NATIONAL BUREAU OF STANDARDS REPORT

6883

MATHEMATICAL RESEARCH RELATED TO
INFORMATION SELECTION SYSTEMS

by

Alan J. Goldman (Project Manager), Bernice K. Bender,
John Edmonds, Karl Goldberg, and Rosalind B. Marimont

Final Report to the
National Science Foundation
(Grant No. G-7579)

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

THE NATIONAL BUREAU OF STANDARDS

Functions and Activities

The functions of the National Bureau of Standards are set forth in the Act of Congress, March 3, 1901, as amended by Congress in Public Law 619, 1950. These include the development and maintenance of the national standards of measurement and the provision of means and methods for making measurements consistent with these standards; the determination of physical constants and properties of materials; the development of methods and instruments for testing materials, devices, and structures; advisory services to government agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; and the development of standard practices, codes, and specifications. The work includes basic and applied research, development, engineering, instrumentation, testing, evaluation, calibration services, and various consultation and information services. Research projects are also performed for other government agencies when the work relates to and supplements the basic program of the Bureau or when the Bureau's unique competence is required. The scope of activities is suggested by the listing of divisions and sections on the inside of the back cover.

Publications

The results of the Bureau's research are published either in the Bureau's own series of publications or in the journals of professional and scientific societies. The Bureau itself publishes three periodicals available from the Government Printing Office: The Journal of Research, published in four separate sections, presents complete scientific and technical papers; the Technical News Bulletin presents summary and preliminary reports on work in progress; and Basic Radio Propagation Predictions provides data for determining the best frequencies to use for radio communications throughout the world. There are also five series of nonperiodical publications: Monographs, Applied Mathematics Series, Handbooks, Miscellaneous Publications, and Technical Notes.

A complete listing of the Bureau's publications can be found in National Bureau of Standards Circular 460, Publications of the National Bureau of Standards, 1901 to June 1947 ($1.25), and the Supplement to National Bureau of Standards Circular 460, July 1947 to June 1957 ($1.50), and Miscellaneous Publication 240, July 1957 to June 1960 (Includes Titles of Papers Published in Outside Journals 1950 to 1959) ($2.25); available from the Superintendent of Documents, Government Printing Office, Washington 25, D. C.

NATIONAL BUREAU OF STANDARDS REPORT

NBS PROJECT		NBS REPORT
1101-40-11412	30 June 1961	6883

MATHEMATICAL RESEARCH RELATED TO
INFORMATION SELECTION SYSTEMS

by

Alan J. Goldman (Project Manager), Bernice K. Bender,
John Edmonds, Karl Goldberg, and Rosalind B. Marimont

Final Report to the
National Science Foundation
(Grant No. G-7579)

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

CONTENTS

PAGE

1. INTRODUCTION ---------- 1

2. EFFICIENT FORMULATION OF SEARCH PROBLEMS ---------- 5

3. ON A CONJECTURE OF CHITTENDEN ---------- 12

4. ON A CONJECTURE OF QUINE ---------- 24

5. MAXIMAL ISOLATED SETS IN BIPARTITE GRAPHS ---------- 30

6. ONE-DIMENSIONAL CATALOGING OF INCOMING MATERIAL --------- 38

7. EXPECTED LENGTHS OF MONOTONE CHAINS ---------- 41

8. APPROXIMATIONS OF FINITE SET STRUCTURES ---------- 53

Mathematical Research Related to Information Selection Systems

Alan J. Goldman (Project Manager), Bernice K. Bender, John Edmonds, Karl Goldberg, and Rosalind B. Marimont

1. INTRODUCTION

This report contains a description of work performed under NSF Grant G-7579 during fiscal 1960 on mathematical research relevant to information selection systems. We have considered three areas of mathematical problems related to mechanization of the information retrieval process:

(1) Efficient Formulation of Search Questions,

(2) Rapid Classification of Incoming Documents on a Given One-Dimensional Scale,

(3) Categorizing and Manipulating Documents and Their Relationships.

Work on the "search formulation" was carried out by Mrs. Bernice K. Bender and Dr. Alan J. Goldman. This research is described in "medium detail" in Section 2. The results were of two kinds. First, an algorithm for partial simplification of Boolean functions was devised and translated into a digital computer code. This work was presented to the American Mathematical Society (Notices AMS 7 (1960), p. 86), and its technical aspects are given in NBS Report 6859 ("Computer Simplification of Boolean Functions") which has already been submitted to the National Science Foundation; these aspects are not repeated in the present report. A paper on this material has been submitted to a technical journal, and is now being revised and enlarged. Computer experimentation with the algorithm is being carried on under other

auspices. Much of the analysis is also applicable to the more general task of selecting a "minimum-cost" covering of a given set from a prescribed family of subsets.

Second, solutions were found to two theoretical questions which had appeared in the technical literature. This research is reported in detail in Sections 3 and 4, and is to be incorporated in a longer paper under preparation. The material of Section 3 has been presented to the American Mathematical Society (Notices AMS 7 (1960), p. 715); John Edmonds has examined some of this material in a more general (graph-theoretic) setting, and his investigation is set forth in Section 5 of this report.

In connection with the "rapid classification" problem, Dr. Karl Goldberg investigated the efficacy of a proposed method for ordering a flow of incoming data by monotonic chains. The results contradict previous conjectures. A summary of this work appears in Section 6 below, while the details are given in Section 7. The material is to be included in a longer paper now being prepared.

As one aspect of our work on the "categorization" problem, it was found appropriate to support in part the preparation of a paper on "Graph-Theoretic and Boolean Method in Computer Programming." The reasons were (a) that the mathematical techniques involved are relevant to the "categorization and manipulation" problem, (b) that advanced programming techniques will almost surely be required for any sophisticated retrieval system, and (c) that computer programs and flow charts themselves form

interesting and important information structures from which it is often quite difficult (e.g., in debugging) to extract needed data. The paper has been published (SIAM Review 2 (1960), p. 259), and reprints have been submitted to the National Science Foundation; its contents are therefore not discussed in this report.

Brief examinations were made of several formulations of the "categorization" problem. One version turns out to be of roughly the same order of difficulty as the classical coloring problems of graph theory. Another formulation leads to a "minimum-cost" covering problem, so that the work mentioned above in connection with Boolean simplification is also useful here. This material is too fragmentary to be reported. John Edmonds observed that relationships between documents in a collection are frequently intransitive; since intransitive relations are considerably more difficult to manipulate than transitive ones, he made a preliminary study of the problem of approximating a given intransitive relation by a suitably-chosen transitive relation. This study is reported in Section 8.

The duration of the project was too short for most of the investigators to obtain results they considered sufficiently substantial and complete to warrant separate publication. On the other hand, the grant provided an essential initial stimulus for several promising lines of research which are now being continued under other auspices and are proving quite fruitful. (One example not mentioned above is the recent work of John Edmonds on "minimum cost" covering problems; cf. Notices AMS 8 (1961), p. 152). These two factors explain the many references above to "longer papers being

prepared" as compared with the single reference to a paper already published. The factors also explain in part the tardiness of this report, which was delayed in hopes of including the "completed story", at least up to the "first publication" stage, of more of the research supported by the grant. The press of the investigators' other duties has prevented this, but of course the National Science Foundation will be informed (and furnished reprints) of subsequent publications resulting from the grant.

2. EFFICIENT FORMULATION OF SEARCH QUESTIONS[1]

The activity reported below deals with efficient formulation of search questions in cases in which the relevant data (about either document topic or document content) can be characterized in terms of n descriptors and all logical combinations of them. If for example the descriptors in question are "cat" and "dog" (so that n=2), then an illustrative situation would be one for which it sufficed to classify documents as dealing with "cats and dogs", or "cats but not dogs", or "dogs but not cats", or "neither dogs nor cats". In such circumstances the description of a class of documents to be retrieved is simply a Boolean function of n variables, and the problem of efficiently representing such descriptions leads to the problem of devising an algorithm for finding the simplest forms of Boolean functions of many variables. We have been especially concerned with methods applicable on small digital computers, or computers for which use of external memory is impossible or particularly undesirable.

