C.1
IBM ASD & RESEARCH LIBRARY
SAN JOSE

A PROBLEM OF MAXIMUM CONSISTENT SUBSETS

R.E. Miller and D.E. Muller

IBM
RESEARCH CENTER

INTERNATIONAL BUSINESS MACHINES CORPORATION
YORKTOWN HEIGHTS, NEW YORK

RC-240

A PROBLEM OF MAXIMUM CONSISTENT SUBSETS

by

Raymond E. Miller and David E. Muller*

International Business Machines Corporation
Research Center
Yorktown Heights, New York

ABSTRACT: We consider the problem of obtaining a class C, of maximum cardinality p, of subsets $S_1, S_2, \ldots S_p$ of the set N of integers 1, 2, ..., n such that:

(i) No subset S_i is contained in any other subset S_j.
(ii) If T is a subset of N such that each pair of integers r, s contained in T is also contained in some S of C, then T itself is contained in at least one S of C.

The problem is formulated as one in linear graphs, various properties resulting from the graphical formulation are investigated, and a method of constructing a class of subsets is given and proved to be maximum. The maximum cardinality p is shown to be 3^k when $n = 3k$, $2 \cdot 3^{k-1}$ when $n = 3k-1$, and $4 \cdot 3^{k-1}$ when $n = 3k + 1$.

* University of Illinois and Consultant for IBM

Research Report
RC-240
March 25, 1960

A PROBLEM OF MAXIMUM CONSISTENT SUBSETS

I. Introduction:

We consider a problem of obtaining a maximum number of subsets of a set of objects such that the class of subsets satisfies certain conditions.[1]

A set formulation and a linear graph formulation of the problem are given. Various properties resulting from the graphical formulation are investigated; a method of constructing a class of subsets is given and proved to be maximum.

II. Set formulation of the problem:

Given the set N of integers $\{1, 2, \ldots, n\}$.

A. Find a set C of subsets $S_1, S_2, \ldots, S_p$ of N such that:

(i) No subset S_i is contained in any other subset S_j.

(ii) If S is a subset of N such that each pair of integers x_k, y_k contained in S is also contained in some S_k of C, then $S \subseteq S_j$ for at least one S_j in C.

Determine a set C of maximum cardinality p when n is fixed, and find this maximum p as a function of n.

---

1 The problem arose in obtaining an upper bound on the number of subsets of possibly equivalent states in a sequential machine in a state reduction procedure in "State Reduction for Sequential Machines," IBM Research Report, RC-121, June 15, 1959, by R. E. Miller.

B. Show how these maximum sets of type C may be formed.

C. Is the solution unique for each n up to permutations of integers?

D. Find the cardinalities of the S_i in the maximum sets.

E. Do all S_i have the same cardinality for given n?

Two examples are now given which give maximum sets for n = 8 and n = 7. This is accomplished by partitioning the set N into several parts and forming all subsets S_i having one and only one member in each partition.

Example 1: Let n = 8. The maximum number of sets is obtained by the partitioning:

123/456/78

The 18 resulting triples are:

147	247	347
148	248	348
157	257	357
158	258	358
167	267	367
168	268	368

Example 2: For n = 7, two sets which are maximum using partitioning are:

(a) 123/45/67 gives the triples:

146	246	346
147	247	347
156	256	356
157	257	357

and

(b) 123/4567 gives the pairs:

14	24	34
15	25	35
16	26	36
17	27	37

III. Linear Graph Formulation - The Maximum Complete Subgraph Problem:

A. Given n, form an n-node linear graph G having the largest number of maximum complete subgraphs (MCSG's) imbedded in the n-node graph.

The ideas and concepts used in this formulation will be developed in Section IV of this paper. There the problem will be stated explicitly and a solution found.

B. Show how the graph G is formed.

C. Is the graph G unique up to a relabelling of the nodes for each n?

D. Find the number of nodes in each MCSG of G.

E. Does each MCSG of G contain the same number of nodes?

IV. The Problem Solution:

The linear graph concepts of the problem are now defined. Various interesting properties are investigated, and a method for constructing a solution to the MCSG problem is given.

Definition 1: an n-node graph G is defined by:

(a) A set $\sigma(G)$ of n objects a, b, . . called nodes, and

(b) a set $\lambda(G)$ consisting of some unordered pairs (a. b) of distinct elements a and b of $\sigma(G)$. Such unordered pairs will be called arcs. No distinction will be made between an arc (a, b) and an arc (b, a).

