REPRINTED FROM

NAVAL RESEARCH LOGISTICS QUARTERLY

OFFICE OF NAVAL RESEARCH

NAVSO P-1278

VOL. 14, NO. 3 SEPTEMBER 1967

MAXIMUM MATCHING IN A CONVEX BIPARTITE GRAPH*

Fred Glover

School of Business Administration
University of Texas

ABSTRACT

A special matching problem arising in industry is shown to be solvable by an algorithm of the form: match objects a_i and b_j if they satisfy a local optimality criterion based on a ranking of currently unmatched objects. When no a_i and b_j remain that can be matched, the largest number of acceptable matches has been found.

We propose a "greedy" algorithm† to solve a maximum cardinality matching problem on a specially structured graph. The prototype for our problem (generalized below) is the following.‡ A manufacturer produces a fixed number of left and right halves of "rods" which must be fitted together.

Let w_i denote the weight of the i^{th} left half and v_j denote the weight of the j^{th} right half, where, say, $i = 1, \ldots, m$, and $j = 1, \ldots, n$. The problem is to assemble as many rods as possible subject to the restriction that the i^{th} left half can be matched to the j^{th} right half only if

(1) $$U_1 \geqq w_i - v_j \geqq L_1$$

(2) $$U_2 \geqq w_i + v_j \geqq L_2 ,$$

where U_1, L_1, U_2, and L_2 are given constants.

Because of the structure of this problem it is natural to index the w_i and v_j so that

$$w_1 \leqq w_2 \leqq \ldots \leqq w_m$$

and

$$v_1 \geqq v_2 \geqq \ldots \geqq v_n$$

*This research was supported by the Miller Institute of Basic Research in Sciences with the University of California.

†This rather suggestive terminology is due to Jack Edmonds [1].

‡I am indebted to Professor Ronald Shephard for posing this problem, which he encountered in an industrial setting.

Suppose m = n. Then we note that

$$U_2 \geqq w_i + v_i \geqq L_2$$

must hold for all i if there exists a matching that produces m rods whose left and right halves satisfy (2).

One may conjecture that the largest number of acceptable rods will be formed by matching w_1 with the first acceptable v_j (i.e., for which w_1 and v_j satisfy (1) and (2)), then matching w_2 with the first remaining acceptable v_j, and so on, skipping those w_i that can't permissibly match with any v_j. This conjecture is incorrect, however, and is also incorrect by instead attempting to match each w_i to the last v_j with which it can be acceptably paired.* In fact, if one permits an enumeration of alternatives, the attempt to match either the first or last of the (remaining) w_i by one of these criteria will still not guarantee that the maximum cardinality matching will be found.

Finally, if one tentatively matches w_i to v_i for all i, it is not in general possible to solve the problem by successive pair interchanges of the indices of the w_i while increasing the number of acceptable matchings at each step.

Before giving a rule that will produce the largest number of acceptable pairings, we generalize the problem as follows.

Let a_i, i = 1, . . . , m denote identifying labels assigned to a collection of m objects, and b_j, j = 1, . . . , n similarly denote identifying labels assigned to a second collection of objects (disjoint with the first). The two sets of objects may be represented by nodes in a bipartite graph, where the edge (a_i, b_j) connects the node a_i to the node b_j only if they can be acceptably matched. We assume that the a_i can be ordered to satisfy the following "convexity" property: for each j, if b_j connects a_h and a_k, $h < k$, then b_j connects every a_i such that $h \leqq i \leqq k$.

We will call a graph that satisfies the convexity property a convex bipartite graph. It is evident that the earlier matching problem satisfies this property by the indicated indexing of the w_i (regardless of the indexing of the v_j). There are also many convex bipartite graphs that cannot be interpreted to represent the restrictions (1) and (2).

The matching problem for an arbitrary (finite) convex bipartite graph G can be stated: find a maximum cardinality set of edges (a_i, b_j) that share no nodes in common. We now state our main result.

THEOREM: The following algorithm yields a maximum cardinality matching on a convex bipartite graph G.

1. Begin with i = 1, and repeat the following instructions until i is incremented to m + 1.

2. Let S be the set of all j such that (a_i, b_j) is an edge in G and b_j has not already been matched to some a_k for $k < i$.

3. If S is empty, increment i by 1 and return to 2. Otherwise,

4. Let $j(\max) = \text{Max}\{p: (a_p, b_j) \text{ is in } G\}$. Match b_v to a_i, where $v \in S$ and $v(\max) = \underset{j \in S}{\text{Min}} \{j(\max)\}$.

Then increment i by 1 and return to 2

*Beginning with the w_i that matches with the fewest v_j, rather than with w_1, does not repair the conjecture.