A digital computer technique requiring relatively little storage and no external storage was devised for the partial simplification of Boolean functions. Functions of as many as 14 variables have been treated on the SEAC computer (2048-word internal memory, no external memory, 44-bit word length). The programming for SEAC and the actual machine usage were supported in part by the Navy Department (Bureau of Ordnance), and in part from the NSF funds for the NBS Research Information Center and Advisory Service on Information Processing. All other work reported below was supported under NSF Grant No. G-7579.

---

[1] The material of Sections 2-4 is due to Mrs. Bernice K. Bender and Dr. Alan J. Goldman.

We have found the "geometric" or "topological" approach to Boolean functions to be far more useful than the "algebraic" formulation, though perhaps this is due rather to the nature of our particular mathematical intuitions than to any inherent superiority in the former method. In the geometrical approach, the n variables of the Boolean function (each of which can be either "true" or "false") become n coordinates (each of which can take on either the value 1 or the value 0), so that the set of all possible combinations of true-or-false values for the variables becomes the set of 2^n vertices of the unit cube I_n in n-dimensional space. A particular Boolean function can be identified with the collection of those true-or-false combinations of the variables which verify the function; thus geometrically a Boolean function of n variables is simply a subset of the vertices of I_n.

If f is a fixed Boolean function, then a face of I_n whose vertices all lie in f will be called a cell. A collection C of cells, such that every vertex of f lies in at least one member of C, is called a covering of f. Any such covering corresponds to a normal disjunctive form for f, i.e., a disjunction of conjunctions of variables and their negatives, each conjunction corresponding to a single cell of the covering. For example, the covering made up of all the individual vertices in f corresponds to the "completely expanded normal form".

In our work (as in that of many workers), the problem of "simplifying" a Boolean function is restricted to the search for a "simplest" normal disjunctive form, which in the geometric language becomes the task of finding a covering which is "minimal" in some prescribed sense. Both on esthetic grounds and from the viewpoint of applications to switching-circuit design, it is necessary that the definition of "minimal"

be such as, roughly speaking, to encourage coverings with few cells, and coverings with mostly cells of high dimension. To realize these ends, we introduce an unspecified decreasing function g(d) as the cost of a d-dimensional cell; the cost of a covering is the sum of the costs of its cells, and a "minimal" covering is one with the least possible cost. Thus we are concerned (as was Quine in his basic papers in the field) with that part of the simplification process which is valid independent of more precise specification of the cost function; this is one aspect of what was meant above in speaking of a computer program for partial simplification of Boolean functions.

A cell contained in no other cell is called basic ("maximal" would be more in keeping with customary mathematical usage); such cells correspond to what Quine calls prime implicants of the Boolean function. Clearly a minimal covering is composed of basic cells only. Furthermore, suppose some vertex in f lies in only one basic cell; such a cell is necessarily in every covering by basic cells, hence in every minimal covering, and so such cells are called essential.

Although the details of the computer program are of importance in establishing the feasibility of the techniques for a small computer, they will not be repeated here; a full description is given in NBS Report 6859, (1960), a preprint of a paper "Computer Simplification of Boolean Functions" by B. K. Bender, A. J. Goldman and R. B. Thomas[2] which has been submitted to a technical journal. In summary, Phase 1 of the program determines all the essential cells of the Boolean function. Note that no further effort need be made to ensure "covering"

---

[2] The last author is a member of the NBS Data Processing Systems Division.

of the vertices in essential cells. Phase 2A determines, for each vertex in f but not in an essential cell, the list of basic cells containing that vertex. If x and y are two such vertices, and if every basic cell which contains x also contains y, then y will automatically be "covered" when a basic cell is selected to cover x in the course of forming a minimal covering. Phase 2B of the program detects such vertices "y" and removes them from the set of vertices requiring covering. Thus Phases 1 and 2 together produce what amounts to Quine's "reduced table of prime implicants" (Amer. Math. Monthly 59, p. 528).

Phase 3 of the program applies certain decision rules to select certain basic cells for inclusion in a minimal covering, and to eliminate other basic cells from further consideration. The rules are not repeated in this report, but three aspects of Phase 3 will be noted here.

First, the aim is that of determining just one minimal covering instead of all such coverings; thus eliminations can be made more freely.

Second, at each stage of the algorithm one considers only the basic cells containing some one vertex of f "remaining to be covered" (i.e., only prime implicants implied by some one abscissa of the Quine table). This has the important advantage of avoiding the difficulty (noted by Quine, Amer. Math. Monthly 66, p. 758) of dealing with the potentially enormous total number of basic cells of the Boolean function; also, in such situations a cell can be specified by only n binary digits. It therefore seems plausible to carry on the simplification process as far as possible with such steps before having recourse to other methods. In general, however, these

single-vertex considerations do not lead all the way to a minimal covering (this is another reason for our use of the phrase "partial simplification").

For the third point, note first that our process is not one of begining with some covering and "modifying it toward minimality", but rather one of gradually building up a single minimal covering (by the selections made in Phase 3) and its complement (by the eliminations made in Phase 3). In such a process it seems almost inevitable that one will generally reach logical branchings at which further progress can be made only by supposing that such-and-such a cell is to be accepted for or rejected from the covering, and then tracing out the consequences of this supposition, presumably encountering other branchings en route. A simple example is given by the Boolean function f of three variables whose vertices V_i (i=1,2,...,6) are

(0,0,0), (1,0,0), (1,0,1), (1,1,1), (0,1,1), (0,1,0);

there are precisely two minimal coverings, one by the edges V_1V_2, V_3V_4 and V_5V_6, and the other by the edges V_2V_3, V_4V_5, V_6V_1. Without a branching there seems no way to "get at" either covering, but a single branching very quickly leads to a minimal covering. In general, however, and for functions of many variables, it seems clear that most of the paths explored in this way will turn out to be false scents, and that the number of such paths will be very great. It therefore seems imperative to carry the simplification process as far as possible before making any branchings, and Phase 3 can be viewed as an attempt in this direction. Because the selection and elimination decisions made in Phase 3 are irrevocable, the decision rules require careful justification (the theorem is stated in op. cit.), and the necessary proofs turn out to be distinctly non-trivial and interesting in their own right.

Our most recent work involved an examination of the effect of varying the order of the steps in Phase 3, and culminated in a proof of the earlier conjecture that the effect is essentially nil, so that there are no problems of "search strategy". A precise statement and proof are still to be detailed, so the result is not absolutely established.

Because SEAC is very slow by modern standards and is not maintained on a regular "production" basis, the computer time available to us permitted little more than code-checking. Under the present NSF grant a start has been made on recoding the program for the Bureau's IBM 704 computer; we hope in a subsequent phase of the project to complete the recoding and to obtain a fair amount of computational experience.

Insight into the geometric structure of Boolean functions is very desirable, both for intrinsic mathematical interest and for providing a better foundation for applications. We feel that current understanding in this field is rather unsatisfactory, since it is based on examination of functions of relatively few variables, whose typical structure-types may be quite atypical in higher dimensions. For example, in functions of few variables the core (the set of all vertices lying in at least one essential cell) generally makes up a substantial part of the function, so that workers in the field of Boolean function simplification have stressed the importance of early location of the core. We conjecture, however, that for functions of many variables which contain many of the vertices of the cube, the core is generally a negligible proportion of the function.

