JOURNAL OF COMBINATORIAL THEORY, Series B 26, 66–74 (1979)

Upper and Lower Bounds for Graph-Diameter Problems with Application to Record Allocation

D. S. HIRSCHBERG* AND C. K. WONG†

* *Department of Electrical Engineering, Rice University, Houston, Texas 77001; and*
† *IBM T. J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598*

Communicated by C. Berge

Received April 12, 1976

This paper studies three graph problems with parameters n, the number of nodes, e, the number of edges, and k, the diameter of the graph. Given any two of these three parameters, the problem is to construct a directed graph which minimizes or maximizes the third. The first problem has its origin in a recent study of record allocation in a paged computer system. It is shown how to construct graphs that are optimal for all three problems in some cases and are asymptotically optimal for other cases. The solution of the second problem answers a question raised by Berge in "The Theory of Graphs and its Application," 1962.

I. INTRODUCTION

In this paper, we study three graph problems with parameters n, the number of nodes, e, the number of edges, and k, the diameter of the graph. Given any two of these three parameters, we seek a directed graph which minimizes or maximizes the third. Specifically, the first problem is: Given n and k, find a directed graph with n nodes, whose diameter is not larger than k, such that it has the minimum possible number of edges. This problem will be referred to as the $\langle n, k\rangle$ problem. The second problem is: Given e and n, find a directed graph with n nodes and at most e edges whose diameter is minimum. This problem will be referred to as the $\langle n, e\rangle$ problem. The last problem is: Given e and k, find a directed graph with at most e edges whose diamter is not larger than k such that it has the maximum possible number of nodes. This problem will be referred to as the $\langle e, k\rangle$ problem.

The first problem (the $\langle n, k\rangle$ problem) has its origin in a recent study of record allocation in a paged computer system. Given n pages of records, one wants to assign page pointers to each page. These pointers point to other pages after whatever operations on the page containing these pointers are completed. Since each pointer corresponds to a page fault, one naturally

0095-8956/79/010066–09$02.00/0

wants to minimize the total number of page pointers. On the other hand, in order to guarantee the performance of the system as a whole, the following constraint is imposed: the number of pointers traced when going from one page to another should not be larger than k. Represented as a graph, this problem becomes the $\langle n, k \rangle$ problem. The other two problems are natural complements of the first one and are of interest in their own rights. In fact, the second problem was mentioned in [1].

Since optimal solutions seem to be difficult to find, we propose some heuristics. We shall show how to construct graphs that are optimal for all three problems in some cases and are asymptotically optimal (in a certain sense) for other cases.

Finally, it should be pointed out that some aspects of these problems have been studied by various authors [3–5].

For definitions concerning graphs, we refer to [2].

II. THE $\langle n, k \rangle$ PROBLEM

This problem is to find a directed graph $G = (V, E)$ with minimum size edge set E such that the vertex set, V, is of size n and there is a path between any two vertices that is of length k or less ($k \leqslant n - 1$). Let $e(n, k)$ be the size of the minimum edge set.

DEFINITION 1. Given n, k ($k < n - 1$), a flower graph $G_{n,k}$ is constructed as follows: Let there be a loop having $\lceil k/2 \rceil + 1$ vertices. Pick one of the vertices and call it CENTER. As long as there are at least $\lfloor k/2 \rfloor$ vertices that have not yet been picked, pick $\lfloor k/2 \rfloor$ of them and, together with CENTER, add edges to make a loop. If there are less than $\lfloor k/2 \rfloor$ vertices unpicked, add edges so as to make a loop using those vertices and CENTER. Finally, assign the same direction to all edges in a loop.

An example for $n = 5$, $k = 3$ is given in Fig. 1.

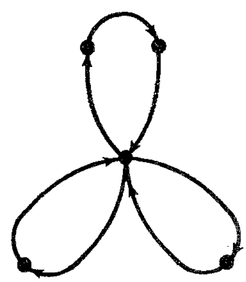

FIG. 1. A Flower Graph for $n = 5$, $k = 3$

THEOREM 1.

(i) $e(n, k) \leqslant n + \lceil (n-1-\lceil k/2 \rceil)/\lfloor k/2 \rfloor \rceil$ *for* $k < n-1$.
(ii) $e(n, n-1) = n$.

Proof. We show (i) by counting the edges in a flower graph $G_{n,k}$ and demonstrating that it satisfies the problem's path requirements.

There is one edge emanating from each vertex except for CENTER. CENTER emanates as many edges as there are loops. The expression $\lceil (n-1-\lceil k/2 \rceil)/\lfloor k/2 \rfloor \rceil + 1$ is precisely the number of loops in the flower graph. Thus the expression on the right-hand side of (i) gives the number of edges in $G_{n,k}$. We need but show that the flower graph satisfies the problem's path requirements.

