Analysis of heuristics for finding a maximum weight planar subgraph

M.E. DYER
Department of Mathematics and Statistics, Teesside Polytechnic, Middlesbrough, Cleveland TS1 3BA, United Kingdom

L.R. FOULDS
Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611, U.S.A.

A.M. FRIEZE
Department of Computer Science and Statistics Queen Mary College, University of London, London, E1 4NS, United Kingdom

Abstract. The problem of identifying a planar subgraph of maximum total weight in an edge-weighted graph has application to the layout of facilities in a production system and elsewhere in industrial engineering.

This problem is NP-hard, and so we confine our attention to polynomial-time heuristics. In this paper we analyse the performance of some heuristics for this problem.

Keywords: Graphs, heuristics, computational analysis, probability, facilities, location

Received September 1983; revised March 1984

North-Holland
European Journal of Operational Research 20 (1985) 102–114

1. Introduction

In this paper we are concerned with a specific graph theoretic problem. The reader is referred to Harary [9] for an explanation of the graph-theoretic terminology used here. Consider a graph $G = (V, E)$ with vertex set V and an edge weighting function $w: E \to \mathbb{R}$ which assigns a real number to each edge in E. For each subset S of E define

$$w(S) = \sum_{e \in E} w(e).$$

Then the problem is to find a planar subgraph of G which has maximum total weight. That is, to find $E' \subseteq E$ such that the graph (V, E') is planar and $w(E')$ is as large as possible. We call this the planar subgraph problem. We will now explain why this problem is of interest in industrial engineering.

1.1. An application in industrial engineering

The facilities layout problem of industrial engineering is concerned with the design of a system of physical facilities such as machines on a factory floor. For further details see, for example, [1,2,10,14,15]. The question of layout design manifests itself in many other applications such as the design of hospitals [4], universities [3] and office blocks [16]. One of the important subproblems involved is the question of which facilities should be located adjacently. It is common to begin the layout process by defining a *relationship chart*. Each entry of this chart (called a *closeness rating*) defines the desirability (measured by an arbitrary real number) of locating a pair of facilities adjacently. The problem is to find a layout which maximizes the sum of the closeness ratings corresponding to adjacent pairs of facilities. This subproblem can be modelled in graph theoretic terms. Consider a weighted graph $G = (V, E)$ with an edge weighting function w. Each vertex in V represents a facility of the system to be laid out and each edge in E corresponds to the possibility of locating the facilities identified by its endpoints adjacent to each other. Each edge weight is the appropriate closeness rating. Each subgraph of G characterizes the adjacency structure of a possible layout. Usually, the facilities are to be laid out on a plane surface such as a building site or a factory floor. When this is the case, it is necessary that any subgraph of G characterizing the adjacency structure of a *feasible* layout must be *planar*. Thus the finding of the adjacency structure of the most desirable layout (ie the one which maximizes the sum of closeness ratings for adjacent facilities) is the graph-theoretic optimization problem described in the previous section. A special case of practical interest occurs when all edge weights are either zero or one. That is, all candidates for adjacency are equally desirable and we wish to maximize the number of desirable adjacencies.

1.2. The complexity of the problem

It is clear from the following theorem that the planar subgraph problem is NP-hard even in the zero–one case.

Theorem. *Given a graph* $G=(V, E)$ *and positive integer* $K \leqslant |E|$ *the problem of determining whether there exists a subset* $E' \subseteq E$ *with* $|E'| \geqslant K$ *such that* $G'=(V, E')$ *is planar is NP-complete.*

This result is due to Liu and Geldmacher [13]. See also Garey and Johnson [70, p. 197] for further information. This result reinforces recent attempts to devise polynomially time-bounded heuristics for the problem with good performance guarantees. In the next section we briefly describe some existing heuristics and in the following section we analyse their performance.

2. Heuristics for the planar subgraph problem

We assume that each edge weight in G is nonnegative. It is also convenient to assume that G is complete by putting in missing edges with zero weight. In this case it is clear that there will be an optimal solution for any instance of the planar subgraph problem which is a *maximal planar graph*,, i.e. a *triangulation*. We thus confine ourselves to heuristics which attempt to identify the triangulation of G of maximum total edge weight. We will describe three heuristics. Heuristics 1 and 2 are variants of a 'one vertex at a time' approach. Heuristic 3 is the well-known Greedy Heuristic [12] applied to this problem.

2.1 Heuristic 1 (H1)

One of the earlier approaches to the planar subgraph problem was presented by Foulds and Robinson [5]. Their heuristic avoids the complicated testing of subgraphs of G for planarity. This is achieved by building up the final triangulation, one vertex at a time, by inserting each new vertex into a face of the existing triangulation. The procedure begins by creating an initial tetrahedron, i.e. a triangulation with four vertices, which is the complete graph on four vertices K_4.

This is done as follows. Define

$$W_1(v)=\sum_{v \in e} w(e), \quad (v \in V). \tag{2.1}$$

Arrange the vertices in V in order of nonincreasing W_1 values. The four vertices with the highest W_1 values are chosen to make up the initial K_4. Vertices are then inserted one at a time in the order created. As each vertex is inserted into the triangulation it is inserted into the face which causes the largest increase in edge weight of the triangulation. To make this clear suppose vertex $v \in V$ is to be inserted next. All faces in the triangulation built up so far are examined. The face xyz with vertices x, y, and z yielding the largest sum (or *score*)

$$w(xv)+w(yv)+w(zv)$$

is identified. Then edges xv, yv, and zv are added to the triangulation along with vertex v and faces xyv, xzv and yzv. Note that xyz is no longer a face of the updated triangulation. This process is continued until all vertices have been inserted.

2.2 Heuristic 2 (H2)

This differs from Heuristic 1 only in the initial ordering of the vertices. If all edge weights are distinct we now define W_2 by

$$W_2(v)=\max_{v \in e} w(e), \quad (v \in V). \tag{2.2}$$

If the edge weights are not distinct and, without loss of generality $V=\{1,2,\ldots,n\}$ then we can implicitly perturb the weight of edge xy to $w(xy) + \varepsilon^x + \varepsilon^y$ where ε is small enough to make all edge weights distinct. This can be achieved explicitly by defining the weight of edge xy, $x > y$ to be $(w(xy), x, y)$ and then using the lexicographic ordering of these triples to compare edge weights.

2.3 Heuristic 3 (H3)

This is the Greedy Heuristic, and was first proposed for the planar subgraph problem by Foulds et al. [6]. It begins with an edgeless graph on the vertices of V. The edges of G are ordered in nonincreasing order of weight. Each edge is accepted in this order as part of the triangulation unless it causes the subgraph being built up to become nonplanar. In this case the edge is rejected. The heuristic terminates when a triangulation spanning V has been constructed. The repeated tests for planarity require significant computational effort despite Hopcroft and Tarjan's linear-time algorithm [11].

Foulds and Robinson [5] have proposed a series of filters which lessen the number of times the planarity algorithm has to be used.