Definition 2: A graph H is said to be a subgraph of G if $\sigma(H) \subseteq \sigma(G)$ and $\lambda(H) \subseteq \lambda(G)$.

Definition 3: A subgraph H of G is said to be normal if $\lambda(H)$ consists of every pair (a, b) of $\lambda(G)$ such that a and b are both in $\sigma(H)$.

A normal subgraph H of G is uniquely determined from a specification of G and the nodes of H.

Definition 4: Given any graph G, we construct its complementary graph G' by letting $\sigma(G') = \sigma(G)$, and $\lambda(G')$ consist of all unordered pairs (a, b) of elements of $\sigma(G)$ which are not in $\lambda(G)$.

Property 1: (Duality Property) For every theorem concerning graphs there is a dual theorem concerning complementary graphs.

Definition 5: A graph G is said to be complete if $\lambda(G)$ contains all unordered pairs which may be constructed from elements of $\sigma(G)$. A graph H is said to be nodal if $\lambda(H)$ is empty.

The complementary graph to a nodal graph is complete and vice-versa. Any set S of nodes uniquely determines a complete graph G and a nodal graph H such that $\sigma(G) = S$ and $\sigma(H) = S$.

Property 2: If H is a normal subgraph of G, then H' is a normal subgraph of G'.

Property 3: Any complete subgraph of a graph G is a normal subgraph of G.

Definition 6: A complete subgraph of a graph G is said to be an MCSG (maximum complete subgraph) if it is not a subgraph of any other complete subgraph of G.

Definition 7: A nodal normal subgraph of a graph G is said to be an MNSG (maximum nodal normal subgraph) if it is not a subgraph of any other nodal normal subgraph of G.

Property 4: If H is an MCSG of G, then H' is an MNSG of G' and vice-versa.

Definition 8: If G is an n-node graph for which $n \geq 1$, let $\nu(G)$ be the number of MCSG's of G and $\nu'(G)$ the number of MNSG's of G.

Property 5: $\nu'(G) = \nu(G')$.

Definition 9: Define the MCSG problem to consist of finding an n-node graph K having as many MCSG's as any other n-node graph. In other words, for fixed n, we wish to find K which will maximize $\nu(K)$.

This statement is the same as that in Section III. In Section V we show that it is also equivalent to the statement in Section II.

Definition 10: Given two graphs G_1 and G_2 whose nodes are distinct, let $G = G_1 + G_2$ (the cardinal sum) be the graph formed by letting $\sigma(G) = \sigma(G_1) + \sigma(G_2)$ and $\lambda(G) = \lambda(G_1) + \lambda(G_2)$.

This form of addition is associative and commutative.

Definition 11: The graph G is said to be disconnected if it may be expressed as the sum of two subgraphs G_1 and G_2.

Theorem 1: If $G = G_1 + G_2$, then G_1 and G_2 are normal subgraphs of G and

$$\nu(G) = \nu(G_1) + \nu(G_2)$$

$$\nu'(G) = \nu'(G_1) \quad \nu'(G_2)$$

Proof: Any MCSG of G must be a subgraph of either G_1 or G_2. Any MCSG of either G_1 or G_2 is an MCSG of G. Hence $\nu(G) = \nu(G_1) + \nu(G_2)$.

Let A be an MNSG of G. Then $\sigma(A)$ may be written as $\sigma(A) = \sigma(A_1) + \sigma(A_2)$, where $\sigma(A_1)$ and $\sigma(A_2)$ lie in $\sigma(G_1)$ and $\sigma(G_2)$ respectively. Since A is nodal and normal in G, if we let A_1 and A_2 be the corresponding nodal subgraphs of G_1 and G_2, they must also be normal in G_1 and G_2 respectively. To show that they are maximum we note that if nodal normal subgraphs B_1 and B_2 of G_1 and G_2 can be found such that A_1 is a subgraph of B_1 and A_2 is a subgraph of B_2, then $B_1 + B_2$ forms a nodal normal subgraph of G, and A can only be maximum if $A = B_1 + B_2$. Since $B_1 + B_2$ is unique we see that there is a one-to-one correspondence between pairs of MNSG's in G_1 and G_2 and MNSG's of G. Hence $\nu'(G) = \nu'(G_1) \ \nu'(G_2)$.