Our preliminary attempts to gain insight into the structure of functions of many variables have led to solutions to two problems posed

in the technical literature. First, we have obtained a simple proof of the conjecture of E. W. Chittenden (Office of Ordnance Research, Transactions of the Third Conference of Arsenal Mathematicians, p. 134) that a maximal isolated subset of the n-cube I_n has 2^{n-1} vertices. The proof is given in Section 3. To describe the second problem, call a covering irredundant if it ceases to be a covering upon deletion of any of its cells (clearly every minimal covering is irredundant) and call a basic cell absolutely superfluous if it appears in no irredundant covering. Quine (Amer. Math. Monthly 66, pp. 758-9) suggested that every absolutely superfluous basic cell lay in the core. We have constructed a counterexample[3] based on a method which promises to be generally useful in constructing functions with prescribed properties. This work is described in Section 4.

Besides classifying structures of Boolean functions, one would like to know (especially for simplification-algorithm purposes) which properties are typical or frequent. This idea has led us to the identification of a number of nontrivial combinatorial-probability problems concerning properties of "random" Boolean functions; for example the "negligible core" conjecture of the next-to-last paragraph leads to the problem of finding the average core size of a random function. Formulations of some of these problems, together with some partial results, are given in the paper "Computer Simplification of Boolean Functions" mentioned above.

---

[3] Professor Quine informed us that he had received several other counterexamples.

3. ON A CONJECTURE OF CHITTENDEN

Let I_n be the set of vertices of the unit cube in n-dimensional space; that is, I_n is the set of all binary sequences of length n. A Boolean function of n variables is simply a subset of I_n.

Call two vertices of I_n neighbors if they differ in precisely one position. A Boolean function f is called connected if, given any two vertices X, X' in f, there exists a sequence

$$X = X_1, X_2, X_3, \ldots, X_{k-1}, X_k = X'$$

of members of f such that X_i and X_{i+1} are neighbors for $1 \leq i \leq k-1$. It is easy to show that if f is not connected, then it has a unique partition into maximal connected subsets called its components; if f is connected, the unique component consists of f itself.

The relevance of these concepts to the simplification of Boolean functions is as follows. By a cell of a Boolean function f we mean a face of I_n all of whose vertices lie in f. A covering of f is a collection C of cells such that each vertex of f lies in at least one cell from C. Each cell is assigned a cost in a manner which need not be described here, and the cost of a covering is defined to be the sum of the costs of its cells. The problem of finding a minimum-cost cover of f is then identical with that of finding a simplest "normal disjunctive form" for f. From the fact that any cell lies entirely in a single component of f, it follows that a covering of f is a disjoint union of coverings of the components of f; in particular a minimum-cost covering of f is a disjoint union of minimum-cost coverings of the components of f. That is, the Boolean simplification problem, for a function f which is not connected, can be split into

independent "smaller" subproblems by splitting f into its components. Such a decomposition is clearly likely to be helpful in attacking the simplification problem.

In this context it is natural to ask how many components a Boolean function of n variables can possible have. This question will be answered below. We shall also investigate the structure of the maximally disconnected functions of n variable, i.e., those with the greatest possible number of components.

A function will be called isolated if each of its components consists of a single vertex, i.e., if no two of its vertices are neighbors. Our results include a proof of the conjecture by E. W. Chittenden ("On the Minimal Representation of Boolean Functions", Transactions of the Tird Conference of Arsenal Mathematicians, U. S. Army Office of Ordnance Research Rept. No. 58-2, p. 134) that a maximal isolated function of n variables contains precisely half of the 2^n vertices of I_n.

LEMMA 1. *There is at least one maximally disconnected function of* n *variables which is isolated.*

To prove the lemma, observe first that because the number of Boolean functions of n variables is finite, a maximally disconnected function f must exist. Now form a function f' by deleting all but one vertex from every component of f. Clearly f' is isolated, and it must be maximally disconnected since it has the same number of components as does f.

LEMMA 2. *If* M_n *is the maximal number of components of a Boolean function of* n *variables, then* $M_n \leq 2^{n-1}$.

This is clear for n=1; assume it true for n=k and consider the case n = k+1. First let f be any function of k+1 variables; we define two functions f_0, f_1 of k variables called the projections of f. The function f_0 is obtained by dropping the final 0 from all members of f which end in 0, while f_1 is obtained by dropping the final 1 from all members of f which end in 1. If c, c_0, c_1 denote the respective cardinalities of f, f_0, f_1 then $c = c_0 + c_1$. Now specialize f to be a maximally disconnected function of k+1 variables which is isolated; the existence of such a function is guaranteed by Lemma 1. Then f_0 and f_1 are also isolated, so that $c_0 \leq M_k$ and $c_1 \leq M_k$, and thus $c \leq 2M_k$. Since $M_k \leq 2^{k-1}$ by inductive hypothesis, we have $M_{k+1} = c \leq 2^k$, and the inductive proof is complete.

For what follows it is convenient to define two special Boolean functions of n variables; f(n) consists of those members of I_n with an even number (possibly zero) of 1's among their entries, while g(n) consists of those members of I_n with an odd number of 1's as entries.

THEOREM 1. The functions f(n) and g(n) which form a partition of I_n, are the only maximal isolated Boolean functions of n variables. Furthermore, $M_n = 2^{n-1}$.

Part of the proof is trivial; it is clear that f(n) and g(n) are isolated and together form a partition of I_n. Since f(n) and g(n) each have 2^{n-1} members, and thus the same number of components, we must have $M_n > 2^{n-1}$, from which the equality follows by Lemma 2.

It only remains to show that if h is any maximal isolated function of n variables, then h is either f(n) or g(n). This is clearly true for n=1; suppose it true for n=k and consider the case n = k+1. The

projections h_0 and h_1 of $\underline{h}$ are isolated functions of $\underline{k}$ variables, so each has at most $M_k = 2^{k-1}$ vertices; since the number of vertices of $\underline{h}$ (namely $M_{k+1} = 2^k$) is the sum of the numbers in h_0 and h_1, each projection must have exactly M_k vertices and is thus a maximal isolated function of $\underline{k}$ variables. The induction hypothesis leads to four possibilities:

$$h_0 = f(k) \text{ and } h_1 = f(k),$$

$$h_0 = g(k) \text{ and } h_1 = g(k),$$

$$h_0 = f(k) \text{ and } h_1 = g(k)$$

$$h_0 = g(k) \text{ and } h_1 = f(k).$$

The first two possibilities are incompatible with the fact that $\underline{h}$ is isolated, the third yields $h = f(k+1)$, and the fourth yields $h = g(k+1)$. Thus the inductive proof is complete.

One might conjecture that a maximally disconnected set is necessarily isolated. A counter-example for n=1 is provided by I_1 itself. The next theorem shows that this is the only counter-example.

THEOREM 2. For $n > 1$, every maximally disconnected Boolean function of n variables is isolated.

To prove this, let f be a maximally disconnected function of $\underline{n}$ variables, where $n > 1$. As in the proof of Lemma 1, form f' by deleting all but one vertex from each component of f. Since f' is isolated and has the same number of components as f, it follows from Theorem 1 that f' is either f(n) or g(n). Without loss of generality assume f' = f(n). If $f \neq f'$, then f includes at least one vertex Y of g(n). One neighbor X_1 of Y is obtained by changing just the first entry of Y, while a second neighbor X_2 is obtained by changing just the second entry of Y. Since

X_1 and X_2 are in f(n) = f' and thus in f, and both are neighbors of the vertex Y of f, it follows that the vertices X_1 and X_2 of f' lie in the same component of f. This however contradicts the construction of f'; we conclude that f = f', so that f is isolated.

An even simpler proof of Lemma 2 and Theorem 1 can be given if we take as known the existence of a Gray code, i.e., an ordering $X_0, X_1, \ldots, X_m$ $(m=2^n-1)$ of the vertices of I_n such that for $1 \leq i \leq m-1$, X_{i-1} and X_{i+1} are neighbors of X_i. From this it is obvious that an isolated set cannot contain both X_i and X_{i+1} and therefore has at most 2^{n-1} vertices, and that if the sets

$$F(n) = \left\{ X_i \,\middle|\, i \text{ is even} \right\} \text{ and } G(n) = \left\{ X_i \,\middle|\, i \text{ is odd} \right\}$$

are isolated then they are the only maximal isolated sets. But the defining property of a Gray code makes it clear that the number of 1's in every member of F(n) has the same parity while the opposite parity holds for all members of G(n). Thus either F(n) = f(n) and G(n) = g(n), or F(n) = g(n) and G(n) = f(n); since f(n) and g(n) are easily seen to be isolated, the proof is complete.