3. Worst case analyses

Let P denote an instance of the planar subgraph problem. For any heuristic H let $E(H, P)$ denote the set of edges chosen by the heuristic and let $E^*(P)$ denote the maximum weight edge set. The ratio

$$R_H(P) = \frac{w(E(H, P))}{w(E^*(P))} \tag{3.1}$$

is a measure of the quality of the solution produced by H. The *worst case ratio* ρ_H is defined to be

$$\rho_H = \inf_P (R_H(P)). \tag{3.2}$$

This value has now been analysed for a large number of heuristics for a variety of combinatorial optimisation problems—see [7] for example. It gives us a guaranteed value for the quality of our solution in comparison with the optimum. ρ_H is clearly no more that 1 but usually it is much less and gives a pessimistic measure of the quality of H. In the next section we also look at a random model and analyse 'typical' values of $R_H(P)$ for a heuristic similar to our first two.

In the following results H1, H2, H3 denote the three heuristics of Section 2. Our first result shows that H1 can be arbitrarily bad in the general case, but has a performance guarantee in the zero-one case.

Theorem 3.1.
(a) $\rho_{\mathrm{H1}} = 0$,
(b) Let

$$\tilde{\rho}_{\mathrm{H1}} = \inf_Q (R_{\mathrm{H1}}(Q))$$

where the infimum now is over problems Q in which $w(e) = 0$ or 1 for all edges e. Then

$$1/6 \leqslant \tilde{\rho}_{\mathrm{H1}} \leqslant 2/9.$$

Proof. Here, as in all our examples, we will restrict the edge set of G to those edges of positive weight.

Let m be a positive integer and let $A_1, A_2, \ldots, A_m$, B be $(m+1)$ disjoint sets of size m. Let $A = \cup_{i=1}^m A_i$, $V = A \cup B$ and let $B = \{b_1, b_2, \ldots, b_m\}$. The set of edges E is defined as follows: There is an edge of weight 1 joining each pair of vertices in A. There is an edge of weight m joining each vertex in A_i to b_i for $i = 1, 2, \ldots$ m.

Using (2.1) we see

$$\begin{aligned} W_1(v) &= m^2 + m - 1, \quad v \in A, \\ &= m^2, \quad v \in B. \end{aligned} \tag{3.3}$$

Now $G = (V, E)$ contains a planar subgraph (V, E^*) with $w(E^*) \geqslant m^3$. (Simply take the edges incident with B.) We see from (3.3) that H1 first uses all the vertices of A and then the vertices of B.

After the vertices of A have been used the weight of the triangulation produced will be $3m^2 - 6$. The vertices of B are then added one by one. The increase in weight or score produced is at most $3m$ each time. Thus H1 produces a graph $(V, \tilde{E})$ where $w(\tilde{E}) \leqslant 6m^2 - 6$.

It follows that

$$\rho_{\mathrm{H1}} \leqslant (6m^2 - 6)/m^3 \quad \text{for all } m > 0,$$

which proves (a). This example can be easily modified so that all the $W_1(v)$ are equal, thus implying that a heuristic of this type which relies on ordering the vertices according to (2.1), in any fashion, will be arbitrarily bad in the worst case.

We first show that $\tilde{\rho}_{\mathrm{H1}} \geqslant 1/6$. Let $G_1 = (V, E_1)$ where $E_1 = \{e \in E : w(e) = 1\}$. In this case $W_1(v)$ is simply the degree $d_{G_1}(v)$ of vertex v and H1 orders the vertices in decreasing order of degree. Suppose this order is $v_1, v_2, \ldots, v_n$.

Let $A = \{i\text{: there exists } j < i \text{ such that } v_i v_j \in E_I\}$ and let $B = V - A$.

We note first that when H1 inserts a vertex of A into the triangulation the score is at least one. Thus if H1 chooses $(V, \tilde{E})$ then $w(\tilde{E}) \geqslant |A|$. Also $w(E') \leqslant 3n - 6$ for any planar subgraph (V, E') of G and so we can prove $\rho_{H1} \geqslant 1/6$ by showing $|A| \geqslant n/2$ or, equivalently, $|A| \geqslant |B|$.

Note next that $x, y \in B$ implies $w(xy) = 0$, i.e. B is a stable set in G_1. Now

$$\begin{aligned}
|A| &= \sum_{a\in A} 1 \\
&= \sum_{a\in A} d_{G_1}(a)/d_{G_1}(a) \qquad (\text{as } d_{G_1}(a) > 0) \\
&\geqslant \sum_{a\in A}\sum_{b\in B} \frac{w(ab)}{d_{G_1}(a)} \\
&\geqslant \sum_{a\in A}\sum_{b\in B} \frac{w(ab)}{d_{G_1}(b)} \qquad (\text{since } w(ab) = 1 \text{ implies } d_{G_1}(b) \geqslant d_{G_1}(a)) \\
&= \sum_{b\in B}\sum_{a\in A} \frac{w(ab)}{d_{G_1}(b)} \\
&= |B| \qquad \text{as } B \text{ is stable in } G_1.
\end{aligned}$$

Thus $|A| \geqslant |B|$. From the above proof we can easily see when $|A| = |B|$. The first inequality is equality if A is stable also, ie G_1 is bipartite, and the second if each component of G_1 is regular.

We show $\bar{\rho}_{H1} \leqslant 2/9$ by considering the following example: Let M_k be the following graph ($k \geqslant 3$): M_k has $3k$ vertices $v(i, j)$, $i = 0, 1, \ldots, k-1$, $j = 0, 1, 2$, and $v(i, j)$ is adjacent to $v(i', j')$ if and only if $j \neq j'$ and $i - i' \equiv 0, \pm 1 \pmod{k}$.

It is easy to see that M_k is 6-regular and connected and consequently has exactly $9k$ edges. It is also easy to see that M_k is 3-colourable, the three colours corresponding to the three j-values. It will be convenient to write R_i, B_i, Y_i for $v(i, 0)$, $v(i, 1)$ and $v(i, 2)$ respectively in order to emphasize this colouring. It also follows that M_k is nonplanar, since it has too many edges, but it can be embedded without crossings in the torus. (In fact it triangulates the torus.) However M_k does contain a planar triangulation T_k containing all the vertices and hence $(9k - 6)$ edges. (The graph T_4 is shown in Figure 1 below.) In fact T_k omits only the six edges connecting R_0, B_0, Y_0 to R_{k-1}, B_{k-1}, Y_{k-1} from M_k.

The example graphs are then essentially the following graphs M_k':

