtain

$$c(M) - c(e^1) \geqslant \frac{1}{2} c(M^*) - \frac{1}{2} c(e^1) ,$$

or

$$c(M) \geqslant \frac{1}{2}c(M^*) + \frac{1}{2}c(e^1) \geqslant \frac{1}{2}c(M^*),$$

as desired.

Case 3: The edge e^1 is adjacent to two edges, say, e^{j1} and e^{j2} of M^* (again $e^1 \notin M^*$). By the choice of e^1, we have

$$c(e^1) \geqslant \max\{c(e^{j1}), c(e^{j2})\} \geqslant \frac{1}{2}[c(e^{j1}) + c(e^{j2})].$$

As $M^* - \{e^{j1}, e^{j2}\}$ is a matching of $\hat{G}$, the last inequality yields

$$(5) \qquad c(\hat{M}^*) \geqslant c(M^*) - [c(e^{j1}) + c(e^{j2})] \geqslant c(M^*) - 2c(e^1).$$

Using (2) and (5) in (1), we obtain

$$c(M) - c(e^1) \geqslant \frac{1}{2}c(M^*) - c(e^1)$$

which gives the required inequality also in Case 3.

To complete the proof of Theorem 1 we consider the complete graph G with $V(G)\colon = \{1, 2, 3, 4\}$; all edge costs are equal to 1 except for $c_{34}\colon = 0$. Here GREEDY-1 can give $M = \{12\}$. Thus $c(M) = 1$ and $c^* = c(M^*) = 2$.

(ii) It is sufficient to consider the following example. Let $V(G)\colon = \{1,2,3,4\}$ and $c_{12}\colon = c_{13}\colon = c_{34}\colon = 1$, $c_{24}\colon = -1$, $c_{14}\colon = c_{23}\colon = 0$. We see that GREEDY-2 can output $M = \{13, 24\}$ with cost $c(M) = 1 - 1 = 0$, but $M^* = \{12, 34\}$ and thus $c^* = 2$.

(iii) It is left to the reader to verify that the proof of (i) can be used if GREEDY-1 is replaced by GREEDY-2 and other trivial changes are done. ■

2. Minimum matchings

Now, we are interested in matchings with the minimum costs. First, no assumptions on G or c_{ij} are put and a heuristic called GREEDY-3 will be used. It arises from GREEDY-1 by the following changes in Step 1: Delete all edges of G with positive or zero costs and order the remaining edges into a non-decreasing sequence $S\colon = (e^1, e^2, \ldots, e^q)$ (i.e. $c(e^1 \leqslant c(e^2) \leqslant \ldots \leqslant c((e^q))$.

For perfect matchings a further heuristic GREEDY-4 will be used. It differs from GREEDY-3 only in Step 1, where GREEDY-4 deletes no edges.

The minimum perfect matching problem is defined for complete graphs with an even number of vertices. This problem has the most applications if all the costs of edges are positive and fulfil the triangle inequality ($c_{ik} \leqslant c_{ij} + c_{jk}$).

(E. g. the Chinese postman problem and others [2, 4, 5].) It was also studied in [6] where two heuristics are given; they provide the worst-case ratio at most $O(\log n)$. We show that GREEDY-4 is much worse.

Theorem 2. For minimum matching problems, the following assertions hold.

(i)The minimum matching problem in arbitrary graphs with arbitrary costs: For ony output M of GREEDY-3 we have

$$c(M) \leqslant \frac{1}{2} c_* .$$

Moreover, coefficient $\frac{1}{2}$ in this inequality is best possible.

(ii) The minimum perfect matching problem with positive costs: For any real number $r > 1$, there exists an example such that GREEDY-4 outputs a perfect matching M with

$$c(M) \geqslant r c_* .$$

(iii) The minimum perfect matching problem with positive costs fulfilling the triangle inequality: For any real number $r \geqslant 1$, there exists an example such that GREEDY-4 outputs M with

$$c(M) \geqslant r c_* .$$

In particular, for any natural number k, there is an example with $n = 2^{k+1}$ such that GREEDY-4 yields a perfect matching M with

$$c(M) = \left[\frac{4}{3} n^{\log_2 3 - 1} - 1\right] c_* \doteq \left[\frac{4}{3} n^{0.58} - 1\right] c_* .$$

Proof. (i) If we change the sings of all costs to the opposite ones, then GREEDY-3 becomes GREEDY-1 and thus Theorem 2 (i) is an immediate corollary of Theorem 1 (i).

(ii) Put $V(G) := \{1, 2, 3, 4\}$, $c_{12} := 1$ and the other edges have $c_{ij} := \varepsilon$, where ε is a small real number with $0 < \varepsilon \leqslant 1/(2r - 1)$. One sees that GREEDY-4 gives $M = \{12, 34\}$ with $c(M) = 1 + \varepsilon$ while $c_* = 2\varepsilon$. The proof follows.