Theorem 2: Each node of G is in at least one MCSG of G and in at least one MNSG of G.

Proof: Each node of G is simultaneously a complete subgraph of G and a nodal normal subgraph of G. Hence, it is a subgraph of at least one maximum subgraph of each of these types.

Theorem 3: Each MCSG of G shares at most one node with each MNSG of G.

Proof: If nodes a and b are nodes of an MCSG then (a, b) is in $\lambda(G)$ while if they are nodes of an MNSG, then (a, b) is not in $\lambda(G)$.

Theorem 4: If (a, b) is in $\lambda(G)$, then there is at least one MCSG of G containing both a and b.

Proof: The two nodes a and b determine a two node complete subgraph of G. It must be a subgraph of at least one MCSG.

We will now proceed with the development of several results which lead to the following four solutions to the MCSG problem which are the only solutions for $n > 1$, up to a permutation of nodes.

Define C_j as the complete j-node graph and let k be a positive integer. Then depending on n, the solution K will be shown later to be one of the following.

(a) When $n = 3k$, the solution K_1 is such that $K'_1 = k \times C_3$. (The cardinal sum of K complete three node subgraphs).

(b) When $n = 3k - 1$, the solution K_2 is such that $K'_2 = (k - 1)C_3 + C_2$.

(c) When $n = 3k + 1$, a solution K_3 is such that $K'_3 = (k - 1)C_3 + 2C_2$.

(d) When $n = 3k + 1$, a second solution K_4 is such that $K'_4 = (k - 1)C_3 + C_4$.

Define the number of MCSG's for each of these graphs as g(n). From Theorem 1 and the formula

$\nu'(C_j) = j$, we obtain the g(n) values shown in Table I.

$$\boxed{\begin{aligned} g(n) &= \nu'(K_1) = 3^k \text{ or} \\ g(n) &= \nu'(K_2) = 2 \cdot 3^{k-1} \text{ or} \\ g(n) &= \nu'(K_3) = \nu'(K_4) = 4 \cdot 3^{k-1} \end{aligned}}$$

Table I

Definition 12: Let f(n) be the maximum value of $\nu(G)$ over all n-node graphs G.

The solution K is thus any n-node graph for which $\nu(K) = f(n)$. Our object is to show that $f(n) = g(n)$, and that K_1, K_2, K_3, and K_4 are the only graphs with this property.

Definition 13: Let G be an n-node graph and let S be any subset of $\sigma(G)$. The graph G - S is then defined by the rules:

(a) $\sigma(G - S) = \sigma(G) - S$, and

(b) $\lambda(G - S)$ consists of all pairs in $\lambda(G)$ which contain no members of S.

Theorem 5: If H is any nodal normal subgraph of G and B is the set of nodes b_j such that (a_i, b_j) is in $\lambda(G)$ for some a_i in $\sigma(H)$, then $\nu'\{G - [\sigma(H) + B]\}$ = the number of MNSG's of G containing H.

Proof: Any MNSG, say J, of G which contains H also defines an MNSG of $G - [\sigma(H) + B]$, namely $J - \sigma(H)$, since J can contain no nodes of B. Also any MNSG of $G - [\sigma(H) + B]$, say R, can be used to define an MNSG of G, namely R + H.

Definition 14: An n-node graph P_n where $n \geq 1$ is called an open path if $\lambda(P_n)$ contains just (a_1, a_2), $(a_2, a_3), \ldots, (a_{n-1}, a_n)$. An n-node graph Q_n is called a closed path if $\lambda(Q_n)$ contains the n-1 pairs listed above and (a_n, a_1) as well. We note that $P_1 = Q_1 = C_1$ is an isolated node. Also, we have $P_2 = Q_2 = C_2$, but $P_3 \neq Q_3 = C_3$.

Definition 15: Define $r(n) = \nu'(P_n)$ and $s(n) = \nu'(Q_n)$.

Theorem 6: When $n \geq 4$, the function $r(n)$ satisfies the difference equation $r(n) = r(n - 2) + r(n - 3)$ and when $n \geq 6$, the function $s(n)$ satisfies $s(n) = 3r(n - 3) - r(n - 5)$ and hence when $n \geq 5$ it satisfies $s(n) = s(n - 2) + s(n - 3)$.

