The Greedy Travelling Salesman's Problem

T. A. Jenkyns
Brock University
St. Catharines, Ont., Canada

ABSTRACT

The Travelling Salesman's Problem is to find a Hamilton path (or circuit) which has minimum total weight W_*, *in a graph (or digraph) with a non-negative weight on each edge. The Greedy Travelling Salesman's Problem is "How much larger than* W_* *can the total weight* G_* *of the solution obtained by the Greedy Algorithm be?". Using the theory of independence systems, it is shown that* G_*-W_* *may be as large as* $f(n,M,W_*)$ *where* n *is the number of vertices and* M *is the maximum edge-weight. The function* f *is determined for the several variations of the Travelling Salesman's Problem and the bound is shown to be best possible in each case.*

1. INTRODUCTION

Two problems concerning graphs which enjoy wide interest and many applications are "the Hamilton Path Problem": find (if one exists) a path through a given graph passing through each vertex exactly once; and "the Travelling Salesman's Problem" [the TSP]: when each edge of the graph carries a non-negative weight, find (if one exists) a Hamilton path whose total weight is as small as possible. Variations of these arise by requiring that the Hamilton paths be closed or not in graphs which are directed or not.

The well-known but naive method of trying to solve the TSP efficiently, dubbed by Jack Edmonds "the greedy algorithm", builds a 'solution' as follows: Order the edges from lightest to heaviest; beginning with the first edge, add the next edge, then add the next edge, then the next as long as no (non-Hamilton)

Networks, Vol. 9 (1979) 363-373 0028-3045/79/0009-0363$01.00

circuit is formed and no 3 edges which meet at a common vertex are taken. However, by taking many of the lightest edges one may be forced to take several of the heaviest edges and the total weight of a greedy solution may be much larger than that of an optimal solution.

The main results of this paper provide bounds for the total weight of a solution to the TSP obtained by the greedy algorithm. The next section obtains such bounds in the general context of the independence systems and subsequent sections give specific bounds for the several variations of the TSP.

2. OPTIMIZATION IN INDEPENDENCE SYSTEMS

An *independence system* [1] is an ordered pair (E,F) where E is a finite set and F is a non-empty family of subsets of E called *independent* sets such that every subset of an independent set is also independent. Independence systems will be defined on the edge sets of graphs so that the greedy algorithm generates a sequence of independent sets.

Let (E,F) be an independence system. For $A \subseteq E$, a *basis* of A is a maximal independent subset of A. Let two functions r_* and r^* be defined on the power set of E as follows:

$r_*(A) = \min\{|B| : B \text{ is a basis of } A\}$ and $r^*(A) = \max\{|B| : B \text{ is a basis of } A\}$. Then $0 \leq r_*(A) \leq r^*(A) \leq |A|$ for all $A \subseteq E$.

If for some number p, $r^*(A) \leq p \cdot r_*(A)$ for all $A \subseteq E$, then (E,F) will be called a *p-system*. It is obvious that every independence system is a p-system for some p such that $1 \leq p < |E|$. Define $p(E,F)$ to be the minimum value of p such that (E,F) is a p-system. In Sections 3 and 4 the value of the parameter p is determined for independence systems relevant to the TSP.

Let w be a non-negative function on E and let $w(A) = \sum_{a \in A} w(a)$ for $A \subseteq E$. The MAX problem for (E,F) with weight function w is the following: find an independent set B^* whose total weight is as large as possible; i.e. let $W^* = \max\{w(A) : A \in F\}$ and find $B^* \in F$ such that $w(B^*) = W^*$. The *(maximizing) greedy algorithm* is then

(1) index the N elements of E in *any* way so that $w(e_1) \geq w(e_2) \geq \ldots \geq w(e_N)$;

(2) set $I_0 = \phi$;

(3) given I_i for $i < N$, set $X = I_i \cup \{e_{i+1}\}$ and set

$$I_{i+1} = \begin{cases} X \text{ if } X \in F \\ I_i \text{ otherwise} \end{cases}$$

After N tests of independence one obtains a set I_N which is a basis of E and which we shall call a "greedy solution" of the MAX problem. The set I_N will depend on the indexing in setp (1). Let G* denote $w(I_N)$.