Let x be in loop L_x and y be in loop L_y. If $L_x = L_y$ then clearly the path $x \to y$ is of length $\leqslant \lceil k/2 \rceil < k$. If $L_x \neq L_y$ then the larger of these loops can have at most $\lceil k/2 \rceil + 1$ vertices, the smaller can have at most $\lfloor k/2 \rfloor + 1$ vertices. If L_x is the larger, then $x \to$ CENTER is of length at most $\lceil k/2 \rceil$ and CENTER $\to y$ is of length at most $\lfloor k/2 \rfloor$. If L_y is the larger, then $x \to$ CENTER is of length at most $\lfloor k/2 \rfloor$ and CENTER $\to y$ is of length at most $\lceil k/2 \rceil$. In either case, the path $x \to$ CENTER $\to y$ is of length at most k.

(ii) is demonstrated by constructing a single loop with all n vertices. There are n edges and the path $x \to y$ for $x \neq y$ is of length at most $n - 1$. Clearly, such a graph is optimal. ∎

THEOREM 2. $e(n, k) \geqslant n - 1 + 2(n-1)/k$ *for* $k < n - 1$ *and hence*

$$e(n, k) \geqslant n - 1 + \lceil 2(n-1)/k \rceil \qquad \textit{for} \quad k < n-1.$$

Proof. In a minimal graph G for the $\langle n, k \rangle$ problem, let m be the maximum length of a chain of nodes all of whom have outdegree $= 1$, and let C be an example of such a chain, let ROOT be the first node in C, let BASE be the node pointed to by the last node in C. (See Fig. 2).

Let T be a spanning tree of G such that the root of T will be ROOT and the path, $P(v)$ in T, from ROOT to any vertex v will be a minimal length path ROOT $\to v$ in G. For v not in C, $P(v)$ will include the m nodes in C as well as BASE.

All edges in G that are not in T will be either back edges or cross edges. Forward edges are not possible by definition of T. Define the *level* of vertex v to be the length of the path in T from ROOT to v. The level of ROOT is zero, the level of BASE is m, and the level of any vertex must be no more than k since G satisfies the $\langle n, k \rangle$ condition. For any leaf, w, the length of the path from BASE to w is $\leqslant k - m$.

Define vertex v to be a *junction node* if there are at least 2 tree edges emanating from v. BASE is an example of a junction node.

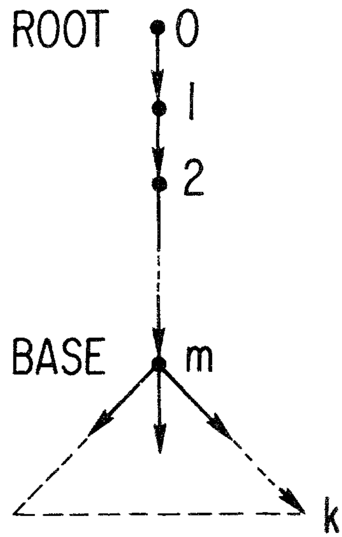

FIGURE 2

Define vertex v to be an *emitter node* if outdegree $(v) \geqslant 2$ and v is not a junction node.

If, for each leaf, w, we backtrack up the tree and consider the set S_w of the first $m + 1$ nodes encountered (including w) then, by the maximality of m, there must be a node with outdegree at least 2 and so one of the following two conditions must hold:

(1) S_w contains a junction node

or

(2) S_w contains an emitter node.

If condition (1) is true we will say that w is a class 1 member, otherwise w is a class 2 member. Let the size of class 1 be L_1, the size of class 2 be L_2, then L = total number of leaves $= L_1 + L_2$.

We shall count the number of nodes in the tree (a total of n) by backtracking from the leaves and, to avoid duplication of nodes in the inventory, by halting the backtracking from a leaf when the nodes higher up the tree will be counted by some other path.

The inventory count is as follows:

ROOT $\rightarrow$ BASE, $= m + 1$
Nodes in class 1 paths, at most half of which are needed to continue at or above their juction node in order to count all nodes in their paths, $\leqslant (k - m) L_1/2 + mL_1/2$
Nodes in class 2 paths, $\leqslant (k - m) L_2$

From the above, we get:

$$n \leqslant m + 1 + kL_1/2 + (k - m)\, L_2 \tag{1}$$

The number of tree edges is $n - 1$, the number of edges from leaves is at least $L_1 + L_2$ (each leaf must have outdegree at least one), and the number of other edges must be at least L_2. Taken all together, this yields:

$$e \geqslant n - 1 + L_1 + 2L_2 \tag{2}$$

From (1) we get $L_1 \geqslant 2[n - m - 1 - (k - m)\, L_2]/k$ which, when substituted into (2), results in:

$$e \geqslant n - 1 + 2(n - 1)/k + 2m(L_2 - 1)/k$$

If $L_2 \geqslant 1$ then $e \geqslant n - 1 + 2(n - 1)/k$; otherwise we have $L_2 = 0$ in which case $e \geqslant n - 1 + 2(n - 1)/k - 2m/k$. We shall show that, even in the case where $L_2 = 0$, $e \geqslant n - 1 + 2(n - 1)/k$.