(John Edmonds has extended Theorems 1 and 2 to a wider class of situations; his results appear as Section 5 in this report.)

A natural generalization of the notion of "isolated function" is that of "cellular function". A function will be called cellular if each of its components is a cell; different components may be cells of different dimensions. The functions f(n) and g(n) defined before Theorem 1 are examples of cellular functions.

THEOREM 3. Let F_d be a d-dimensional face of I_n, $0 \leq d < n$. There exists a Boolean function $f(n; F_d)$ of n variables with

F_d as a component, 2^{n-d-1} components, and every component a d-dimensional cell. The complement $g(n; F_d) = I_n - f(n; F_d)$ of $f(n; F_d)$ also has 2^{n-d-1} components, each a d-dimensional cell.

These functions $f(n; F_d)$, $g(n; F_d)$ can be regarded as the analogues, for cellular functions, of the functions f(n) and g(n) for isolated functions. For example our later results will imply as a very special case that these functions (for various F_d) give the only solutions to the problem of constructing a cellular function with the maximal number of components under the restriction that every component is a d-dimensional cell.

We will prove the theorem by explicitly constructing $f(n; F_d)$ and verifying its properties. For simplicity we assume F_d consists of all vertices $(x_1,\ldots,x_n)$ of I_n which satisfy

$$x_{d+1} = x_{d+2} = \ldots = x_n = 0;$$

any other d-dimensional face can be treated similarly. The cube I_n can be written as a Cartesian product $I_n = I_d \times I_{n-d}$, so that for example $F = I_d \times \{0_{n-d}\}$, where 0_{n-d} is a sequence of n-d 0's. We define

$$f(n; F_d) = \bigcup \{ I_d \times \{Z\} \mid Z \in f(n-d) \}.$$

That is, for each of the 2^{n-d-1} elements Z of f(n-d) (a function of n-d variables), we form the Cartesian product $I_d \times \{Z\}$, which is a subset of $I_d \times I_{n-d} = I_n$. The union of these subsets is taken as $f(n; F_d)$. F_d itself is obtained as such a subset by choosing $Z = 0_{n-d}$. Each subset, as the Cartesian product of a d-dimensional face by a

single point, is again a d-dimensional face. To see that these faces (which are each connected) are in fact the components of their union $f(n; F_d)$, it suffices to observe that since any two elements Z, Z' of the isolated set f(n-d) differ in at least two entries, no vertex of $I_d \times \{Z\}$ can be a neighbor of any vertex in $I_d \times \{Z'\}$.

Since the complement $I_{n-d} - f(n-d)$ of f(n-d) is the maximal isolated set g(n-d) which contains 2^{n-d-1} vertices, it follows that the complement of $f(n; F_d)$ is

$$g(n; F_d) = I_n - f(n; F_d) = \bigcup \{ I_d \times \{Z\} \mid Z \in g(n-d) \},$$

whose 2^{n-d-1} components are the individual d-dimensional faces $I_d \times \{Z\}$ for Z in g(n-d). Thus Theorem 3 is proved.

LEMMA 3. Let f_0 and f_1 be the projections of a cellular function f of k+1 variables. Then

(a) the components of f_0 are the projections h_0 (if non-empty) of the components h of f,

(b) the components of f_1 are the projections h_1 (if non-empty) of the components h of f,

(c) if h is a component of f, then either h_0 is empty or h_1 is empty or $h_0 = h_1$,

(d) f_0 and f_1, if non-empty, are cellular, and

(e) if a component p of f_0 meets a component q of f_1, then there is a component h of f such that $p = q = h_0 = h_1$.

To begin the proof, let h be any component of f. Since f is cellular, h is a cell of some dimension d. That is, there is a set S_h of k+1-d integers from among $\{1,\ldots,k,k+1\}$, and a binary sequence $\{c_i \mid i \in S_h\}$, such that h consists of all vertices $(x_1,\ldots,x_k,x_{k+1})$

of I_{k+1} which satisfy

$$x_i = c_i \text{ for all } i \in S_h.$$

If $k+1 \in S_h$ and $c_{k+1} = 0$, then h_1 is empty and h_0 consists of all vertices $(x_1,\ldots,x_k)$ of I_k which satisfy

$$x_i = c_i \text{ for all } i \in S_h - \{k+1\};$$

thus h_0 is a d-dimensional face. If $k+1 \in S_h$ and $c_{k+1} = 1$, then h_0 is empty and h_1 consists of all vertices of I_k which satisfy

$$x_i = c_i \text{ for all } i \in S_h - \{k+1\};$$

thus h_1 is a d-dimensional face. If k+1 is not an element of S_h, then h is the disjoint union of the set of vertices of I_{k+1} satisfying

$$x_{k+1} = 0,\ x_i = c_i \text{ for all } i \in S_h,$$

and the set of vertices of I_{k+1} satisfying

$$x_{k+1} = 1,\ x_i = c_i \text{ for all } i \in S_h;$$

thus $h_0 = h_1$ is the (d-1)-dimensional face of I_k consisting of all vertices $(x_1,\ldots,x_k)$ obeying $x_i = c_i$ for all $i \in S_h$. This proves (c) of the lemma, and shows that h_0 and h_1 are faces of I_k, and therefore cells of f_0 and f_1 respectively.

To prove (a) of the lemma, it now suffices to show that if h_0 and h_0^* are non-empty projections of distinct components h, h* of f, then no vertex $(x_1,\ldots,x_k)$ of h_0 is a neighbor of any vertex $(x_1^*,\ldots,x_k^*)$ of h_0^*. If such a "neighboring" occurred, then the vertex $(x_1,\ldots,x_k,0)$ of h would be a neighbor of the vertex $(x_1^*,\ldots,x_k^*,0)$ of h*, which is impossible since h and h* are distinct components. Thus (a) of the lemma holds; the proof of (b) is analogous, and (d) follows directly from (a), (b), and the fact (see the last paragraph) that the projections of faces are again faces.

Now suppose a component p of f_0 and a component q of f_1 have a vertex $(x_1,\ldots,x_k)$ in common. By what has already been proved, there exist components h_p and h_q of f such that $(h_p)_0 = p$ and $(h_q)_1 = q$. Thus $(x_1,\ldots,x_k,0)$ is a vertex of h_p and $(x_1,\ldots,x_k,1)$ is a vertex of h_q. Since these two vertices are neighbors, we must have $h_p = h_q = h$, say. Since neither $h_0 = p$ nor $h_1 = q$ is empty, it follows by (c) that $p = q$. This completes the proof of the lemma.

We can now generalize Theorems 1 and 2:

THEOREM 4. The maximum number of vertices in a cellular Boolean function of n variables (except for I_n itself) is 2^{n-1}. For $0 \leq d < n$ and each d-dimensional face F_d of I_n, $f(n; F_d)$ is the only cellular function with F_d as a component which contains 2^{n-1} vertices.