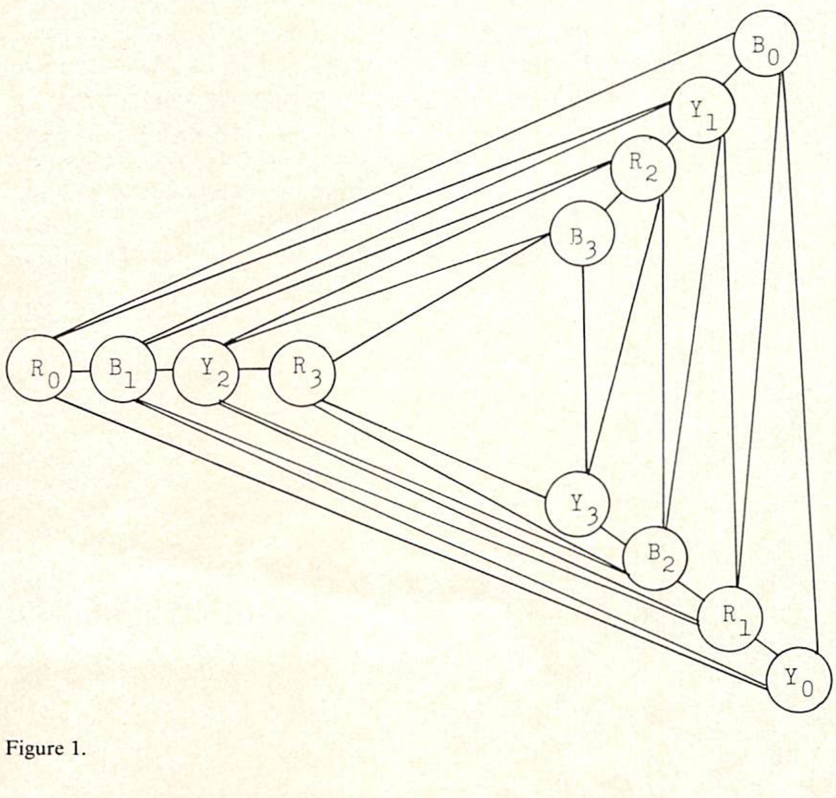

Figure 1.

M'_k has the same vertex set and all the edges of M_k and in addition the k edges of the form (R_i, B_p) such that $(p-i) \equiv 6 \pmod k$. Thus all R_i, B_i are now of degree 7, but the Y_i still have degree 6. M'_k has $10k$ edges, and obviously still contains T_k.

In order to simplify the arguments, we will augment M'_k with the following fixed graph F:

F has eight vertices, $v_1, v_2, \ldots, v_8$ and v_r is adjacent to v_s if and only if $r \leqslant 3$ or $s \leqslant 3$. Hence v_1, v_2, v_3 are of degree 7 and $v_4, v_5, \ldots, v_8$ are of degree 3. F has 18 edges and a planar subgraph of 15 edges (Figure 2).

The graph G_k is then the disjoint union of F with M'_k. Obviously G_k is not connected, but it could be made 3-connected if desired, without significantly changing the argument below adding the edges (v_4, Y_0), (v_5, Y_1), (v_6, Y_2). Now G_k has $(10k+18)$ edges and contains a planar subgraph with $(9k-6+15) = 9(k+1)$ edges.

For convenience we will make the assumptions $k > 7$ and $k \equiv 1 \pmod 3$. The first assumption is simply that k needs to be large enough to avoid special cases, and the second is made to allow the R_i vertices of M'_k to be traversed in increments of 3 in $i \pmod k$. In fact it is only necessary to have $k \not\equiv 0 \pmod 3$, but we will choose $k \equiv 1$ for definiteness. We now assume the vertices of G_k are ordered as follows, consistent with nonincreasing degree:

$v_1, v_2, v_3,$

$R_0, R_3, R_6, R_{3i}, \ldots, R_{k-1}, R_2, R_5, \ldots,$

$R_{3i-1}, \ldots, R_{k-2}, R_1, R_4, \ldots, R_{3i+1}, \ldots, R_{k-3},$

$B_0, B_1, B_2, \ldots, B_{k-1},$

$Y_0, Y_1, Y_2, \ldots, Y_{k-1},$

$v_4, v_5, \ldots, v_8.$

The initial triangle is v_1, v_2, v_3 with a score of 3. Now consider the introduction of the R_i vertices. None of these is adjacent to any of v_1, v_2, v_3 or to each other. Thus they each score zero and hence can be placed in any triangle. R_0 subdivides the initial triangle and we will then assume R_j is placed in the triangle with vertices v_2, v_3, R_{j-3}. (We assume subscript arithmetic is mod k throughout.) See Figure 3 for the triangulation after all the R vertices have been introduced.

Now consider the insertion of the B vertices. Since none of these is adjacent to v_1, v_2, v_3, they could score at most 1 if placed in any triangle with two of these as vertices. The other triangles have vertices of the form v_2 (or v_3), R_i, R_{i+3}. No B_j is adjacent to both R_i and R_{i+3} and thus could score at most 1 in any of these triangles. Furthermore, if any such triangle was previously subdivided by vertices this could not improve matters, since the B vertices are all nonadjacent to B_j. Thus each B_j scores exactly 1, since it must score at least 1 by being placed in any triangle with an adjacent R vertex. We will therefore assume B_i is placed in the triangle with vertices v_2, R_{i-6}, R_{i-3} and scores 1 from its adjacency to R_{i-6}. There is one exception to this: For B_3 there is no triangle with vertices v_2, R_{k-3}, R_0. Thus B_3 will be placed in the triangle v_2, v_3, R_{k-3} and again scores 1 from its adjacency to R_{k-3}. Figure 4 shows the triangulation after all B vertices have been inserted.

We must now insert the Y vertices. The trian-

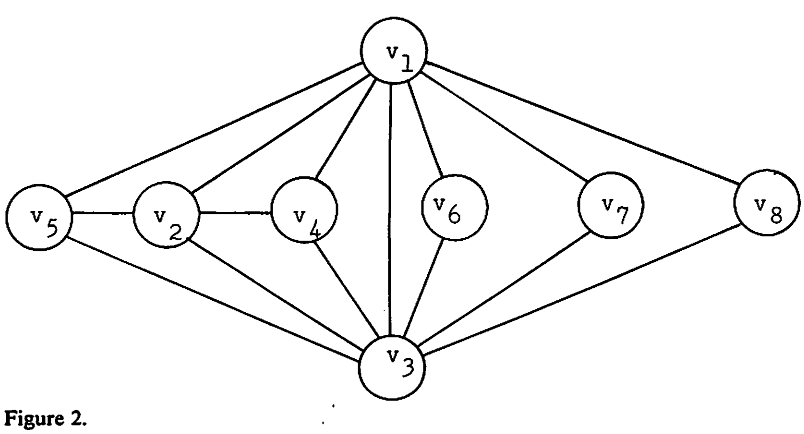

Figure 2.

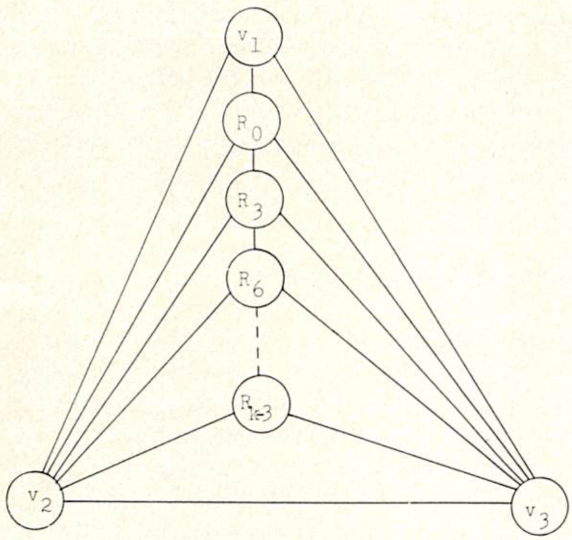

Figure 3.

gles without at least two vertices from v_1, v_2, v_3 are of one of the forms:

R_i, R_{i+3}, B_{i+6};

v_2, R_i, B_{i+6}; v_2, R_{i+3}, B_{i+6};

v_3, R_{k-3}, B_3; v_3, R_i, R_{i+3}.