Edmonds has shown in [2] that every such I_N is optimal for all weight functions w if and only if (E,F) is a matroid [a 1-system]. In [3] and [4] this theorem is generalized to

Theorem 1: If $p(E,F) = p$ then $W^*/p \le G^* \le W^*$ for all weight functions on E and equality holds for certain weight functions.

Thus, the parameter p of an independence system determines the effectiveness of the greedy algorithm for solving the MAX problem (for all w).

The MIN problem for (E,F) with weight function w is the following: find a basis of E whose total weight is as small as possible; i.e. let $W_* = \min\{w(B): B \text{ is a basis of } E\}$ and find B_*, a basis of E, such that $w(B_*) = W_*$. The *minimizing greedy algorithm* is

(1') index of N elements of E in *any* way so that $w(e_1) \le w(e_2) \le \ldots \le w(e_N)$;

and obtain a greedy solution I_N by steps (2) and (3) of the maximizing greedy algorithm. Let G_* denote $w(I_N)$.

Theorem 2: Suppose $p(E,F) = p$, $M = \max\{w(e): e \in E\}$ and all bases of E have cardinality k. Then for all weight functions w

$$W_* \le G_* \le \left(\frac{1}{p}\right)W_* + \left(\frac{p-1}{p}\right)(k\cdot M),$$

and the upper bound is realized for certain weight functions.

Proof: Define a second weight function w' on E by $w'(e) = M - w(e)$. Then w' is non-negative. Suppose B_* is an optimal solution to the MIN problem with weight function w. Then B_* is an optimal solution to the MAX problem with weight function w'.

Suppose now that I_N is a greedy solution obtained by the minimizing greedy algorithm for w. Then I_N would also be obtained as a greedy solution for the MAX problem with weight function w' and by Theorem 1,

$$M|B_*| - w(B_*) = w'(B_*) \leq p\cdot w'(I_N) = p\{M|I_N| - w(I_N)\};$$

so
$$G_* = w(I_N) \leq \left(\frac{1}{p}\right)\{W_* - M.k\} + M\cdot k$$
$$= \left(\frac{1}{p}\right)W_* + \left(\frac{p-1}{p}\right)(k\cdot M).$$

Let A be a subset of E such that $p = r^*(A)/r_*(A)$; let J be a minimum cardinality basis of A, and K be a maximum cardinality basis of A. Define a weight function on E by

$$w(e) = \begin{cases} x \text{ if } e \in J \cup K \\ M \text{ otherwise, where } 0 < x << M. \end{cases}$$

Suppose step (1') of the greedy algorithm puts each element of J before any element of $E \setminus J$. Then $J \subseteq I_N$ and $K \subseteq B_*$ so $W_* = |K|x + \{k - |K|\}M$ and

$$G_* = |J|x + \{k - |J|\}M$$
$$= |J|\left(\frac{W_* - \{k - |K|\}M}{|K|}\right) + \{k - |J|\}M$$
$$= \left(\frac{1}{p}\right)W_* + \left(\frac{p-1}{p}\right)(k\cdot M).$$ □

Since $w(B) \leq k\cdot M$ for all bases of E, the upper bound given in Theorem 2 is a weighted average of W_* and the most pessimistic estimate of W_*. Furthermore, G_*-W_* may be as large as $\left(\frac{p-1}{p}\right)(k\cdot M-W_*)$ and may be forced to be arbitrarily large.