The statement is clearly true for n = 1. We assume it true for n = k, and complete the inductive proof by showing that it holds for n = k+1. Since a function of the form $f(k+1; F_d)$ has $2^{k+1-d-1}$ d-dimensional cells (each with 2^d vertices) as components, we see that the maximum number of vertices in a cellular Boolean function of k+1 variables is at least 2^k. To prove the opposite inequality, let f be any cellular function of k+1 variables (other than I_{k+1}). If either one of f_0 or f_1 is I_k, then by (e) of Lemma 3 the other must be empty, since otherwise we would have $f_0 = f_1 = I_k$ and thus $f = I_{k+1}$. Thus if $f_0 = I_k$ then f consists of all vertices of I_{k+1} obeying $x_{k+1} = 0$, while if $f_1 = I_k$ then f consists of all vertices obeying $x_{k+1} = 1$; in either case f contains 2^k vertices. If both f_0 and f_1 are proper subsets of I_k, then by the inductive hypothesis and (d)

of Lemma 3 we can conclude that f_0 and f_1 each contain at most 2^{k-1}; the number of vertices in f is the sum of the number in f_0 and the number in f_1, so that f contains <u>at most</u> 2^k vertices.

Now let F_d be a d-dimensional face of I_{k+1}, where $0 \leq d < k+1$, and let f be a cellular Boolean function of k+1 variables with 2^k vertices and with F_d as a component. For simplicity we assume F_d is the set of vertices $(x_1,\ldots,x_k,x_{k+1})$ of I_{k+1} which satisfy

$$x_{d+1} = x_{d+2} = \ldots = x_k = x_{k+1} = 0.$$

If $f_0 = I_k$ then (see the previous paragraph) f would consist of all vertices of I_{k+1} obeying $x_{k+1} = 0$, and so $f = f(k+1; F_k) = f(k+1; F_d)$. If $f_1 = I_k$, then (see the last paragraph) f_0 would be empty, which is impossible since f_0 contains the non-empty set $(F_d)_0$. Thus we can assume that f_0 and f_1 are <u>proper</u> cellular functions of <u>k</u> variables; thus by inductive hypothesis each contains at most 2^{k-1} vertices, but since f contains 2^k vertices each of f_0, f_1 must contain exactly 2^{k-1}.

By Lemma 3, f_0 is a cellular function with $(F_d)_0$ as a component. It follows from the inductive hypothesis that $f_0 = f(k; (F_d)_0)$. We will show in the next paragraph that f_1 is the complement (in I_k) of f_0. Note that f is the union of (i) the set obtained by adjoining a final <u>0</u> to each element of $f_0 = f(k; (F_d)_0)$ and (ii) the set obtained by adjoining a final <u>1</u> to each element of $f_1 = g(k; (F_d)_0)$. Since $(F_d)_0$ is d-dimensional (this follows from the fact that $(F_d)_1$ is empty and from the proof of Lemma 3), we see that the first set is obtained by adjoining a final <u>0</u> to each element of

$$\bigcup \{ I_d \times \{Z\} \mid Z \in f(k-d) \} ,$$

while the second is obtained by adjoining a final 1 to each element of

$$\bigcup \{ I_d \times \{Z\} \mid Z \in g(k-d) \} .$$

But adjoining a final 0 to each element Z of f(k-d) and a final 1 to each element Z of g(k-d) simply yields f(k+1-d), and so we have

$$f = \{ I_d \times \{ Z^* \} \mid Z^* \in f(k+1-d) \} = f(k+1; F_d),$$

as was to be proved.

It remains to show that f_1 is the complement of f_0. Since f_0 and f_1 each contain 2^{k-1} vertices, it suffices to prove f_1 a subset of the complement of f_0. Thus we must show that no component p of f_0 can meet a component q of f_1. If such a meeting occurred, then by (e) of Lemma 3 there would be a component h of f such that $p = q = h_0 = h_1$, and by (b) and (c) of Lemma 3 it follows that p is a component of f_1 as well as f_0. But by inductive hypothesis f_0 is the only cellular function of k variables with 2^{k-1} vertices having p as a component, and so we must have $f_1 = f_0$. Thus $(F_d)_0$ is a component of f_1 as well as f_0. Thus the face of I_{k+1} consisting of all vertices $(x_1,\ldots,x_k,x_{k+1})$ obeying

$$x_{d+1} = x_{d+2} = \ldots = x_k = 0,$$

which properly contains F_d, would lie in f, contradicting the hypothesis that F_d is a component of f. This completes the proof of Theorem 4.

COROLLARY. A cellular Boolean function with the maximum number of vertices has all its component cells of the same dimension d, and its complement is again such a function

(with the same "d"). There are $2\binom{n}{d}$ such functions for each $d < n$, and thus in all there are $2(2^n-1)$ cellular functions of n variables with the maximum number of vertices.

4. ON A CONJECTURE OF QUINE

The notation and terminology of the preceding section are also to apply here. A cell of the Boolean function f is called basic ("maximal" would be a better term) if it is not properly contained in any other cell of f. In the "Boolean simplification" situations described earlier in this report, "costs" are assigned to cells in such a way that a minimum-cost covering necessarily consists only of basic cells.

A basic cell of f is called essential if some vertex of this cell lies in no other basic cell of f. Clearly every covering of f by basic cells (in particular, every minimum-cost covering) must include all the essential cells. The union of all the essential cells of f will here be called the core of f; some authors use this term, instead, to refer to the collection of all essential cells of f.

A covering C of f is called redundant if some cell in C is contained in the union of the other cells of C. If C is not redundant, it is called irredundant. Clearly every minimum-cost covering is irredundant (assuming the "cost" of every cell is positive). This motivates our calling a basic cell absolutely superfluous if it appears in no irredundant covering of f by basic cells.

A category of cells whose "absolute superfluity" is especially clearly seen is provided by those basic cells which are not themselves essential but are contained in the core of f. Quine[4] asked whether all absolutely

---

[4] W. V. Quine, "On Cores and Prime Implicants of Truth Functions," Amer. Math. Monthly 66 (1959). See pp. 758-759.

superfluous cells were of this type. We shall give a counter-example for this conjecture.

A weaker version of the conjecture would be the assertion that every absolutely superfluous basic cell must intersect the core. A still weaker version would be the claim that if f has vacuous core (i.e., has no essential cells) then it has no absolutely superfluous cells. We shall also give counter-examples for these two conjectures. Of course a counter-example for the weakest version serves to disprove all three, but it is interesting to see how the complexity of counter-examples must increase as the conjecture is progressively weakened. To justify this phrase "must increase," we should prove that the number of variables in each of our counter-examples is as small as possible. For the present we leave this as a conjecture.

Each of the three counter-examples will be discussed in the same way. The Boolean function will at first be described "abstractly", i.e. as a collection of basic cells without regard to a specific "embedding" in a discrete unit cube. The basic cells are assumed to have no intersections other than those explicitly listed. Also, the function will be shown in fact to provide a counter-example for the conjecture under discussion. Finally, an embedding of the function in a discrete unit cube will be given; we must leave to the reader the verification that the Boolean function thus specified is really the same one previously discussed abstractly.

The three Boolean functions (one for each counter-example) have the following structure in common. Each has a basic 1-cell A, whose vertices

will be denoted X and Y, and two basic 2-cells B and C, which intersect in a 1-face having X as one vertex. The vertex other than X of this 1-face will be denoted Z. For each function to be constructed, Z will lie in no basic cells other than B and C, so that for every covering, the vertex X of A will lie in some cell of the covering (viz., either B or C) other than A. The functions will also be constructed so that for any covering by basic cells the other vertex, Y, of A lies in some cell of the covering other than A. Thus A will be an absolutely superfluous basic cell; it turns out to be the only such cell for the examples to be described below.

For the initial example, the "common structure" is supplemented in two ways. First, a basic 1-cell A' with Y as one vertex is adjoined. The vertex of A' other than Y will lie in no basic cell except A'; therefore A' will be an essential cell and for any covering by basic cells Y will be in a cell of the covering (viz., A') other than A. As noted above, this will ensure that A is absolutely superfluous. Second, B and C will be prevented from being essential cells in such a way that the vertex X lies in no basic cells other than A, B, and C. Since the vertex X of the absolutely superfluous cell A lies in no essential cell, we will indeed have an example of an absolutely superfluous basic cell not contained in the core. B is prevented from being essential by adjoining a basic 2-cell B' which intersects B in the 1-face opposite (X, Z), while C is prevented from being essential by adjoining a basic 2-cell C' which intersects C in the 1-face opposite (X, Z).