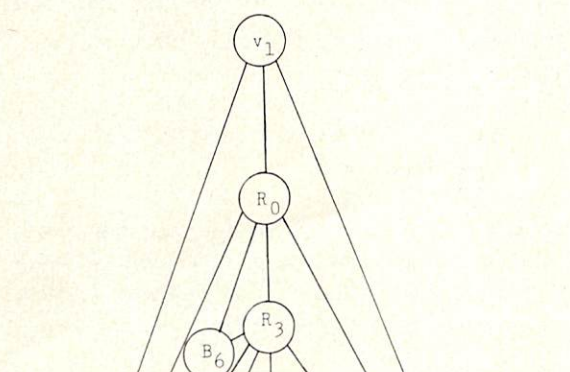

Figure 4.

Clearly no Y vertex is adjacent to more than one vertex in any of these triangles. Again it cannot help if such a triangle has been previously subdivided by other Y vertices. Thus each Y vertex scores exactly 1. We need not consider exactly where they are placed in the triangulation. Finally the 5 vertices $v_4, v_5, \ldots, v_8$ are inserted and can each score at most 3. (In fact they cannot all score 3 but we will not refine the estimate further.) The total score is thus at most

$$3 + k \times 0 + k \times 1 + k \times 1 + 5 \times 3 = 2k + 18,$$

and the ratio of this to the number of edges in the optimal planar subgraph in G_k is therefore at most

$$(2k + 18)/9(k + 1) \to 2/9 \quad \text{as } k \to \infty.$$

There is an obvious gap between this upper bound and the lower bound of 1/6 for $\tilde{\rho}_{H1}$. It is likely that neither is tight. In fact similar methods to those used in this example can (probably) be used to slightly reduce the upper bound below 2/9 using more complicated triangulations of the torus and additional edges and vertices. However it appears difficult to obain anything approaching 1/6 by this type of example.

We now consider our second heuristic H2 and show that this has a performance guarantee even in the general weighted case. We also have here an exact result.

Theorem 3.2. $\rho_{H2} = \frac{1}{6}$.

We prove this by a sequence of lemmas. First let W_2 be defined as in (2.2) and for $S \subseteq V$ let $W_2(S) = \sum_{v \in S} W_2(v)$.

Lemma 3.1. *If* $G = (V, E)$ *is a planar graph then* $w(E) \leqslant 3W_2(V)$.

Proof. Since G is planar it has a vertex of degree $\leqslant 5$. We argue by induction using such a vertex, v_0 say.

(i) If v_0 has degree $\leqslant 3$, then clearly the weight of all its incident edges is at most $3W_2(v_0)$. Let G' be the graph obtained by deleting v_0 and its incident edges, $G' = (V', E')$. Thus

$$\begin{aligned} w(E) &\leqslant w(E') + 3W_2(v_0) \\ &\leqslant 3W_2'(V') + 3W_2(v_0) \end{aligned}$$

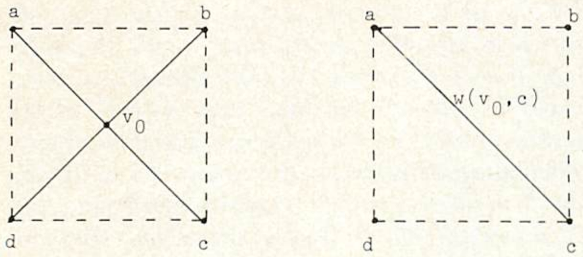

Figure 5.

by the inductive hypothesis, where W_2' is the W_2 function on G'. (Note that the w function coincides.) Clearly

$$W_2'(V') \leqslant W_2(V') \quad \text{and}$$
$$W_2(V) = W_2(V') + W_2(v_0),$$

giving the result.

(ii) If v_0 has degree 4, let its adjacent vertices be a, b, c, d where $w(v_0, a)$ is of maximum weight amongst the four edges, and a, b, c, d are in cyclic order around v_0. Form G' by deleting v_0 and its incident edges, but inserting an edge (a, c) of weight $w(v_0, c)$. See Figure 5. Clearly G' is still planar and the argument goes through similarly to case (i), since 3 edges have been deleted.

(iii) v_0 has degree 5. This is similar to case (ii), but we insert two edges. (See Figure 6.) Again a is the vertex such that $w(v_0, a) = W_2(v_0)$.

As a basis for the induction, when $|V| = 2$, $w(E) = W_2(V)/2$.

Corollary 3.1. *If G is an arbitrary graph and $T \subseteq E$ is the set of edges in a maximum weight planar subgraph, then $w(T) \leqslant 3W_2(V)$.*

Proof. Obvious, since the vertex weights in the whole of G are all at least as large as those in the subgraph (V, T). These inequalities are strict if all

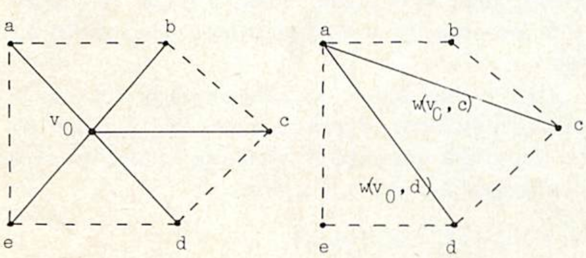

Figure 6.

edge weights are positive. Assume now that $G = (V, E)$ is our complete weighted graph with unequal non-negative edge weights. (See below for the case of equal weights.) The vertices are sorted in order of $W_2(v)$ nonincreasing before applying the heuristic. Let $\tilde{E}$ be the edges of G selected by the heuristic.

Lemma 3.2. $w(\tilde{E}) \geqslant \frac{1}{2} W_2(V)$.

Proof. By induction on the number of vertices, $n = |V|$. Let v_n be the last vertex in the ordering. Thus $W_2(v) \geqslant W_2(v_n)$ for all $v \in V$ and, in view of the unequal edge weights hypothesis, there is at most one vertex such that $W_2(v) = W_2(v_n)$ and $v \neq v_n$. Clearly in this case this vertex must be v_{n-1}, and $w(v_n, v_{n-1}) = W_2(v_n) = W_2(v_{n-1})$. Consider v_n. There are two cases:

(i) $W_2(v_{n-1}) = W_2(v_n)$. In this case delete v_n, v_{n-1} and all their incident edges (all of which have weight $\leqslant W_2(v_n)$) to give G'. In G' the weights of $v_1, v_2, \ldots, v_{n-2}$ are unchanged therefore. Let E' be the edges in the triangulation after v_{n-2} is inserted. Thus by the induction

$$w(E') \geqslant \tfrac{1}{2} W_2(V').$$

Now as v_{n-1} and v_n are inserted, between them they must score at least $W_2(v_n)$. Thus

$$\begin{aligned} w(E) &\geqslant w(\tilde{E}') + W_2(v_n) \\ &\geqslant \tfrac{1}{2} W_2(V') + \tfrac{1}{2}(W_2(v_{n-1}) + W_2(v_n)) \\ &= \tfrac{1}{2} W_2(V). \end{aligned}$$

(ii) $W_2(v_{n-1}) > W_2(v_n)$. In this case delete only v_n and its incident edges to give G', with E' being the triangulation after v_{n-1} is inserted. Now v_n must score at least $W_2(v_n)$. Then

$$\begin{aligned} w(E) &\geqslant w(\tilde{E}') + W_2(v_n) \\ &\geqslant \tfrac{1}{2} W_2(V') + W_2(v_n) \\ &\geqslant \tfrac{1}{2} W_2(V), \end{aligned}$$

since $W_2(v_n) \geqslant 0$.
As basis, when $n = 3$ it is obvious we must have

$$w(\tilde{E}) = w(E) = w(e_1) + w(e_2) + w(e_3)$$

where $w(e_1) > w(e_2) > w(e_3)$ since

$$W_2(V) = 2w(e_1) + w(e_2),$$
$$\tfrac{1}{2} W_2(V) = w(e_1) + \tfrac{1}{2} w(e_2)$$

and

$$w(\tilde{E}) - \tfrac{1}{2}W_2(V) = \tfrac{1}{2}w(e_2) + w(e_3) \geq 0$$

as required.