3. HAMILTON PATHS IN UNDIRECTED GRAPHS

Let G be an undirected graph with vertex set V and edge set E. Let $n = |V|$ and assume $n \geq 3$. For $J \subseteq E$ let: $G[J]$ denote the smallest subgraph of G with edge set J; $V_i(J) = \{\nu \in V: \nu \text{ is incident to } i \text{ edges in } J\}$; and $c(J)$ = the number of components of $G[J]$.

For $i = 1,2,\ldots,n$ let $C_i = \{X \subseteq E\colon G[X]$ is a circuit of length $i\}$. Let $C_0 = \{X \subseteq E\colon |X| = 3$ and $V_3(X) \neq \phi\}$. Define two independence systems on E as follows:

$F_C = \{X \subseteq E\colon X$ contains no member of $C_0 \cup C_1 \cup \ldots \cup C_{n-1}\}$

and $F = \{X \subseteq E\colon X$ contains no member of $C_0 \cup \quad \cup \ldots \cup C_n\}$.

If $G = K_n$ then X is a basis of E wrt F['wrt' will be used for 'with respect to'] if and only if G[X] is a Hamilton path; and X is a basis of E wrt F_C if and only if G[X] is a Hamilton circuit.

Lemma 1: $p(E,F) \leq p(E,F_C)$.

Proof: If $p(E,F) = 1$ then from the definitions $p(E,F) \leq p(E,F_C)$. Suppose then that $1 < p(E,F)$ and $A \subseteq E$, J is a minimum cardinality basis of A wrt F, K is a maximum cardinality basis of A wrt F and $p(E,F) = |K|/|J|$. Let $m = |J|$, then $m+1 < |K|+1 \leq n$. For all $a \in A \backslash J$, $J \cup \{a\}$ contains an element of $C_0 \cup C_1 \cup \ldots \cup C_{m+1}$ so $J \cup \{a\} \notin F_C$. Thus J is a basis of A wrt F_C. Because $K \in F \subseteq F_C$, K is contained in a basis K' of A wrt F_C. Then

$$p(E,F_C) \geq \frac{|K'|}{|J|} \geq \frac{|K|}{|J|} = p(E,F). \qquad \square$$

Lemma 2: Let H be any subset of edges of G, let J be any basis of H wrt F_C and let K be any subset of H. Then if $K \in F_C$, $|K| \leq 2|J| - c(J)$.

Proof: Let $A = \{k \in K\colon k$ is incident to a vertex in $V_2(J)\}$ and let $B = \{k \in K \backslash A\colon k$ joins the ends of a component of $G[J]\}$. Clearly $K \cap J \subseteq A \cup B$. Because J is a basis of H, for any $k \in K \backslash J$, $X = J \cup \{k\} \notin F_C$, so G[X] has a vertex of degree 3 or contains a non-Hamilton circuit. Thus $K \backslash J \subseteq A \cup B$. Because $K \in F_C$, at most two members of K are incident to any vertex of G, and at most one member of K joins two given vertices of G. Thus

$$|K| \leq |A| + |B| \leq 2|V_2(J)| + c(J).$$

If $G[J]$ is a Hamilton circuit, $|K| \leq |J| = |V_2(J)| = n \geq 3$ and $c(J) = 1$. If $G[J]$ is not a Hamilton circuit, $|V_2(J)| = |J| - c(J)$. In either case, $|K| \leq 2|J| - c(J)$. □

Theorem 3: $p(E,\mathcal{F}_C) \leq 2 - \frac{1}{a}$ *where* $a = \left[\frac{n+1}{2}\right]$ *and equality holds when* $G = K_n$.

Proof: Let H be any non-empty subset of edges of G, let J be a minimum cardinality basis of H wrt $\mathcal{F}_C$ and let K be a maximum cardinality basis of H wrt $\mathcal{F}_C$. If $|J| > a$ then $|K| \leq n \leq 2a$ so

$$\frac{|K|}{|J|} \leq \frac{2a}{a+1} = 2 - \frac{2}{a+1} \leq 2 - \frac{1}{a} .$$

If $|J| \leq a$ then $1 \leq c(J)$ so from Lemma 2

$$\frac{|K|}{|J|} \leq 2 - \frac{c(J)}{|J|} \leq 2 - \frac{1}{a} .$$

Thus $\frac{r^*(H)}{r_*(H)} \leq 2 - \frac{1}{a}$ and $(E,\mathcal{F}_C)$ is a $\left(2 - \frac{1}{a}\right)$-system.