Proof: Consider P_n first. The number of MNSG's containing a_1 is $r(n - 2)$ by Theorem 5. Those not containing a_1 must contain a_2 and the number of these is $r(n - 3)$ by Theorem 5. Hence $r(n) = r(n - 2) + r(n - 3)$.

Next consider Q_n. The number of MNSG's containing a_1 is $r(n - 3)$ by Theorem 5. The number containing a_2 is $r(n - 3)$ and the number containing a_3 is $r(n - 3)$. The overlapping of these categories is characterized by just those MNSG's containing both a_1 and a_3. They number $r(n - 5)$ and since they were counted twice we subtract them, giving $s(n) = 3r(n - 3) - r(n - 5)$. The restrictions on n result from the corresponding restriction in the definition of ν (G).

The difference equation $s(n) = s(n - 2) + s(n - 3)$ is a simple consequence of the linearity of the two previous expressions. This derivation, however, is only valid if $n \geq 9$ and to obtain it for $n \geq 5$ we must inspect individual cases. These are given in Table II.

Definition 16: Define the degree of a node a of a graph G as the number of pairs in λ (G) containing a.

Definition 17 Define the degree of a graph G as maximum degree over all nodes in σ(G).

A graph of degree two or less must be a cardinal sum of open and closed paths The proof rests on conventional graph theory and will not be treated here. We note that K'_1, K'_2, and K'_3 are of degree two or less, while K'_4 is of degree three.

Theorem 7: If K is any n-node graph which is a solution to the MCSG problem and K' is of degree greater than two, then $f(n) \leq f(n-1) + f(n-4)$.

Proof: Let a be a node of K' of degree three or greater. Those MNSG's of K' containing a are no more numerous than $f(n-4)$ by Theorem 5. On the other hand, if node a is not contained in a given MNSG, then this MNSG of K' is also an MNSG of K' - a. Such MNSG's can thus be no more numerous than $f(n-1)$.

Theorem 8: Given any integer m we can always find an integer $p > m$ for which all p-node solutions K will have complementary graphs K' of degree two or less.

Proof: Assume we can find K' of degree greater than two for all $p > m$. Pick α so that $f(j) \leq \alpha v^j$ for all $j \leq m + 3$. Take $v = 1.4 < 3^{1/3}$. Assume that $f(j) \leq \alpha v^j$ whenever $m \leq j \leq h$. Then $f(h) \leq f(h-1) + f(h-4) \leq \alpha v^{h-1} + \alpha v^{h-4} < \alpha v^h$, by direct computation. Thus we obtain $f(p) \leq \alpha v^p$ for all p. We know, however, from the g(n) values of Table I that $f(3k) \geq g(3k) = 3^k = (3^{1/3})^{3k}$, so if we make p large enough we obtain a contradiction by virtue of $v < 3^{1/3}$. Therefore K' is of degree no greater than two for some $p > m$.

Theorem 9: If K is a solution to the MCSG problem and K' is of degree two or less, then K' is the cardinal sum of subgraphs containing no more than three nodes.

Proof: We will show that when $n \geq 4$ we always have $r(n) < g(n)$ and $s(n) < g(n)$.

Hence by Theorem 1 if an open of closed path of more than three nodes existed in K', we could replace this path by either K'_1, K'_2, or K'_3.

To show that $r(n) < g(n)$ and $s(n) < g(n)$ we check in Table II that these relationships hold when $4 \leq n \leq 8$. Also when $n \geq 4$ we have $g(n) = 3g(n-3)$ and when $n \geq 2$ we have $g(n) < 2g(n-1)$, because of the formulas given for $g(n)$ in Table I. Hence for $n \geq 4$ we have $r(n) = r(n-2) + r(n-3) \leq g(n-2) + g(n-3) < 3g(n-3) = g(n)$, by induction on n. The same argument also yields $s(n) < g(n)$.

n	r(n)	s(n)	g(n)
1	1	1	--
2	2	2	2
3	2	3	3
4	3	2	4
5	4	5	6
6	5	5	9
7	7	7	12
8	9	10	18

Table II

Theorem 10: The solution referred to in Theorems 8 and 9 is unique up to a permutation of nodes and is either K_1, K_2, or K_3, depending on p.

Proof: By Theorem 9 the K' of Theorem 8 is a sum of graphs of degree two or less and hence is a sum of graphs containing no more than three nodes. If, however, K' contains as many as three two node graphs which are terms in the sum, then these three may be replaced by the sum of two three node graphs with the resulting increase in $\nu(K)$ by a factor of 9/8. Hence we can have no more than two two-node graphs.