It follows directly from Corollary 3.1 and Lemma 3.2 that $\rho_{H2} \geq 1/6$ when restricted to unequal edge weights.

As previously mentioned, if any edge weights are equal we must resolve the ties. This can be done by numbering the vertices initially from 1 to n in arbitrary order, and letting $w'(i,j) = w(i,j) + \varepsilon^i + \varepsilon^j$ if $(i,j) \in E$, where ε is a 'small' perturbation. If the problem is solved with the weights w' this corresponds to resolving ties as follows: If (i,j), (k,l) are two different edges of G with $i<j$, $k<l$ then $w'(i,j) > w'(k,l)$ corresponds to either:

(i) $w(i,j) > w(k,l)$,

(ii) $w(i,j) = w(k,l)$ and $i>k$,

(iii) $w(i,j) = w(k,l)$, $i=k$ and $j>l$.

Thus the ordering of the edges (i,j) with $i<j$ is lexicographic in the triples $[w(i,j),i,j]$. With this ordering the unequal weights case carries over. The weight for vertex i is then

$$\text{lex max}\{[w(i,j),i,j] : (i,j) \in E \text{ and and smaller of } i,j \text{ is last}\},$$

and the vertex ordering respects this lexicographic order.

To prove $\rho_{H2} \leq \frac{1}{6}$ we consider the following

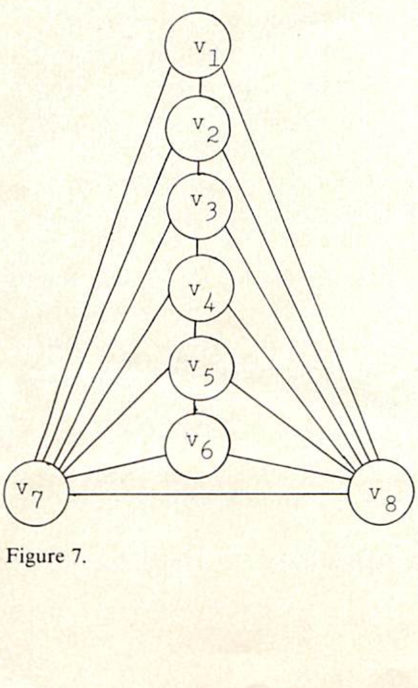

Figure 7.

example. Let the graph L_k be as follows: L_k has $2k+2$ vertices: $v_1, v_2, \ldots, v_{2k+2}$, v_i is adjacent to v_j $(i<j)$ if $j = i+1$, $2k+1$ or $2k+2$.

Now L_k is a planar trianguation with $6k$ edges. See Figure 7 for L_3.

We assume that all edges of L_k have essentially unit weight, but are perturbed slightly so that $w(v_{2i-1}, v_{2i})$ for $i = 1, 2, \ldots, (k+1)$ is decreasing with increasing i, and all these weights exceed any other edge weight. Thus $W(v_{2i-1}) = W(v_{2i})$ and $W(v_{2i})$ is decreasing with i.

Again to simplify the arguments, we will augment L_k with a single disjoint triangle K_3 with vertices V_1, V_2, V_3. The edge weights of K_3 are again essentially unity, but perturbed so that they exceed any edge weight in L_k. We will assume without loss of generality that $W_2(V_1) = W_2(V_2) > W_2(V_3)$, and these vertex weight exceed any in L_k.

Let G'_k be this augmented (weighted) graph, which is planar with $(6k+3)$ edges. Thus G'_k has total edge weight approximately $(6k+3)$. We will ignore the effects of the perturbation in weight calculations except insofar as they determine the vertex ordering; clearly in the limit as the perturbations tend to zero this is valid. Again it will be observed that G'_k is not connected, but could be made so by the addition of suitable weighted edges without changing significantly the following argument.

We will now assume the vertices are ordered as follows, consistent with nonincreasing vertex weight:

$$V_1, V_2, V_3, v_2, v_1, v_4, v_3, v_6, v_5, \ldots, v_{2i},$$
$$v_{2i-1}, \ldots, v_{2k}, v_{2k-1}, v_{2k+2}, v_{2k+1}.$$

The initial triangle is K_3 with a total score of 3. Then v_2 will subdivide this triangle, scoring 0. Now v_1 can score at most 1 from its adjacency to v_2. We therefore assume it is placed in the triangle v_1, V_2, V_3 scoring 1.

Now inductively we will suppose that immediately prior to the insertion of v_{2i} $(i \leq k)$ we have constructed within K_3 the central path $V_1, v_2, v_1, \ldots, v_{2i-2}, v_{2i-3}$ with each of these vertices connected to both V_2 and V_3 (see Figure 8) and that the total score is $(i-1)+3$.

Now consider the insertion of v_{2i} and v_{2i-1}. When v_{2i} enters it scores 0, since it is not adjacent to any vertex currently in the triangulation. We will thus assume it is placed in the triangle V_2, V_3,

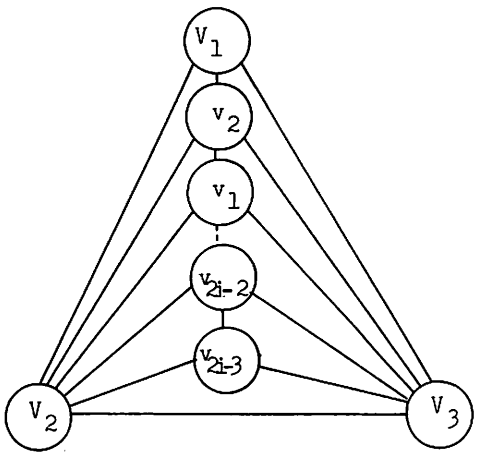

Figure 8.

v_{2i-3} extending the central path.