For $1 \leq i \neq j \leq n$ let $[i,j]$ denote the edge of K_n joining i and j. Then let $J = \{[i,i+1]: i = 1,2,\ldots,a\}$ and $I = \{[i,a+i]: i = 1,2,\ldots,a-1\} \cup \{[i+1,a+i]: i = 1,2,\ldots,a-1\}$.

Suppose n is odd; i.e., $n = 2a-1$. Let $I_1 = I \cup \{[1,a]\}$. Then both J and I_1 are bases of $H_1 = J \cup I_1$ wrt $\mathcal{F}_C$. Hence

$$\frac{r^*(H_1)}{r_*(H_1)} \geq \frac{|I_1|}{|J|} = \frac{2a-1}{a} = 2 - \frac{1}{a} .$$

Suppose n is even; i.e., $n = 2a$. Let $I_2 = I \cup \{[a,2a]\}$. Then both J and I_2 are bases of $H_2 = J \cup I_2$ wrt $\mathcal{F}_C$. Hence

$$\frac{r^*(H_2)}{r_*(H_2)} \geq \frac{|I_2|}{|J|} = \frac{2a-1}{a} = 2 - \frac{1}{a} .$$

Thus if G contains a subgraph isomorphic to $K_{2a-1}[H_1]$ or $K_{2a}[H_2]$, $p(E,\mathcal{F}) = 2 - \frac{1}{a}$. □

The bound on the worst-case behaviour of the greedy algorithm for solving the TSP, when $G = K_n$ and we are searching for Hamilton *circuits*, is given by Theorem 2 and may be realized. In general, p is almost 2, so G_*-W_* may be almost as large as $\frac{1}{2}(nM - W_*)$; in particular, if $n = 25$ then $p = 25/13$ and G_*-W_* may be as large as $\frac{12}{25}(25M - W_*)$.

Lemma 3: Let H be any subset of edges of G, let J be any basis of H wrt F, and let K be any subset of H. Then if $K \in F$, $|K| \leq 2|J| - c(H)$.

The proof of Lemma 3 is the same as that of Lemma 2 without the possibility that G[J] might be a Hamilton circuit.

Theorem 4: $p(E,F) \leq 2 - \frac{1}{a}$ where $a = \left[\frac{n}{2}\right]$ and equality holds when $G = K_n$.

Proof: Let H be any non-empty subset of edges of G, let J be any minimum cardinality basis of W wrt F, and let K be a maximum cardinality basis of W wrt F. If $|J| > a$ then $|K| \leq n-1 \leq 2a$ so

$$\frac{|K|}{|J|} \leq \frac{2a}{a+1} = 2 - \frac{2}{a+1} \leq 2 - \frac{1}{a}.$$

If $|J| \leq a$ then $1 \leq c(J)$ so from Lemma 3

$$\frac{|K|}{|J|} \leq 2 - \frac{c(J)}{|J|} \leq 2 - \frac{1}{a}.$$

Thus $\frac{r^*(H)}{r_*(H)} \leq 2 - \frac{1}{a}$ and (E,F) is a $\left(2 - \frac{1}{a}\right)$-system.

