SIAM J. ALG. DISC. METH.
Vol. 2, No. 1, March 1981

0196-5212/81/0201-0010 $01.00/0

COMPUTING THE MINIMUM FILL-IN IS NP-COMPLETE*

MIHALIS YANNAKAKIS†

Abstract. We show that the following problem is NP-complete. Given a graph, find the minimum number of edges (fill-in) whose addition makes the graph chordal. This problem arises in the solution of sparse symmetric positive definite systems of linear equations by Gaussian elimination.

1. Introduction and terminology. A graph is a pair $G=(N, E)$, where N is a finite set of *nodes* and E, a set of unordered pairs (u, v) of distinct nodes, is a set of *edges*. Two nodes u and v are *adjacent* if $(u, v) \in E$. The *neighborhood* $\Gamma(v)$ of a node v is the set of nodes that are adjacent to v. The *degree* $d(v)$ of v is the number of nodes adjacent to v. A graph is a *clique* if every two nodes are adjacent. A set of nodes is *independent* if no two of them are adjacent.

If $S \subseteq N$ is a subset of nodes, the *subgraph* of G *induced* by S, denoted as $\langle S\rangle$, is the graph (S, E_S), where $E_S=\{(u, v) \in E \mid u, v \in S\}$. The graph $G-S$, formed by deleting a subset $S \subseteq N$ of nodes from G, is $\langle N-S\rangle$. A graph $G=(N, E)$ is *bipartite* if N can be partitioned into two sets P, Q of independent nodes; we will write the bipartite graph as (P, Q, E). The bipartite graph (P, Q, E) is a *chain graph* if the neighborhoods of the nodes in P form a chain; i.e., there is a bijection $\pi:\{1,2, \cdots,|P|\} \leftrightarrow P$ (an *ordering* of P) such that $\Gamma(\pi(1)) \supseteq \Gamma(\pi(2)) \supseteq \cdots \supseteq \Gamma(\pi(|P|))$. It is easy to see [Y] that then the neighborhoods of the nodes in Q form also a chain, and thus the definition is unambiguous.

A graph is *chordal* (or *triangulated*) if every cycle of length $\geqq 4$ has a *chord*, i.e., an edge connecting two nonconsecutive nodes of the cycle. Chordal graphs are important in connection with the solution of sparse symmetric positive definite systems of linear equations by Gaussian elimination [R]. From the symmetric $n \times n$ matrix $M=(m_{ij})$ of coefficients of such a system we can construct a graph $G=(N, E)$ with n nodes, where node v_i corresponds to the ith row and column of M and $(v_i, v_j) \in E$ iff $m_{ij} \neq 0$. The *elimination* of node v_i from G is performed by (1) adding edges so that $\Gamma(v_i)$ becomes a clique, and (2) deleting v_i from the augmented graph. The added edges correspond to the new nonzero elements that are created when we eliminate the ith variable, assuming no lucky cancellations. (See [R] for a detailed exposition of this graph-theoretic modeling.) If π is an ordering of N, the *fill-in* $F(\pi)$ produced by π is the set of new edges that are added when we eliminate $\pi(1)$ from G, then eliminate $\pi(2)$ from the resulting graph, $\pi(3)$ from the new graph, etc. The ordering π is a *perfect elimination ordering* if $F(\pi)=\varnothing$. Chordal graphs come into the picture because of the following two properties [R]. (1) A graph has a perfect elimination ordering if and only if it is chordal. Thus, "chordal" is a *hereditary* property (i.e., deleting nodes from a chordal graph does not violate the property), and every chordal graph has a node v such that $\langle\Gamma(v)\rangle$ is a clique; v is called a *simplicial* node. (2) If π is an elimination ordering of a graph $G=(N, E)$, then the augmented graph $G_\pi=(N, E \cup F(\pi))$ is chordal: π is a perfect elimination ordering of G_π.

In this paper we examine the problem of finding an elimination ordering which produces a minimum fill-in, or equivalently, finding the minimum set of edges whose addition renders the graph chordal. We shall show that this problem is NP-complete.

* Received by the editors June 26, 1980.

† Bell Laboratories, Murray Hill, New Jersey 07974.

(For an exposition of NP-completeness see [GJ].) The NP-completeness of the minimum fill-in problem was conjectured in [RTL] and [RT], but a proof had not been found, and it is one of the open problems in [GJ]. The version of the problem on directed graphs was shown to be NP-complete in [RT].