Now v_{2i-1} is inserted. It is adjacent to only v_{2i-2} and v_{2i} in the triangulation. However v_{2i-2} and v_{2i} are not both vertices of the same triangle, since they are separated by v_{2i-3} in the central path. Thus v_{2i-1} can score at most 1. We therefore place it in the triangle V_2, V_3, v_{2i} and it scores 1 from its adjacency to v_{2i} since

$$w(v_{2i-1}, v_{2i}) > w(v_{2i-1}, v_{2i-2}).$$

The total score has therefore increased to $(i+3)$ and the inductive hypothesis is clearly maintained. It follows that after v_{2k-1} is inserted the total score is $(k+3)$. Finally v_{2k+2} and v_{2k+1} are introduced and score at most 3 each. (In fact v_{2k+2} can score at most 2 but this refinement is unimportant.) Hence the total score is at most $(k+9)$. The ratio of the weight of the triangulation constructed to the optimal (i.e. the whole of G'_k) is therefore at most

$$(k+9)/(6k+3) \to \tfrac{1}{6} \quad \text{as } k \to \infty.$$

An objection to this example which might be raised is that the vertices v_{2k+1}, v_{2k+2} have very high degree, and that a slight modification of the heuristic to take account of this might perform better. However this does not seem to be the case. We have more complicated examples, in which the vertex degrees remain bounded, which tend to the ratio 1/6 in the limit.

We now consider Heuristic H3. We again have an exact result.

Theorem 3.3. $\rho_{H3} = \frac{1}{3}$.

Proof. As already observed, Heuristic H3 is in fact the Greedy Heuristic [12] applied to the independence system consisting of the sets of edges of G which induce planar graphs.

It follows from general results on the greedy heuristic for independence systems [12] that the worst case occurs when all edge weights are 0 or 1. Thus, if the graph has n vertices, let U_n be the subgraph of unit-weight edges. The problem is then that of determining a maximum (edge) cardinality planar subgraph in U_n. Suppose U_n has c connected components, then its maximum cardinality planar subgraph has at most $(3n-6c)$ edges. The greedy heuristic always constructs an edge maximal planar subgraph (EMPS) of U_n. However, any such EMPS of U_n must contain a spanning tree of each of the components of U_n, so the cardinality of the heuristic subgraph is at least $(n-c)$. Thus the worst-case ratio is at least

$$(n-c)/(3n-6c) > \tfrac{1}{3} \quad \text{for any } c \geqslant 1.$$

We will now exhibit a family of graphs G_k^* for which there is an EMPS containing (as $k \to \infty$) only one-third the number of edges in the maximum planar subgraph. Since the edges in this EMPS could be ordered first, it then follows that the ratio 1/3 is a tight worst-case bound for the greedy heuristic in this problem.

The graph G_k^* is based on the graph M_k used in Theorem 3.1. We will assume for convenience $k = 4r$ (i.e. $k \equiv 0 \pmod 4$). Let M_k^* be the graph obtained by disconnecting from the rest of M_k the two triangles R_0, B_0, Y_0 and R_{2r}, B_{2r}, Y_{2r}. This removes 24 edges from M_k. Thus M_k^* has $(9k-24)$ edges and is planar, since it is a subgraph of T_k.

We now add to M_k^* the following edges to give G_k^*:

(i) All edges of the form (R_i, R_{i+1}), (B_i, B_{i+1}), (Y_i, Y_{i+1}) for $i = 0, 1, \ldots, (k-1)$. (Subscript arithmetic is again mod k.) There are $3k$ such edges forming 3 disjoint circuits which we will label C_R, C_B, C_Y respectively.

(ii) The three edges (R_{3r}, B_r), (B_{3r}, Y_r), (Y_{3r}, B_r). Thus G_k^* has $3k$ vertices and $(12k-21)$ edges. It contains a planar subgraph with at least $(9k-24)$ edges.

Let H_k be the subgraph of G_k^* containing all the

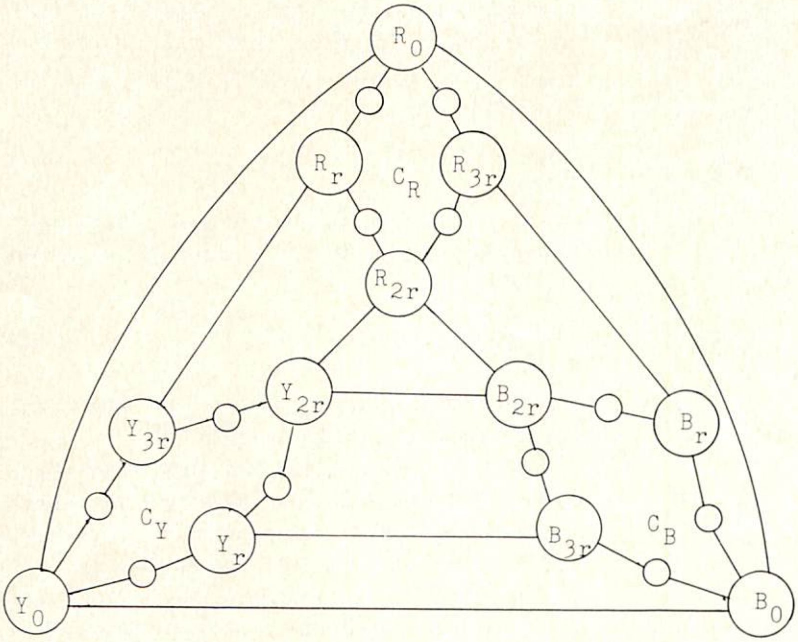

Figure 9.

vertices, the edges added in (i), (ii) above (i.e. those not in M_k^*), and the two triangles with vertices R_0, B_0, Y_0 and R_{2r}, B_{2r}, Y_{2r}. Thus H_k has $(3k+9)$ edges, and is in fact planar. (See Figure 9.)

We now show that H_k is an EMPS of G_k^*. The edges of G_k^* not in H_k are the remaining edges of M_k^*. The symmetry, assume that an edge of M_k^* of the form (R_i, B_j) can be added to H_k without destroying planarity. We must have $|i-j| \leqslant 1$ and $i, j \neq 0, 2r$.

Consider the subgraph of H_k induced by the R, B vertices, then it follows that the situation is as shown in Figure 10.

We will assume R_i is on the chain $R_0, R_1, \ldots, R_{2r}$ of C_R and B_j on the chain $B_0, B_1, \ldots, B_r$ of C_B. Symmetrical arguments hold in the other possible cases. Therefore we complete the argument by showing that the graph of Figure 10 is nonplanar. Suppose we omit all vertices on the chain R_{2r+1}, $R_{2r+2}, \ldots, R_{3r-1}$ from this graph, and consider any chain of degree 2 vertices as an edge. We obtain the graph shown in Figure 11. However this graph is the bipartite complete graph $K_{3,3}$. (The two vertex sets are R_0, B_j, B_{2r} and B_0, R_i, B_r.)

Thus the graph of Figure 10 is nonplanar by Kuratowski's criterion. Thus H_k is an EMPS of

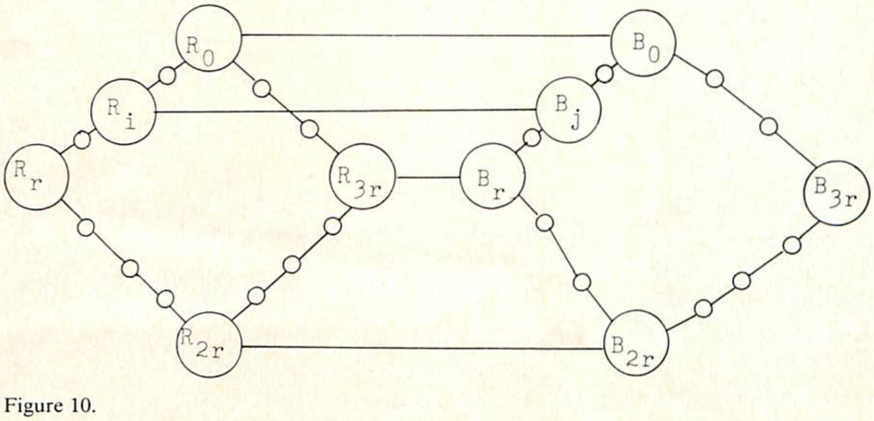

Figure 10.