To show this bound may be obtained, again consider $G' = K_{2a}[H_2]$ and sets J and I_2 defined in the proof of Theorem 3. Suppose n is even; i.e., $n = 2a$. Then both J and I_2 are bases of H_2 wrt F. Hence

$$\frac{r^*(H_2)}{r_*(H_2)} \geq \frac{|I_2|}{|J|} = \frac{2a-1}{a} = 2 - \frac{1}{a}.$$

Thus if C contains a subgraph isomorphic to G' [in particular, if $G = K_{2a}$ or $G = K_{2a+1}$], then $p(E,F) = 2 - \frac{1}{a}$. □

The bound on the worst-case behavior of the greedy algorithm for solving the TSP, when $G = K_n$ and we are searching for Hamilton *paths*, is given by Theorem 2 and may also be realized; in particular, if $n = 25$ then $a = 12$, $p = 23/12$ and G_*-W_* may be as large as $\frac{11}{23}(24M - W_*)$.

4. HAMILTON PATHS IN DIRECTED GRAPHS

Let D be a directed graph with vertex set V and edge set E. Let $n = |V|$ and assume $n \geq 4$. Let D_n denote the digraph on n vertices where there is an edge from every vertex to every other vertex.

For $i = 1,2,\ldots,n$ let $C_i' = \{X \subseteq E\colon D[X]$ is a dicircuit of length $i\}$. Let $C_0' = \{X \subseteq E\colon |X| = 2$ and both edges in X are direct *to* the same vertex$\}$ and let $C_0'' = \{X \subseteq E\colon |X| = 2$ and both edges in X are directed *from* the same vertex$\}$. Define two independence systems on E as follows:

$\mathcal{F}_C = \{X \subseteq E\colon X$ contains no member of $C_0'' \cup C_0' \cup C_1' \cup \ldots \cup C_{n-1}'\}$

and $\mathcal{F} = \{X \subseteq E\colon X$ contains no member of $C_0'' \cup C_0' \cup C_1' \cup \ldots \cup C_n'\}$.

If $D = D_n$ then X is a basis of E wrt $\mathcal{F}$ if and only if D[X] is a Hamilton dipath; and X is a basis of E wrt $\mathcal{F}_C$ if and only if D[X] is a Hamilton dicircuit. As was the case for undirected graphs, here too, $p(E,\mathcal{F}) \leq p(E,\mathcal{F}_C)$.

Lemma 4: Let H be any subset of edges of D, let J be any basis of H wrt $\mathcal{F}_C$ and K be any subset of H. Then if $K \in \mathcal{F}_C$, $|K| \leq 3|J|$.

Proof: Let $A' = \{k \in K\colon$ for some $j \in J$, k and j are directed *to* the same vertex$\}$, let $A'' = \{k \in K\colon$ for some $j \in J$, k and j are directed *from* the same vertex$\}$, and let $B' = \{k \quad K\colon$ k joins the ends of a component of D[J] forming a non-Hamilton dicircuit$\}$. Clearly $K \cap J \subseteq A'$. Because J is a basis of H, for any $k \in K \backslash J$, $X = J \cup \{k\} \notin \mathcal{F}_C$, so $k \in A' \cup A'' \cup B'$. Because J and $K \in \mathcal{F}_C$, no two edges in J or in K are directed *to* or *from* the same vertex. Thus

$$|K| \leq |A'| + |A''| + |B'| \leq |J| + |J| + c(J) \leq 3|J|. \qquad \square$$

Unlike the undirected case, where the value of p dependent on $n = |V|$, we have

Theorem 5: $p(E,F_C) \leq 3$, *and equality holds when* $D = D_n$ *for* $n \geq 4$.

Proof: From Lemma 4, for all non-empty sets $H \subseteq E$, $\frac{r^*(H)}{r_*(H)} \leq 3$ so (E,F_C) is a 3-system.

For $1 \leq i \neq j \leq n$ let (i,j) denote the edge of D_n *from* vertex i *to* vertex j. Then let $J = \{(2,3)\}$ and $I = \{(4,3), (3,2), (2,1)\}$. Both J and I are bases of $H = J \cup I$ wrt F_C; hence if D contains a subgraph isomorphic to $D' = D_4[H]$ then $p(E,F_C) = 3$. □

Theorem 6: $p(E,F) \leq 3$ *and equality holds when* $D = D_n$ *for* $n \geq 4$.