2. The reduction. We will make use of chain graphs. Two edges (u, v), (x, y) are said to be *independent* in a graph G if the nodes u, v, x, y are distinct and the subgraph of G induced by them consists of exactly these two edges. The following lemma from [Y] is easy to prove.

LEMMA 1. *A bipartite graph is a chain graph if and only if it does not contain a pair of independent edges.*

Let $G=(P, Q, E)$ be a bipartite graph. From G we construct another graph $C(G)=(N, E')$ by making P and Q cliques; i.e., $E'=E\cup\{(u, v)|u, v\in P\}\times \cup\{(u, v)|u, v\in Q\}$.

LEMMA 2. *Let G be a bipartite graph. $C(G)$ is chordal if and only if G is a chain graph.*

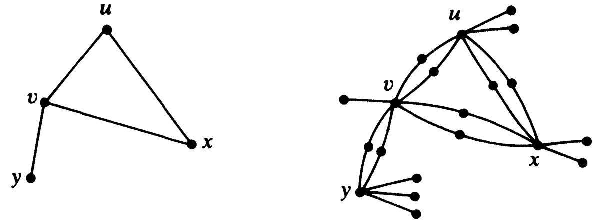
Proof (*only if*). Suppose that G is not a chain graph. Then it has two independent edges (u, v) and (x, y) by Lemma 1. Suppose without loss of generality that $u, x\in P$ and $v, y\in Q$. Then these two edges together with (u, x) and (v, y) form a chordless cycle of length 4 in $C(G)$.

(*if*). Suppose that G is a chain graph, and let π be an ordering of P such that $\Gamma(\pi(1))\supseteq\Gamma(\pi(2))\supseteq\cdots\supseteq\Gamma(\pi(p))$, where $p=|P|$. Since the property of being a chain graph is hereditary, it suffices to show that $C(G)$ has a simplicial node. The neighborhood of $\pi(p)$ in $C(G)$ is $\Gamma'(\pi(p))=\Gamma(\pi(p))\cup[P-\pi(p)]$. In $C(G)$ the subgraphs $\langle P-\pi(p)\rangle$ and $\langle\Gamma(\pi(p))\rangle$ are cliques, the latter because $\Gamma(\pi(p))\subseteq Q$ and $\langle Q\rangle$ is a clique. Also, since $\Gamma(\pi(p))\subseteq\Gamma(v)$ for every $v\in P$, all nodes of P are adjacent to all nodes of $\Gamma(\pi(p))$. Therefore $\langle\Gamma'(\pi(p))\rangle$ is a clique, and $\pi(p)$ is a simplicial node of $C(G)$. □

LEMMA 3. *It is* NP-*complete to find the minimum number of edges whose addition to a bipartite graph $G=(P, Q, E)$ gives a chain graph.*

Proof. The reduction is from the Optimal Linear Arrangement Problem. A linear arrangement of a graph $G=(N, E)$ is an ordering π of N. For an edge $e=(u, v)$ of G, let $\delta(e, \pi)=|\pi^{-1}(u)-\pi^{-1}(v)|$. The cost $c(\pi)$ of the linear arrangement π is $c(\pi)=\sum_{e\in E}\delta(e, \pi)$. The optimal linear arrangement problem is to decide, given a graph G and an integer k, whether there exists a linear arrangement π of G with cost $c(\pi)\leqq k$. This problem was shown to be NP-complete in [GJS].

Let $(G=(N, E); k)$ be an instance of the optimal linear arrangement problem. We construct a bipartite graph $G'=(P, Q, E')$ as follows. P has one node for every node of G (i.e., $P=N$); Q has two nodes e_1, e_2 for every edge e of G, and a set $R(v)$ of $n-d(v)$ nodes for every node v of N, where $n=|N|$ and $d(v)$ is the degree of v in G. If $e=(u, v)$ is an edge of G, then the nodes e_1, e_2 that correspond to e are adjacent to u and v. The nodes in $R(v)$ are adjacent to v. In Fig. 1 we show an example of this construction.



(a) *A graph G.*

(b) *The graph G'.*

FIG. 1

Let $l(G)$ be the minimum cost of a linear arrangement of G, and $h(G')$ the minimum number of edges whose addition to G' gives a chain graph. We claim that

$$h(G') = l(G) + \frac{n^2(n-1)}{2} - 2m, \tag{1}$$

where n, m are respectively the numbers of nodes and edges of G. Thus, $l(G) \leqq k$ iff $h(G') \leqq k + (n^2(n-1)/2) - 2m$.