If $m \leqslant k/2$ then $e \geqslant n - 2 + 2(n - 1)/k$ whereas if $m \geqslant k/2$ then there are $m + 1$ nodes in the chain from ROOT to BASE and every other node will be in some path from BASE to some leaf, the length of this path being at most $k - m$. Thus $n \leqslant m + 1 + (k - m)L = k + 1 + (k - m)(L - 1) \leqslant k + 1 + (L - 1)k/2$ where L is the total number of leaves. This implies that $L \geqslant 2(n - 1)/k - 1$. There are $n - 1$ tree edges and each leaf has outdegree at least one (a non-tree edge). Thus $e \geqslant n - 1 + L$ and so $e \geqslant n - 2 + 2(n - 1)/k$.

Thus, if we count only tree edges and allow for each leaf to have outdegree of precisely one, we get $e \geqslant n - 2 + 2(n - 1)/k$ (which is exactly one less than the minimum shown for the case when $L_2 \geqslant 1$.) But we shall show that there must be at least one edge unaccounted for (other than tree edges and one edge per leaf).

Since ROOT must be reachable, there must be an edge entering ROOT, say from vertex v. If we assume that the only non-tree edges are from leaves, then v is a leaf.

If v has outdegree $= 1$ then there is a chain, $v \to \text{ROOT} \to \text{BASE}$, of length $m + 1$ each of whose nodes has outdegree $= 1$, contradicting the assumption that C was maximal. Therefore v has outdegree > 1 which means there is an extra, previously unaccounted for, non-tree edge. Therefore, $e \geqslant n - 1 + 2(n - 1)/k$, always. ∎

COROLLARY 3. *The flower graph is optimal for k even, and is asymptotically optimal when k (odd) is fixed and n tends to infinity.*

DEFINITION 2. A circular graph is a directed graph that consists of 6 line segments that are symmetrically bonded at three junction nodes. Each line segment has m internal nodes and all edges within a line segment are directed as shown in Figure 3. The values of n, e and k are related via parameter m:

$$n = 6m + 3$$
$$e = 6m + 6$$
$$k = 3m + 1$$

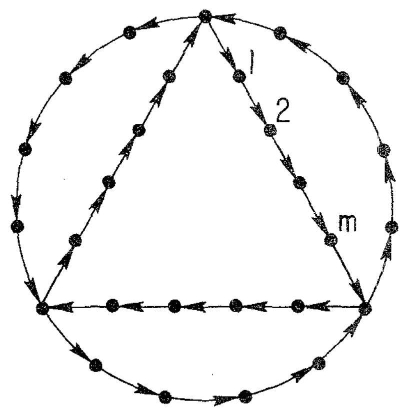

FIG. 3. A circular Graph

COROLLARY 4. (i) *When k is even, if a circular graph and a flower graph have the same values of n and k then they also have the same number of edges.*
(ii) *Circular graphs are always optimal.*

III. THE $\langle n, e \rangle$ PROBLEM

Given n and e ($n \leqslant e$), the problem is to find a directed graph with n nodes and at most e edges such that its diameter is minimum. Diameter is defined as the number of edges in the longest path between any two nodes. Let $k(n, e)$ be the minimum diameter.

DEFINITION 3. For $n \leqslant e$, a flower graph $G_{n,e}$ is constructed by taking one center and $e - n + 1$ loops and distributing the nodes to the loops as evenly as possible. Directions of all edges in one loop are the same.

THEOREM 5. *Let q be the quotient of $(n-1)/(e-n+1)$ and r the remainder. Then for $n < e$,*

$$\begin{aligned} k(n, e) &\leqslant 2q, && \text{if } r = 0 \\ &\leqslant 2q + 1, && \text{if } r = 1 \\ &\leqslant 2q + 2, && \text{if } r \geqslant 2 \end{aligned}$$

and for $n = e$,

$$k(n, n) = n - 1.$$

Proof. Construct a flower graph $G_{n,e}$. For $n < e$, if $r = 0$, each loop has exactly q nodes (other than the center). Thus the diameter is precisely $2q$. If $r = 1$, one loop has $q + 1$ nodes while other loops have q nodes each. Thus, the diameter is $2q + 1$. If $r \geqslant 2$, at least two loops have $q + 1$ nodes each while the others have at most $q + 1$ nodes each. Thus the diameter is $2q + 2$.