Isolated nodes of K' may always be eliminated if $p > 1$, since $g(n)$ is a monotone increasing function of n, and if k isolated nodes were present we would have $f(n) = g(n - k)$. Therefore, the only possible solutions are K_1, K_2, or K_3 if K' is of degree two or less.

Theorem 11: K_1 is a unique solution to the MCSG problem if $n = 3k$.

Proof: Apply Theorem 8 and let $m = 3k + 1$. Thus $p > 3k + 1$. It follows from Theorem 10 that the corresponding K is either K_1, K_2, or K_3. Hence K' may be written as a sum containing the term kC_3. This part of the disconnected graph K' must, in itself, define a solution to the MCSG problem for $n = 3k$ because of Theorem 1. This solution is unique since, from Theorem 10, K is unique.

Theorem 12: K_3 and K_4 are the only solutions when $n = 3k + 1$, $k \geq 1$.

Proof: Consider K_4 first, with $n = 3k + 1$ and $g(n) = 4 \cdot 3^{k-1}$. If K' for $n = 3k + 1$ is of degree greater than two, then from Theorem 7 we have $f(n) \leq f(n - 1) + f(n - 4)$. Using Theorem 11, this means $f(n) \leq 3^k + 3^{k-1} = 4 \cdot 3^{k-1}$. Therefore, $g(n) = f(n)$ and K_4 is a solution if any solution K with K' of degree greater than two exists. Since K_3 is the only possible solution K with K' of degree two or less, and since $\nu(K_3) = \nu(K_4)$ we see that K_3 and K_4 are both solutions.

No other solution K can occur for this n, since K' would have at least one node a of degree three or greater. But $f(n) = f(n - 1) + f(n - 4)$, and therefore $K - a = K_1$. Therefore K would be K_4.

Theorem 13: K_2 is a unique solution when $n = 3k - 1$, $k \geq 1$.

Proof: Suppose a solution K exists for this n, which is not K_2. Then K' must be of degree greater than two by Theorem 10. This means that $f(n) \leq f(n-1) + f(n-4) = 4 \cdot 3^{k-2} + 4 \cdot 3^{k-3} = 16 \cdot 3^{k-3}$. However, $g(n) = 2 \cdot 3^{k-1} = 18 \cdot 3^{k-3}$ and since $f(n) \geq g(n)$ we see that K' cannot be of degree greater than two. Hence, K_2 is the unique solution if $k \geq 3$. It is also a unique solution if $1 \leq k < 3$, because K'_2 in these cases is a term in the sum forming K'_2 for larger k.

V. Equivalence of the two formulations of the MCSG problem:

To show that the formulations of the MCSG problem given in Sections II and III are equivalent, we must show an isomorphism between n-node graphs and sets C satisfying the conditions (i) and (i) of Section IIA.

1. Given an n-node graph G, we may set the elements of $\sigma(G)$ in one-to-one correspondence with the integers $1, 2, \ldots, n$. We let $p = \nu(G)$. If $A_1, A_2, \ldots, A_p$ are the MCSG's of G, then we let $S_1, S_2, \ldots, S_p$ be the sets of integers corresponding to $\sigma(A_1), \sigma(A_2), \ldots, \sigma(A_p)$.
   Conditions (i) and (ii) of Section IIA then follow from the definition of MCSG and Theorem 4.

2. Given a set C satisfying (i) and (ii) of Section IIA, we construct an n-node graph G by the following rules.
   Let $\sigma(G) = \{a_1, a_2, \ldots, a_n\}$, any set of n objects. Put (a_i, a_j) in $\lambda(G)$ if and only if there is an element S_k of C such that i and j are both in S_k. Given any element S_k of C we

can then form a set of nodes S_k corresponding to it. From the construction of $\lambda(G)$, there must be a complete subgraph A_k of G such that $\sigma(A_k) = S_k$.

If A_k is any complete subgraph of G, then since the conditions of Section IIA (ii) are satisfied for the set of integers corresponding to the nodes of A_k we see that $\sigma(A_k)$ corresponds to an S such that $S \subseteq S_j$ for some S_j in C. Since set inclusion among subsets S corresponds to the subgraph relationship among complete subgraphs of G, we see that Section IIA (i) implies that the complete subgraphs A_k corresponding to the sets S_k must be maximum.

RC-