First observe that an ordering π of N specifies uniquely a minimal set $H(\pi)$ of edges whose addition makes G' a chain graph with the neighborhoods of the nodes in $P(=N)$ ordered according to π. For every node x in Q, let $\sigma(x) = \max\{i \mid (x, \pi(i)) \in E'\}$. Then $H(\pi) = \{(x, \pi(j)) \mid x \in Q, j < \sigma(x)\} - E'$. Conversely, suppose that F is a set of edges such that $G'(F) = (P, Q, E' \cup F)$ is a chain graph and let π be an ordering of the nodes in P according to their neighborhoods in $G'(F)$. It is easy to see that $F \supseteq H(\pi)$, and therefore if F is a minimal augmentation then $F = H(\pi)$. Let $h(\pi) = |H(\pi)|$. In order to show (1), it suffices thus to show that for every ordering π of N, $h(\pi) = c(\pi) + (n^2(n-1)/2) - 2m$, where $c(\pi)$ is the cost of the linear arrangement π of G.

Let π be an ordering of N. For every $v \in N$ and $x \in R(v)$, $H(\pi)$ contains $\pi^{-1}(v) - 1$ edges incident to x. Let $e = (u, v)$ be an edge of G, and suppose without loss of generality that $\pi^{-1}(u) < \pi^{-1}(v)$. The number of edges of $H(\pi)$ incident to each of the two nodes e_1, e_2 that correspond to e is $\pi^{-1}(v) - 2 = \pi^{-1}(u) + [\pi^{-1}(v) - \pi^{-1}(u)] - 2 = \pi^{-1}(u) + \delta(e, \pi) - 2$; thus, the number of edges of $H(\pi)$ incident to e_1 and e_2 is $\pi^{-1}(v) + \pi^{-1}(u) + \delta(e, \pi) - 4$. Consequently,

$$\begin{aligned} h(\pi) &= \sum_{v \in N} \sum_{x \in R(v)} [\pi^{-1}(v) - 1] + \sum_{e=(u,v) \in E} [\pi^{-1}(v) + \pi^{-1}(u) + \delta(e, \pi) - 4] \\ &= \sum_{v \in N} (n - d(v))(\pi^{-1}(v) - 1) + \sum_{v \in N} d(v)\pi^{-1}(v) + \sum_{e \in E} \delta(e, \pi) - 4m \\ &= \sum_{v \in N} n[\pi^{-1}(v) - 1] + \sum_{v \in N} d(v) + c(\pi) - 4m \\ &= c(\pi) + \frac{n^2(n-1)}{2} - 2m, \end{aligned}$$

since $\sum_{v \in N} d(v) = 2m$, and

$$\sum_{v \in N} [\pi^{-1}(v) - 1] = 0 + 1 + 2 + \cdots + (n-1) = \frac{n(n-1)}{2}. \qquad \square$$

THEOREM 1. *The minimum fill-in problem is* NP-*complete.*

Proof. Follows from Lemmas 2 and 3. □

REFERENCES

[GJ] M. R. GAREY AND D. S. JOHNSON, *Computers and Intractability: A Guide to the Theory of NP-completeness*, W. H. Freeman, San Francisco, 1979.

[GJS] M. R. GAREY, D. S. JOHNSON AND L. STOCKMEYER, *Some simplified NP-complete graph problems*, Theoret. Comp. Sci., 1 (1976), pp. 237–267.

[R] D. J. ROSE, *A graph- theoretic study of the numerical solution of sparse positive definite systems of linear equations*, in Graph Theory and Computing, R. Read, ed., Academic Press, New York, 1973, pp. 183–217.

[RT] D. J. ROSE AND R. E. TARJAN, *Algorithmic aspects of vertex elimination on directed graphs*, SIAM J. Appl. Math., 34 (1978), pp. 176–197.

[RTL] D. J. ROSE, R. E. TARJAN AND G. S. LUEKER, *Algorithmic aspects of vertex elimination on graphs*, SIAM J. Comput., 5 (1976), pp. 266–283.

[Y] M. YANNAKAKIS, *Node-deletion problems on bipartite graphs*, SIAM J. Comput., 10 (1981).