A realization of this example as a Boolean function of 5 variables is given by the following list, in which for instance (01xx0) denotes the set of vertices $(x_1, x_2, x_3, x_4, x_5)$ of I_5 such that

$x_1 = x_5 = 0$; $x_2 = 1$; x_3 and x_4 unspecified:

A': (000x1)
A : (0000x)
B': (01xx0)
B : (0xx00)
C : (x0x00)
C': (10xx0)

The second example retains the basic cells A, B, B', C, C' from the first one. In addition, there are four new basic 2-cells D, D', E, E' whose arrangement with respect to the vertex Y of A is just like the arrangement of B, B', C, C' with respect to the vertex X of A. Explicitly, D and E intersect in a 1-face which has Y as one vertex (call its other vertex W), D is prevented from being essential because it is intersected by D' in the 1-face opposite (Y, W), and E is prevented from being essential because it is intersected by E' in the 1-face opposite (Y, W). By the same arguments used above for X (i.e., "by symmetry"), Y (like X) lies in no basic cell, and lies in some basic cell other than A of any covering by basic cells. Since X and Y both have these properties, A is an example of an absolutely superfluous basic cell which is disjoint from the core.

A realization of this example as a Boolean function of 7 variables is given by the following list:

A : (0000x00)
B': (01xx000)
B : (0xx0000)
C : (x0x0000)
C': (10xx000)
D': (00xx110)
D : (000x1x0)
E : (000x10x)
E': (00xx101)

The third example retains the cells A, B, C, D, E of the second one. As before, Z lies in no basic cells other than B and C (so that every covering by basic cells includes either B or C), and W lies in no basic cells other than D and E (so that every covering by basic cells includes either D or E); also as before, this implies that A is absolutely superfluous. Let X' and Z' be the vertices of the 1-face of B opposite to (X, Z), with X' opposite X and Z' opposite Z, and let X" and Z" be the vertices of the 1-face of C opposite to (X, Z), with X" opposite X and Z" opposite Z. B and C are prevented from being essential by adjoining a chain F_1, F_2, F_3, F_4, F_5, of basic 1-cells from X' to Z', and a chain G_1, G_2, G_3, G_4, G_5 of basic 1-cells from X" to Z".

Similarly, let Y' and W' be the vertices of the 1-face of D opposite to (Y, W), with Y' opposite Y and W' opposite W, and let Y" and W" be the vertices of the 1-face of E opposite to (Y, W), with Y" opposite Y and W" opposite W. D and E are prevented from being essential by adjoining a chain H_1, H_2, H_3, H_4, H_5 of basic 1-cells from Y' to Y", and a chain J_1, J_2, J_3, J_4, J_5 of basic 1-cells from W' to W". The resulting function

has no essential cells, and so A is an example of an absolutely superfluous basic cell in a Boolean function with vacuous core.

A realization of this example as a Boolean function of 8 variables is given by the following list:

A : (00000x00)
B : (0xx00000)
F_1 : (001x0000)
F_2 : (0011x000)
F_3 : (0x111000)
F_4 : (011x1000)
F_5 : (0110x000)
C : (xx000000)
G_1 : (100x0000)
G_2 : (1001x000)
G_3 : (1x011000)
G_4 : (110x1000)
G_5 : (1100x000)
D : (000x01x0)
H_1 : (00x00110)
H_2 : (0010x110)
H_3 : (001x1110)
H_4 : (00x11110)
H_5 : (0001x110)
E : (000x010x)
J_1 : (00x00101)
J_2 : (0010x101)
J_3 : (001x1101)
J_4 : (00x11101)
J_5 : (0001x101).

5. MAXIMAL ISOLATED SETS IN BIPARTITE GRAPHS[5]

In this section, some of the results of Section 3 ("On a conjecture of Chittenden") will be examined in a more general setting.

A (finite, unoriented) graph is a finite nonempty set of elements called vertices, together with a collection of unordered pairs of distinct vertices. Each pair in the collection is called an edge, and is said to join the two vertices composing it; each of the two vertices forming an edge is said to meet the edge. A subgraph of a graph consists of a subset of its vertices and a subcollection of its edges, each in the subcollection joining two vertices of the subset.

A connected graph is one such that for every proper subset of the vertices, there is at least one edge joining a vertex of the subset and a vertex of its complement. Any graph can be uniquely partitioned into connected subgraphs, the components of the graph.

A path in a graph is an alternating sequence of vertices and edges, beginning with a vertex and ending with a vertex, such that each edge in the sequence joins the vertices immediately preceding and following it in the sequence. The relation of being "joined by a path", i.e. of being contained in a common path, is clearly an equivalence relation on the set of vertices of the graph, and the resulting equivalence classes can easily be shown to be the components of the graph.

A bipartition of a graph is a complementary pair of subsets S, S' of the vertices, such that every edge joins a vertex of S to a vertex of S'. A bipartite graph is a graph which admits a bipartition.

---

[5] The material of this section is due to John Edmonds.

Since a bipartition of a graph "induces" a bipartition of each subgraph, it follows that every subgraph of a bipartite graph is also bipartite.

A connected bipartite graph has a unique bipartition.[6] To see this, suppose (S,S') and (T,T') are bipartitions such that S and T have a common vertex v_o. (This assumption involves no loss of generality.) Each vertex v is the last vertex of a path whose first vertex is v_o; the vertices of this path are on the one hand alternately in S and in S', and on the other hand alternately in T and in T', so that in particular v is either in both S and T or in both S' and T'. Thus S = T and S' = T', as was to be proved. The same idea can in principle be used to test a given connected graph for bipartiteness; one places a fixed vertex v_o in S, and joins v_o to each other vertex v by a path whose vertices are alternately placed in S and in S'. The graph is bipartite if and only if the classification of each vertex v (as being in S or in S') is independent of the path from v_o to v employed in obtaining the classification. It readily follows that a graph is bipartite if and only if every circuit (a path beginning and ending with the same vertex) has an even number of edges.

A subset of the vertices of a graph is called isolated if no two vertices of the subset are joined by an edge. For example, the sets S, S' of a bipartition are each isolated. Berge[7] has given an

---

[6] A bipartite graph with k components thus has 2^k bipartitions, formed in an obvious way from the unique bipartitions of the components.

[7] C. Berge, "Two Theorems in Graph Theory", Proc. Nat. Acad. Sci. U.S.A. 43 (1957), pp. 842-844.

algorithm for finding a maximal isolated subset (i.e., one with a maximum number of vertices) of any graph. A vertex-cover of a graph is a set of vertices such that every edge meets at least one vertex of the set; the complement of an isolated set is a vertex-cover and vice versa, so that the complement of a maximal isolated set is a minimal vertex-cover and vice versa. Therefore Berge's algorithm can be used to find a minimal vertex-cover; Berge in fact formulates his problem in this way.

The cubic graph C_n has as vertices the 2^n vertices of the unit cube in n-dimensional space, and as edges the neighboring pairs of these vertices. That is, a vertex of C_n is a sequence of n numbers, each either 0 or 1, and two vertices form an edge if and only if they differ in precisely one position. C_n is a connected bipartite graph, whose bipartition consists of the set S of vertices with an even number of zeros and the set S' of vertices with an odd number of vertices.

Theorem 1 of Section 3 asserts that the only maximal isolated sets of C_n are the sets S, S' of its bipartition. This suggests investigating what other connected bipartite graphs have the same property. An obvious necessary condition is that S and S' have the same number of vertices. That this condition is not sufficient can be seen by taking $S = \{a,b\}$ and $S' = \{a',b'\}$, with (a,a'), (b,b') and (a,b') as edges; here S and S' are maximal isolated sets, but so is $\{a',b\}$. For an example in which S and S' are not maximal isolated sets, take $S = \{a,b,c\}$ and $S' = \{a',b',c'\}$, with edges (a,a'), (a,b'), (a,c'), (b,c') and (c,c'); here $\{a',b',b,c\}$ is a maximal isolated set. It will be shown, however, that the class of connected bipartite graphs with the property in question (viz., that the bipartition

sets S, S' are the only maximal isolated sets) contains many more than the cubic graphs.