Proof: Since $p(E,F) \leq p(E,F_C)$, from Theorem 5, (E,F) is a 3-system. Furthermore, J and I as defined in the proof of Theorem 5 are bases of H wrt F. Thus if D contains a subgraph isomorphic to D' then $p(E,F) = 3$. □

Thus if we are searching for Hamilton dicircuits in D_n or for Hamilton dipaths in D_{n+1}, G_*-W_* may be as large as $\frac{2}{3}(nM - W_*)$.

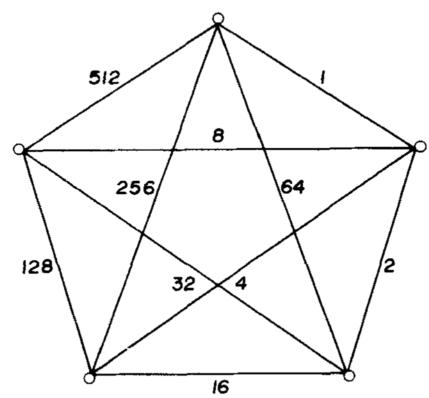
5. CONCLUDING OBSERVATIONS

In this section we make a few remarks concerning the two questions: When is the greedy solution optimal? and when is it "pessimal"?

As was observed in Section 2 the greedy solution I_N is optimal for all weight functions w if and only if (E,F) is matroid. However, for certain specific weight functions it may also be optimal. For instance, if every basis of E has the same weight, as heppens when w is constant, $W_*=G_*$. Note that $W_* = \left(\frac{1}{p}\right) W_* + \left(\frac{p-1}{p}\right)$ (kM) if and only if p=1 or W_*=kM. When all edge weights are distinct the greedy solution is unique, and for $p > 1$, $G_* < \left(\frac{1}{p}\right) W_* + \left(\frac{p-1}{p}\right) kM$.

In the construction where $G_* = \left(\frac{1}{p}\right) W_* + \left(\frac{p-1}{p}\right) (kM)$ the edge-weights took only two values x and M where $0 < x \ll M$, and some ordering of the edges weighted x gave the worst greedy solution while another gave an optimal solution. Thus, even when the number of different edge-weights is small, it might be feasible and worthwhile to consider all possible orderings of the edges, except those weighted M, in step (1') of the greedy algorithm and select the "best" greedy solution.

Finally consider the weighted K_5 below.



The edge weights are distinct powers of 2 so all Hamilton circuits have different total weights. These range from $W_* = 217$ to $W^* = 806$, and $G_* = 659$. Here $W_* \ll G_* \ll \left(\frac{3}{5}\right) W_* + \left(\frac{2}{5}\right) 5M = 1154.2$. This example shows that 5M cannot be replaced by W^* nor by the sum of the weights of the 5 heaviest edges (992) in the upper bound formula given in Theorem 2. This example also shows that the *best* greedy solution may still be far from optimal so even a clever, greedy, travelling salesman has a problem.

ACKNOWLEDGEMENT

The author gratefully acknowledges financial support from the National Research Council of Canada.

REFERENCES

1. J. Edmonds, "Minimum Partition of a Matroid into Independent Subsets", *J. Res. Nat. Bur. Standards*, 69B (1965), pp. 67-72.

2. J. Edmonds, "Matroids and the greedy algorithm" - mimeographed notes (Int'l Symposium on Math. Programming, Princeton, 1966.)

3. T. A. Jenkyns, "The Efficacy of the 'greedy' Algorithm", *Proc. 7th SE Conference Combinatorics, Graph Theory, and Computing*, 1976, pp. 341-350.

4. B. Korte and D. Hausmann, An Analysis of the Greedy Heuristic for Independence Systems, Working paper, Institut fur Okonometric und Operations Research, University Bonn (1976).

Received October 1977; Revised July 1978