PROOF: Let G_r, $r = 1, \ldots, m$ represent the graph consisting of $a_1, \ldots, a_r$, all of the b_j, and all edges (a_i, b_j) for $i \leqq r$.

Correspondingly, E_r be the set of all edges (a_i, b_j) and their nodes, in the matching prescribed by the theorem applied to G_r. Finally, let E_r^* be the set of all (a_i, b_j), and their nodes, in a maximum cardinality matching on G_r, such that E_r^* has as many edges in common with E_r as possible. (If $E_r^* \neq E_r$, then it may not be unique.)

It is obvious that $E_1^* = E_1$. We assume $E_r^* = E_r$ for $r \leqq m - 1$ and prove it is true for $r = m$. More precisely, we assume $E_r^* = E_r$, $r \leqq m - 1$, when the graphs G_r contain any specified subset of the nodes b_j. (It is still clear that $E_1^* = E_1$.) Then if we show $E_m^* = E_m$ when each G_r contains all b_j, the same argument applies to any subset of the b_j, and the proof by induction will be complete.

First, note that $E_1 \subset E_2 \subset \ldots \subset E_m$ (relative to all the b_j). If $a_m \in E_m$, then $E_{m-1} \neq E_m$ and the induction assumption immediately implies $E_m = E_m^*$. Suppose now that $E_m^* \neq E_m$. By the foregoing remark and the induction assumption we have

(i) $a_m \in E_m^* - E_m$,

(ii) if $b_j \in E_{m-1}$, then (a_m, b_j) is not an edge in G,

(iii) there is a $b_k \in E_{m-1}$ such that $(a_m, b_k) \in E_m^*$.

Let E_r' denote the set of edges and their nodes in the matching prescribed by the theorem applied to G_r after removing b_k (defined in (iii)). Then by assumption, the edges of E_{m-1}' together with (a_m, b_k) must be a maximum matching on G_m. Identify the node a_p such that $(a_p, b_k) \in E_{m-1}$. Then, evidently $E_r' = E_r$ for all $r < p$. Note that $a_p \in E_p'$. For if not, E_r' will be the same as E_r excluding (a_p, b_k) and its nodes, for all $r \geqq p$, and hence E_{m-1}' will have one less edge than E_{m-1}, which is impossible. Thus there is a $b_q \in E_p'$ such that $(a_p, b_q) \in E_p'$. Also, by the rule of the theorem for selecting b_k to match to a_p in E_p, it follows that b_q and b_k both connect all a_r for $r \geqq p$. Thus, we can alternately let b_q match a_m in a maximum cardinality matching, and allow b_k to remain matched to a_p. Since $b_q \in E_{m-1}$ (by (ii)) we may assign b_q the role given b_k, and repeat the foregoing argument for a larger value of p. Eventually we must have $p \geqq m$, which is impossible. This completes the proof.

REMARK: We will call G "doubly convex bipartite" if it exhibits the convexity property both as defined earlier and also when the role of the a_i and b_j are interchanged (relative to an appropriate indexing of the b_j). Then, for such a G, the set S of the foregoing algorithm can be replaced by the $S' = \{j_1, j_2\}$, where $j_1 = \min_{j \in S} \{j\}$ and $j_2 = \max_{j \in S} \{j\}$.

PROOF: Assume on the contrary that neither j_1 nor j_2 qualifies to be v defined at step 4 of the algorithm. Then

$$v(\max) < s = \text{Min}\,\{j_1(\max), j_2(\max)\}.$$

Since the current value of i must satisfy $i < s$, and since b_{j_1} connects a_p for some $p \geqq s$, and b_{j_2} connects a_q for some $q \geqq s$, the convexity property implies that b_{j_1} and b_{j_2} both connect a_s. But also $j_1 < v < j_2$, and by the double convexity of G, a_s must connect b_v, contradicting $v(\max) < s$. This proves the remark.

The graph G corresponding to the problem of maximizing the number of rods whose halves satisfy (1) and (2) is doubly convex bipartite, as the indexing proposed for the w_i and v_j makes clear. By reference to (1) and (2) and the results above we observe that the choice of v for this problem can be determined from the following numerical computation:

$$\delta_v = \text{Min} \{\delta_{j_1}, \delta_{j_2}\}$$

where

$$\delta_j = \text{Min} \{U_1 + v_j, U_2 - v_j\}$$

and j_1 and j_2 are respectively the smallest and largest of the unmatched v_j indices such that

$$\text{Min} \{w_i - L_1, U_2 - w_i\} \geqq v_j \geqq \text{Max} \{w_i - U_1, L_2 - w_i\}.$$

REFERENCE

[1] Edmonds, Jack, "Optimum Branchings," National Bureau of Standards, Washington, D.C. (Presented under the title "Optimum Arborescences" at the International Seminar on Graph Theory and Its Applications, Rome, July 1966).