Since an equivalent property is that S and S' are the only minimal vertex-covers, it is of interest to give an alternate characterization of the cardinality of a minimal point-cover of a bipartite graph. For this purpose, number the vertices of S and those of S'; then form a matrix whose entry in the i-th row and j-th column is one of the i-th vertex of S is joined to the j-th vertex of S', and is zero otherwise. A theorem of Konig[8] asserts that for a matrix of zeros and ones, the minimum number of lines[9] that contain all the ones is equal to the maximum number of ones with no two on the same line. A set of lines containing all the ones corresponds to a vertex-cover, while a set of ones with no two on the same line corresponds to an isolated edge-set (i.e., a set of edges no two of which are met by a common vertex). Thus the cardinality of a minimal vertex-cover of a bipartite graph is equal to that of a maximal isolated edge-set.

THEOREM 1. Suppose each set of a bipartition (S,S') of a graph consists of m vertices. S and S' are maximal isolated sets if and only if there is a one-to-one correspondence between them such that each pair of corresponding vertices is an edge.[10]

---

[8] D. Konig, Theorie der endlichen und unendlichen graphen, Chelsea, New York (1950).

[9] A line of a matrix is either a row or a column.

[10] This condition asserts that the rows of the matrix to which Konig's theorem was applied can be permuted to as to make the main diagonal consist entirely of ones.

Proof. S and S' are maximal isolated sets if and only if they are minimal vertex-covers, which by Konig's theorem is true if and only if there exists a maximal isolated edge-set $E = \{e_1, \ldots, e_m\}$ with m members. If such a set E exists, then for $i = 1, 2, \ldots, m$ set $e_i = (a_i, a_i')$ with a_i in S and a_i' in S'; because E is an isolated edge-set, it follows that $a_i \neq a_j$ and $a_i' \neq a_j'$ if $i \neq j$, so that the required one-to-one correspondence is obtained. Conversely, if such a correspondence exists then the vertices of S and S' can be numbered

$$S = \{a_1, \ldots, a_m\}, \quad S' = \{a_1', \ldots, a_m'\}$$

in such a way that $e_i = (a_i, a_i')$ is an edge for $i = 1, 2, \ldots, m$. The edges $\{e_1, \ldots, e_m\}$ form an isolated edge-set E which need only be proved maximal to complete the demonstration. If E were not maximal, there would exist an isolated edge-set $E^* = \{e_1^*, \ldots, e_{m+1}^*\}$ with $m + 1$ members. For $i = 1, 2, \ldots, m$, let $e_i^* = (b_i, b_i')$ with b_i in S and b_i' in S'; then $b_i \neq b_j$ and $b_i' \neq b_j'$ for $i \neq j$, so that

$$S = \{b_1, \ldots, b_m\}, \quad S' = \{b_1', \ldots, b_m'\}.$$

Any other edge $e_{m+1}^* = (b_r, b_s')$ would be met by b_r (which also meets e_r^*) and by b_s' (which also meets e_s^*), contradicting the assumption that E^* is an isolated edge-set. This completes the proof of the theorem.

A theoretical solution to the problem can now be given. By the preceding theorem one need only consider connected graphs with a bipartition whose sets admit numberings

$$(*) \qquad S = \{a_1, \ldots, a_m\}, \quad S' = \{a_1', \ldots, a_m'\}$$

such that (a_i, a_i') is an edge for $i = 1, 2, \ldots, m$. If there exists a maximal isolated set T other than S and S', then by the theorem T contains exactly m vertices, and a suitable simultaneous permutation of S

and S' will make

$$T = \{a_1, \ldots, a_k, a'_{k+1}, \ldots, a'_m\}$$

for some integer k with $1 \leq k \leq m - 1$. That is, if the matrix M to which Konig's theorem was applied earlier is formed using the numberings (*), then there would be an m x m permutation matrix P such that PMP' has a k x (m - k) block of zeros in the upper right-hand corner. Therefore the graph has the property in question if and only if no such permutation exists. For large m it would be impracticable actually to test each of the m! permutations, so that unless a more efficient algorithm can be found the solution can only be regarded as a "theoretical" one.

A more concrete result will now be given. A regular graph of degree d is a graph each of whose vertices meets precisely d edges. For example, the cubic graph C_n is a regular graph of degree n, so that the following theorem has Theorem 1 of Section 4 as an immediate corollary.

THEOREM 2. For a connected bipartite regular graph, the sets S and S' of the unique bipartition are the only maximal isolated sets.

Proof. Let S and S' contain m and m' vertices respectively. Since the graph is regular of some degree d, the number of edges is on the one hand dm (because each vertex of S meets d edges) and on the other hand dm'; hence m = m'. Using arbitrary numberings for the vertices of S and S', form the matrix M to which Konig's theorem was applied earlier. M contains precisely dm ones, with precisely d ones on each line; thus any set of lines which cover all the ones must contain at least m lines, i.e. every vertex-cover contains at least m vertices. Therefore S and S' are minimal vertex-covers, and hence maximal isolated sets. If some other maximal isolated set existed, then as noted earlier S and S' could be

renumbered so that the matrix M assumed the partitioned form

$$\begin{bmatrix} A & O \\ B & C \end{bmatrix}$$

where "O" denotes a k x (m - k) matrix of zeros. Since each column of M contains exactly $\underline{d}$ ones, the same holds for C, but each row of C can contain at most $\underline{d}$ ones since each row of M contains exactly $\underline{d}$ ones. Therefore each row of C contains precisely $\underline{d}$ ones, and so B = O, an (m - k) x k matrix of zeros. This, however, would imply that the first $\underline{k}$ vertices of S and the first $\underline{k}$ vertices of S' are joined by edges only to each other, and that the same holds for the last m - k vertices of S and the last m - k vertices of S'. Since this contradicts the hypothesis that the graph is connected, the theorem is proved.

Theorem 2 does not describe all graphs with the desired property (as will be clear from the following corollary), but it does give many more than the cubic graphs, since whenever $1 < d \leq m$ there exists a connected bipartite graph with 2m vertices which is regular of degree $\underline{d}$. Such a graph can be constructed by taking $S = \{a_1, \ldots, a_m\}$, $S' = \{a'_1, \ldots, a'_m\}$, and for i = 1, 2, ..., m making (a_i, a'_j) an edge for j = i, i + 1, ..., i + d - 1 (where j-values greater than $\underline{m}$ are read modulo $\underline{m}$).

COROLLARY. <u>For a connected bipartite graph which has a connected regular subgraph containing all its vertices, the sets</u> S <u>and</u> S' <u>of the unique bipartition are the only maximal isolated subsets</u>.

<u>Proof</u>. The unique bipartition of the original graph must be identical with that of the subgraph, and any isolated subset of the vertices of the original graph is also isolated when considered as a

subset of the vertices of the subgraph. Thus the result follows from Theorem 2.

It would be interesting to know whether the corollary describes all graphs with the desired property.

6. ONE-DIMENSIONAL CATALOGING OF INCOMING MATERIAL[11]

Suppose we have a one-dimensional method for classifying incoming material. By this we mean something like the Dewey Decimal System; each document is assigned a single number in some way, and a set of documents is "organized" by rearranging it so that the corresponding numbers are in increasing order. In what follows, we shall forget about the documents themselves and the problem of deciding what numbers to assign to them. Instead, we consider an idealized situation in which the only problem is that of arranging an incoming stream of numbers in ascending order.