For $n = e$, $G_{n,e}$ reduces to a single loop with diameter $n - 1$. Clearly, it is an optimal graph. ▮

THEOREM 6. *For $n < e$, $k(n, e) \geqslant 2q$, if $r = 0$*

$$\geqslant 2q + 1, \text{ if } r > 0.$$

Proof. Let $k = k(n, e)$. By Theorem 2, $e \geqslant n - 1 + 2(n - 1)/k$ must hold. Solving this inequality for k results in the statemen of this theorem. ▮

COROLLARY 7. *The flower graph is optimal for $r = 0$ or 1. For $r \geqslant 2$, it may differ from the optimal by at most one.*

COROLLARY 8. (i) *When n and e are of the form $n = 6m + 3$, $e = 6m + 6$, the circular graph and the flower graph have the same diameter.*

(ii) *Circular graphs are always optimal.*

It should be pointed out that Theorem 6 answers a question raised at the end of Chapter 13 in [1].

IV. THE $\langle e, k \rangle$ PROBLEM

Given e, k ($k \leqslant e$), the problem is to find a directed graph with at most e edges whose diameter is not larger than k such that it has the maximum possible number of nodes. Let $n(e, k)$ be the maximum number of nodes.

DEFINITION 4. For $k < e - 1$, a flower graph $G_{e,k}$ is constructed as follows:

(i) form a loop of $\lceil k/2 \rceil + 1$ edges,

(ii) form loops of $\lfloor k/2 \rfloor + 1$ edges all having a node, CENTER, in common with the first loop,

(iii) the remaining edges (if there are at least two of them) are used to form one final loop, again having CENTER in common, and

(iv) assign the same directions to all edges in a loop.

THEOREM 9. *For $k < e - 1$, let q be the quotient of $e - \lceil k/2 \rceil - 1$ divided by $\lfloor k/2 \rfloor + 1$, and r the remainder. Then*

$$\begin{aligned} n(e, k) &\geqslant e - q, && \text{if } r = 0 \\ &\geqslant e - q - 1, && \text{if } r > 0. \\ n(e, e-1) &= e \\ n(e, e) &= e + 1. \end{aligned}$$

Proof. For $k < e - 1$, construct a flower graph $G_{e,k}$. Then q is the number of loops of $\lfloor k/2 \rfloor + 1$ edges each. If $r = 0$, the total number of loops is $q + 1$. Thus, the total number of nodes is $e - (q + 1) + 1 = e - q$. If $r > 0$, the total number of loops is $q + 2$. Thus, the total number of nodes is $e - (q + 2) + 1 = e - q - 1$.

For $k = e - 1$, construct a single loop with e edges with the same direction on all edges. Then it has e nodes and is clearly optimal. ∎

THEOREM 10. $n(e, k) \leqslant \lfloor (ek + k + 2)/(k + 2) \rfloor$, *for $k < e - 1$.*

Proof. Solving the inequality $e \geqslant n - 1 + 2(n - 1)/k$ in Theorem 2 for n, one obtains the desired inequality. ∎

COROLLARY 11. *The flower graph is optimal for k even, and is asymptotically optimal when k(odd) is fixed and n tends to infinity.*

Proof. If k is even, then $q = \lfloor (2e - k - 2)/(k + 2) \rfloor$.

If $r = 0$,

$$q = \frac{2e - k - 2}{k + 2} \quad \text{and} \quad e - q = \frac{ek + k + 2}{k + 2}.$$

If $r > 0$,

$$\begin{aligned} e - q - 1 &= e - \left\lfloor \frac{2e - k - 2}{k + 2} \right\rfloor - 1 \\ &= \left\lceil e - \frac{2e - k - 2}{k + 2} \right\rceil - 1 \\ &= \left\lceil \frac{ek + k + 2}{k + 2} \right\rceil - 1 \\ &= \left\lfloor \frac{ek + k + 2}{k + 2} \right\rfloor \end{aligned}$$

since $(ek + k + 2)/(k + 2)$ is not an integer. The result follows from Theorem 10. Similar arguments can be applied to the case when k is odd. ∎

COROLLARY 12. (i) *when k is even, circular graphs and flower graphs that have the same number of edges and the same value of k also have the same number of nodes.*

(ii) *circular graphs are always optimal.*

REFERENCES

1. C. BERGE, "The Theory of Graphs and Its Applications," Methuen, London, 1962.
2. C. BERGE, "Graphs and Hypergraphs," new ed., Elsevier and North-Holland, Amsterdam, 1975.
3. J. W. MOON, On the diameter of a graph, *Michigan Math. J.* **12** (1965), 349–351.
4. G. KATONA AND E. SZEMERÉDI, On a problem of graph theory, *Studia Sci. Math. Hungar.* **2** (1967), 23–28.
5. F. KRAMER AND H. KRAMER, Schranken für den Diameter eines Graphen, *Studia Sci. Math. Hungar.* **5** (1970), 283–287.