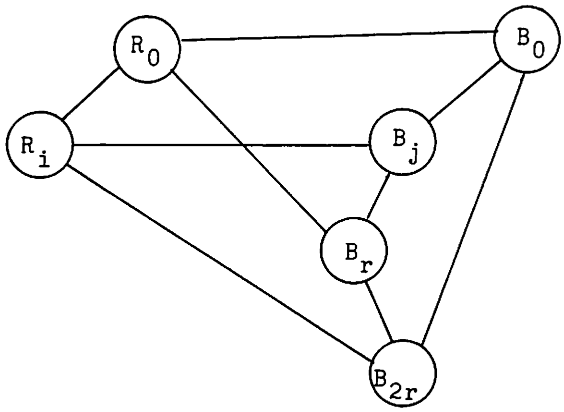

Figure 11.

G_k^*. The ratio of the number of edges in H_k to those in M_k^* is

$$(3k+9)/(9k-24) \to \tfrac{1}{3} \quad \text{as } k \to \infty,$$

completing the analysis. Since H3 clearly has a much better worst-case performance guarantee than H1 or H2, we ought to point out that while H1 and H2 can be very simply implemented in $O(n^2)$ time, it is not clear that H3 can be implemented in anything better than $O(n^3)$ time, even using a sophisticated planarity testing algorithm [11].

4. A random model

We consider now a simple random model for our problem. In particular we assume that the edge weights are distributed independently as uniform [0,1] random variables. This distributional assumption may seem unduly restrictive, but our results generalise easily to independent edge weights chosen from any probability density restricted to a bounded interval of the nonnegative reals. However, we restrict ourselves to the uniform [0,1] case for simplicity of presentation. The heuristic we consider is a simplification of H1 or H2 above. The vertices are considered in their given order $1, 2, \ldots, n$ and then inserted into the best position in the current triangulation as in H1 or H2. We will label this heuristic H0. (Its worst-case behaviour can be shown to be arbitrarily bad even in the zero–one case.) Let V_n, V_n^* be the random variables which take the value of the heuristic and optimal solutions respectively under our random model. Then, in the notation of (3.1), $R_{\text{H0}} = V_n/V_n^*$ is also a random variable. We will show that

$$\lim_{n\to\infty} \Pr\left(R_{H0} \leqslant 1 - n^{-0.1}\right) = 0, \tag{4.1}$$

i.e. H0 is asymptotically arbitrarily good with probability tending to 1. To prove (4.1) we need only show

$$\lim_{n\to\infty} \Pr\left(V_n \leqslant 3n - n^{0.9}\right) = 0 \tag{4.2}$$

and use the fact that $V_n^* \leqslant 3n-6$ in view of the upper bound on the edge weights.

Let T_k be the triangulation after vertex v_k is added. To prove (4.2) we examine the increase in weight of the current triangulation T_{k-1} when we add v_k, $k = 4, 5, \ldots, n$. Now for any triangle f of T_{k-1} the score Z from placing v_k in f is the sum of three independent uniform [0,1] variables. It is easy to show that

$$\Pr(Z \leqslant 3-a) = 1 - a^3/6. \tag{4.3}$$

We associate now an auxiliary graph $G_k = (F_k, A_k)$ with T_k. The vertices of F_k are the triangles of T_k and $f_1 f_2 \in A_k$ if the triangles f_1, f_2 have a common vertex. Let $\alpha(G_k)$ denote the size of the largest stable set of G_k. We will show later that if $m = 18n^{0.3}\log n$, then for large n

$$\Pr\left(\alpha(G_k) < m \quad \text{for any } k \geqslant n^{0.9}\right) = O(n^{-2}). \tag{4.4}$$

Let $Z_k = w(T_k) - w(T_{k-1})$ be the score of vertex v_k. It follows from (4.3) that

$$\begin{aligned}&\Pr\left(Z_k \leqslant 3 - n^{-0.1} \mid \alpha(G_{k-1}) \geqslant m\right)\\ &\quad\leqslant \left(1 - n^{-0.3}/6\right)^m\\ &\quad\leqslant \exp\left(-mn^{-0.3}/6\right) = n^{-3}.\end{aligned}$$

If now

$$Z_s = \sum_{k=n^{0.9}+1}^{n} Z_k$$

is the total score of all vertices v_k for $k > n^{0.9}$, it follows using Boole's inequality that

$$\Pr\left(Z_s \leqslant 3n - n^{0.9} \mid \alpha(G_k) \geqslant m \quad \text{for all } k \geqslant n^{.9}\right) \leqslant n^{-2}$$

which, together with (4.4) implies (4.2).

It remains only to prove (4.4). Let D_k, D_k' denote the maximum degrees of T_k, G_k respectively. Since the degree of f in G_k is exactly six less than the sum of the degrees of its three vertices in T_k, we have $D'_k \leqslant 3D_k - 6$. Now, from an easy inequality on the chromatic number of a graph [9], it follows that

$$\begin{aligned}\alpha(G_k) &\geqslant |F_k|/(D'_k+1) = (2k-4)/(D'_k+1)\\ &\geqslant 2(k-2)/(3D_k-5)\\ &\geqslant 2(k-2)/(3D_n-5)\end{aligned}$$

since $D_k \leqslant D_n$ for $k \leqslant n$.

Now if $D_n \leqslant 6n^{1/2}\log n$ and $k \geqslant n^{0.9}$ it follows that, for large enough n, $\alpha(G_k) \geqslant m$. Thus if we can prove

$$\Pr(D_n > 6n^{1/2}\log n) = \mathrm{O}(n^{-2}) \tag{4.5}$$

we will have proved (4.4) and hence (4.1).

Let d_{jk} be the degree of v_j in T_k. Since v_{k+1} is equally likely to be placed in any of the $(2k-4)$ triangles of T_k and v_j is adjacent to d_{jk} of these, we have the following conditional distribution for $d_{j,k+1}$.

$$\begin{aligned}&\Pr(d_{j,k+1} = p+1 | d_{jk} = p) = p/(2k-4),\\ &\Pr(d_{j,k+1} = p | d_{jk} = p) = 1 - p/(2k-4).\end{aligned} \tag{4.6}$$

We also have $d_{jj} = 3$. For notational convenience, we will suppress the suffix j and write $X_k = d_{jk}$. Also for any integer x write

$$x^{(i)} = x(x+1)(x+2)\cdots(x+i-1),$$

and let $\mu_k^{(i)} = E(X_k^{(i)})$. Clearly $\mu'^{i}_k = E(X_k^i)$, the ith moment of X_k about the origin, satisfies $\mu'^{i}_k \leqslant \mu_k^{(i)}$.