(iii) It is suficient to prove the second part of (iii). To be brief, we put all edges of constructed graphs into three classes and call them heavy, thin and dotted edges, respectively. First let us define graphs $H_k (k = 1, 2, \ldots)$ as follows (cf. *Fig. 1*). Each H_k is an alternating path consisting of heavy edges and thin edges, which alternate; the first and the last edges are heavy. H_k has $n := 2^{k+1}$ vertices. Each heavy edge has cost 1. The middle (thin) edge of H_1 has cost 1 and hence the path H_1 has cost $a_1 := 3$. In general, for every $k \geqslant 2$, the path H_k can be constructed by taking two copies of H_{k-1} and joining them by a new thin edge of cost 3^{k-1}. Therefore H_k has cost $a_k = 2a_{k-1} + 3^{k-1} = 3^k$. Further, a cycle F_k is constructed from H_k by joining the two ends of the path H_k by a new thin edge of cost 3^k (see *Fig. 1*). Finally, a complete graph G_k is produced from

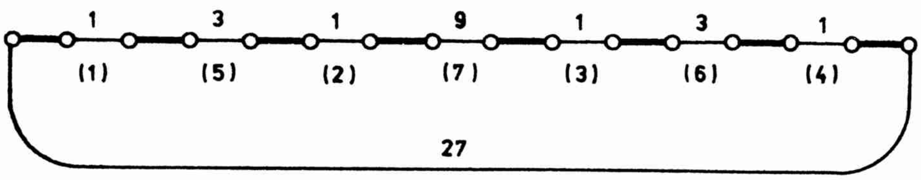

Fig. 1. The cycle F_3 consisting of the path H_3 and the thin edge of cost 27. The numbers in brackets show the order in which GREEDY-4 can choose the edges to M.

F_k by adding new dotted edges between any nonadjacent vertices of F_k. If the cost of each edge of F_k is considered as a length, then the distance $d(i, j)$ of two vertices i and j in F_k is assigned to the dotted edge ij. Thus all the costs in G_k are positive and fulfil the triangle inequality.

Since the least cost of an edge is 1, all the heavy edges realize a unique minimum perfect matching M_* of G_k and hence $c_*: = c(M_*) = 2^k$. Further, one sees that GREEDY-4 can choose a perfect matching M of G_k consisting of all the thin edges. Indeed, this is true if $k = 1$. If $k \geqslant 2$, then at first we choose (in any order) all thin edges of cost 1, then all thin edges of cost 3, then those of cost 3^2, etc.; the edge of cost 3^k is chosen as the last edge of M. One can easily verify that at that moment of taking a thin edge uv to M, every cheaper edge at u or v has the other end covered by an edge chosen before. A simple calculation gives $c(M) = 2.3^k - 2^k$. Thus $c(M)/c_* = 2(3/2)^k - 1 = (4/3)(3/2)^{\log_2 n} - 1 = = (4/3)n^{\log_2 3 - 1} - 1$, as desired. ■

Note that we have not found the worst-case ratio in (iii). Nevertheless, we conjecture that the above ratio is asymptotically the worst one.

REFERENCES

1. Behzad, M.—Chartrand, G.—Lesniak-Foster, L.: Graphs and digraphs, Prindle, Boston 1979.
2. Christofides, N.: Graph theory: an algorithmic approach, Academic, London 1975.
3. Edmonds, J.: Maximum matching and a polyhedron with 0, 1 vertices, J. Res. Nat. Bur. Standards 69 B (1965), 125—130.
4. Lawler, E. L.: Combinatorial optimization: networks and matroids, Holt, New York 1976.

5. Papadimitriou, C. H.—Steiglitz, K.: Combinatorial optimization: algorithms and complexity, Prentice-Hall, Englewood Cliffs, N. J., 1982.
6. Plaisted, D. A.: Heuristic matching for graphs satisfying the triangle inequality, J. Algorithms 5 (1984), 163—179.

Author's address:
Ján Plesník
Katedra numerických a optimalizačných metód MFF UK
Mlynská dolina
842 15 Bratislava

Received: 9. 10. 1985

SÚHRN

ANALÝZA NAJHORŠIEHO PRÍPADU JEDNEJ PAŽRAVEJ HEURISTIKY PRE OPTIMOVÉ PÁRENIA V GRAFOCH

Ján Plesník, Bratislava

Ukazuje sa, že jedna prirodzená pažravá heuristika dáva párenie, ktorého cena je aspoň 1/2 ceny maximového párenia. Avšak pre problém minimového úplného párenia v grafoch s kladnými cenami spĺňajúcimi trojuholníkovú nerovnosť takáto heuristika nezaručuje dobrú aproximáciu.

РЕЗЮМЕ

АНАЛИЗ НАИХУДЩЕГО СЛУЧАЯ У ОДНОЙ ЖАДНОЙ ЭВРИСТИКИ ДЛЯ ОПТИМАЛЬНОГО ПАРОСОЧЕТАНИЯ В ГРАФАХ

Ян Плесник, Братислава

Доказывается, что одна жадная эвристика находит паросочетание, цена (вес) которого не меньше, чем 1/2 цены максимального паросочетания. Но для проблемы минимального полного паросочетания в графах с положительными ценами, удовлетворяющими неравенству треугольника, такая эвристика не гарантирует хорошее приближение.