Now, from (4.6),

$$\begin{aligned}E(X_{k+1}^{(i)} | X_k = p) &= (p+1)^{(i)}\frac{p}{2k-4}\\ &\quad + p^{(i)}\left\{1 - \frac{p}{2k-4}\right\}\\ &= p^{(i)}\left\{1 + \frac{i}{2k-4}\right\}.\end{aligned}$$

Thus

$$\begin{aligned}\mu_{k+1}^{(i)} &= \left[1 + \frac{i}{2k-4}\right]\mu_k^{(i)}\\ &= \left[1 + \frac{i}{2k-4}\right]\left[1 + \frac{i}{2k-6}\right]\\ &\quad \cdots\left[1 + \frac{i}{2j-4}\right]\mu_j^{(i)}.\end{aligned}$$

However

$$\mu_j^{(i)} = 3.4\cdots(i+2) = \tfrac{1}{2}(i+2)!$$

and

$$\begin{aligned}&\left[1 + \frac{i}{2k-4}\right]\cdots\left[1 + \frac{i}{2j-4}\right]\\ &\quad \leqslant \exp\left[\frac{i}{2k-4} + \cdots + \frac{i}{2j-4}\right]\\ &\quad = \exp\left[\frac{i}{2}\left\{\frac{1}{k-2} + \cdots + \frac{1}{j-2}\right\}\right]\\ &\quad \leqslant \exp\left[\frac{i}{2}\log\left(\frac{k-2}{j-3}\right)\right]\\ &\quad = \left[\frac{k-2}{j-3}\right]^{i/2}.\end{aligned}$$

Thus

$$\mu_{k+1}^{(i)} \leqslant \tfrac{1}{2}(i+2)!\left[\frac{k-2}{j-3}\right]^{i/2},$$

or

$$\mu_k^{(i)} \leqslant \tfrac{1}{2}(i+2)!\left[\frac{k-3}{j-3}\right]^{i/2}. \tag{4.7}$$

Consider the moment generating function $M(t) = E(\mathrm{e}^{tX_k})$ of X_k.

We have

$$\begin{aligned}M(t) &= \sum_{i=0}^{\infty}\mu'^{i}_k\frac{t^i}{i!} \leqslant \sum_{i=0}^{\infty}\mu_k^{(i)}\frac{t^i}{i!}\\ &\leqslant \tfrac{1}{2}\sum_{i=0}^{\infty}(i+1)(i+2)\left[t\left[\frac{k-3}{j-3}\right]^{1/2}\right]^i\\ &\quad \text{(using (4.7))}\\ &= \left[1 - t\left[\frac{k-3}{j-3}\right]^{1/2}\right]^{-3}\\ &\quad \text{(for small enough } t\text{)}.\end{aligned}$$

Now it follows from a well-known generalisation of the Markov inequality (see, for example, Grimmett and Stirzaker [8]) that

$$\Pr(X_k \geqslant x) \leqslant \mathrm{e}^{-tx}M(t).$$

Thus

$$\Pr(X_k \geqslant x) \leqslant \mathrm{e}^{-tx}\left[1 - t\left[\frac{k-3}{j-3}\right]^{1/2}\right]^{-3},$$

and putting, for example,

$$t = \frac{1}{2}\left[\frac{j-3}{k-3}\right]^{1/2}$$

gives

$$\Pr(X_k \geqslant x) \leqslant 8 \exp\left[-\frac{1}{2}\left[\frac{j-3}{k-3}\right]^{1/2} x\right],$$

and since $j \geqslant 4$ and $k \leqslant n$ we obtain

$$\Pr(X_k \geqslant x) \leqslant 8 \exp\left[-\tfrac{1}{2} n^{-1/2} x\right]$$

Hence

$$\Pr\left(X_n \geqslant 6n^{1/2} \log n\right) \leqslant 8 \exp(-3 \log n) = 8\, n^{-3}.$$

Therefore, using Boole's inequality,

$$\Pr\left(D_n \geqslant 6n^{1/2} \log n\right) \leqslant n \Pr\left(X_n \geqslant 6n^{1/2} \log n\right) = \mathrm{O}(n^{-2}) \qquad (4.8)$$

which establishes (4.5). We note that since

$$\Pr\left(V_n \leqslant 3n - n^{0.9}\right) = \mathrm{O}(n^{-2}),$$

The Borel–Cantelli lemma [8] implies that

$$\Pr\left(\lim_{n\to\infty} V_n / V_n^* = 1\right) = 1$$

which is a somewhat stronger result than (4.1).

Finally we may observe that the degree inequality (4.8) does not depend strongly on our distributional assumptions. It will be true for any joint distribution of the edge-weights which is invariant under permutation of edges. Thus our results have many easy generalisation.

5. Conclusions

We have analysed the worst-case behaviour of some heuristics for determining a maximum weight planar subgraph in a graph with nonnegative edge weights. Whilst we have proved that the worst case of these heuristics may be quite bad, we have also shown that (under certain assumptions) even a very simple heuristic might be expected to perform almost optimally 'on the average'. The one tantalising gap in our analysis is the disparity between the upper and lower bounds in Theorem 3.1(b). It appears to be a difficult problem to resolve this.

References

[1] Apple, J.M., *Plant-layout and materials handling*, Ronald Press, New York, 2nd ed., 1963.

[2] Buffa, E.S., Armour, G.C., and Vollman, T.E., "Allocating facilities with CRAFT," *Harvard Business Review 42* (1964) 136–159.

[3] Dickey, J.W., and Hopkins, J., "Campus building arrangements using TOPAZ," *Transportation Research* 6 (1972) 59–68.

[4] Elshafei, A.N., "Hospital layout as a quadratic assignment problem," *Operational Research Quaterly* 28 (1977) 167–179.

[5] Foulds, L.R., and Robinson, D.F., "Graph theoretic heuristics for the plant layout problem," *International Journal of Production Research* 16 (1978) 27–37.

[6] Foulds, L.R., Gibbons, P.B., and J.W. Giffin, "The facilities layout problem: An experimental comparison of graph theoretic heuristics", to appear.

[7] Garey, M.R., and Johnson, D.S., *Computers and Intractability, Freeman, San Francisco*, 1979.

[8] Grimmett, G., and Stirzaker, D., *Probability and Random Processes,* Clarendon Press, Oxford, 1982.

[9] Harary, F., *Graph Theory,* Addison-Wesley, Reading, 1969.

[10] Hilliers, F.S., and Connors, M.H., "Quadratic assignment problem and the location of indivisible facilities", *Management Science* 13 (1966) 42–57.

[11] Hopcroft, J.E., and Tarjan, R.E., "Efficient planarity testing," *Journal of the Association of Computing Machinery* 21 (1974) 549–568.

[12] Korte, B., and Hausmann, D., "An analysis of the Greedy heuristic for independence systems", *Annals of Discrete Mathematics* 2 (1978) 65–74.

[13] Liu, P.C., and Geldmacher, R.F., "On the deletion of nonplanar edges of a graph", Stevens Institute of Technology, 1978.

[14] Mather, R., *Systematic Layout Planning,* Cahners, Boston, 2nd ed., 1973.

[15] Moore, J.M., *Plant Layout and Design,* MacMillan, New York, 1963.

[16] Vollman, T.E., Nugent, C.E., and Zartler, R., "A computerized model for office layout", *Journal of Industrial Engineering* 19 (1968) 321–327.