It is assumed for simplicity that the probability of equality of any two of the numbers is negligible. If there are n numbers in the stream, then since only the order of their magnitudes (and not the magnitudes themselves) are relevant in arranging them, the numbers can be replaced by 1, 2, ..., n in some order. We suppose that the stream of numbers is random in the sense that all of the $n!$ possible orderings of 1, 2, ..., n are equally likely.

Assume the stream of n numbers is being fed into a computer with M fast-access memory cells (where $n \gg M$) and an unlimited slow-access auxiliary memory. Clearly we wish to utilize the limited fast-access memory optimally before restoring the data in the slow-access auxiliary memory. One suggestion is to order the first M numbers monotonically in the fast memory, put them in the auxiliary memory and keep a record of the largest number found (call it N). Then bring in M more numbers,

---

[11] The material of Sections 6 and 7 are due to Dr. Karl Goldberg.

order those which are greater than N, add them to the previously stored chain,[12] and start a new chain with the numbers less than N. Of course these chains must still be intertwined later, but at least some progress has been made toward arranging the numbers in increasing order. One measure of this progress is given by the lengths of the chains, or more precisely by their expected lengths, as compared with the expected lengths in the original stream (i.e., the case M = 1, in which no use is made of the fast memory). It is these standards of comparison, the expected chain-lengths for M = 1, which are evaluated in the next section. The results of this evaluation are summarized below.

If $E_r(n)$ denotes the expected length of the r-th chain, then the results can be summarized as follows:

$$E_r(n) = \sum_{k=r}^{n} \frac{1}{k!} \sum_{j=0}^{r} (-1)^j \binom{k+1}{j} (r-j)^k \qquad (1 \leq r \leq n,\ n \geq 1)$$

$$E_r = \lim_{n\to\infty} E_r(n) = r \sum_{k=0}^{r} \frac{(-1)^k}{k!} (r-k)^{k-1} e^{r-k} \qquad (r \geq 1)$$

$$E_r - 2 = O(|\rho+1|^{-r}) \qquad (r \to \infty)$$

where ρ and $\bar{\rho}$ are the non-zero roots of $e^z = 1 + z$ closest to the origin: $1 + \rho = 8.07556\ldots$. The rapidity of the convergence $E_r \to 2$ is illustrated by

---

(12) By a chain we mean an increasing sequence of numbers.

$$E_1 = e-1 = 1.71828\ 18285$$

$$E_2 = e^2 - 2e = 1.95249\ 24429$$

$$E_3 = e^3 - 3e^2 + \frac{3}{2}e = 1.99579\ 13691$$

$$E_4 = e^4 - 4e^3 + 4e^2 - \frac{2}{3}e = 2.00003\ 88505$$

$$E_5 = e^5 - 5e^4 + \frac{15}{2}e^2 + \frac{5}{24}e = 2.00005\ 75752$$

As an example of how $E_r(n)$ is calculated from its definition, take $n = 3$ and let L_1, L_2, L_3 denote the lengths of the first, second and third chains respectively:

Arrangement	L_1	L_2	L_3
1, 2, 3	3		
1, 3, 2	2	1	
2, 1, 3	1	2	
2, 1, 3	2	1	
3, 1, 2	1	2	
3, 2, 1	1	1	1
	10	7	1

Since each arrangement has probability 1/6:

$$E_1(3) = 10/6,\ E_2(3) = 7/6,\ E_3(3) = 1/6.$$

7. PROOFS OF THE RESULTS IN SECTION 6.

In this section we shall derive the results stated in the previous section. We shall use a general approach both to emphasize the reasoning behind the formulas, and because this generality is interesting in itself. Finally, because this is not a mathematical note, we shall skip many details of mechanical manipulation which the reader may want to fill in for himself.

Let $x_1, x_2, \ldots, x_n$ be a sequence of distinct real numbers. If

$$x_1 < x_2 < \ldots < x_{a_1} \quad \text{but} \quad x_{a_1} > x_{a_1+1} \quad \text{or} \quad a_1 = n$$

then we say that the sequence *begins with a chain of length* a_1. If the remaining chain $x_{a_1+1}, x_{a_1+2}, \ldots, x_n$ begins with a chain of length a_2, and so on, we say that the sequence has a chain structure $(a_1, a_2, \ldots, a_k)$. Note that a chain structure is an ordered partition of the number of elements in the sequence: $a_1 + a_2 + \ldots + a_k = n$ with each $a_i \geq 1$.

By letting 1 correspond to the smallest number in the sequence, 2 correspond to the next smallest, and so on, the original sequence can be represented as a permutation of the integers 1, 2, ..., n; every rearrangement of the original sequence corresponds to exactly one permutation. We formulate our problem exactly by asking: *what is the expected length of the r-th chain in a permutation of degree n?* We denote this number by $E_r(n)$. We shall show that $\lim_{n\to\infty} E_r(n)$ exists for each r and denote this limit by E_r. It can be interpreted as *the expected length of the r-th chain in an infinite sequence of random numbers*. We shall show that $\lim_{r\to\infty} E_r = 2$, and this is the

expected length of any chain in such a sequence.

Let $c_n(a_1, a_2, \ldots, a_k)$ denote the number of permutations of degree n with chain structure $(a_1, a_2, \ldots, a_k)$. For example, $c_n(n) = c_n(1,1,\ldots,1)=1$. We proceed to find a recursion for these numbers.

Choose any a_1 of the integers 1, 2, ..., n and arrange them in monotonically increasing order. This can be done in $\binom{n}{a_1}$ different ways. Follow this sequence with any sequence of the remaining $n - a_1$ integers which has a chain structure $(a_2, \ldots, a_k)$. If the first element of this latter sequence is less than the last element of the sequence with a_1 members, then the resulting combined sequence will have a chain structure $(a_1, a_2, \ldots, a_k)$. Otherwise the chain structure will be $(a_1+a_2, a_3, \ldots, a_k)$. Since any permutation of degree n with either of the latter two chain structures can be obtained in this manner, we have the recursion

$$(1)\quad \binom{n}{a_1} c_{n-a_1}(a_2, \ldots, a_k) = c_n(a_1, a_2, \ldots, a_k) + c_n(a_1+a_2, a_3, \ldots, a_k),$$

valid for $k \geq 2$, all n, and all a_i.

Now let $p(a_1, a_2, \ldots, a_k)$ denote the probability that a permutation of degree $a_1 + a_2 + \ldots + a_k$ has a chain structure $(a_1, a_2, \ldots, a_k)$. For example, $p(n) = 1/n!$ In general:

$$p(a_1, a_2, \ldots, a_k) = c_n(a_1, a_2, \ldots, a_k)/n! \qquad n = a_1 + a_2 + \ldots + a_k$$

We can rewrite (1) in terms of these probabilities as

$$(2)\quad p(a_1)p(a_2, \ldots, a_k) = p(a_1, a_2, \ldots, a_k) + p(a_1+ a_2, a_3 \ldots, a_k)$$

We can also express $E_r(n)$ in these terms:

$$(3)\qquad E_r(n) = \sum_{k=r}^{n} \sum_{\sum a_i = n} a_r p(a_1, a_2, \ldots, a_k)$$

where the symbols below the second summation sign indicate that the sum is to be taken over all positive $a_1, a_2, \ldots, a_k$ such that $a_1 + a_2 + \ldots + a_k = n$

Instead of directly computing $E_r(n)$ in this case we shall divert our attention to a somewhat more general discussion in which, however, the various arguments are exactly the same as those we would need in the specific case.

We assume that for every vector $(a_1, a_2, \ldots, a_k)$ with positive integral entries we are given a complex number[13] $p(a_1, a_2, \ldots, a_k)$, and that the set of such numbers satisfies (2). We know $p(n)$, for every n, and we wish to compute $E_r(n)$ as defined by (3).

Using (2), it is easy to show by induction on k, that

$$(4)\quad p(a_1, a_2, \ldots, a_k) = \sum_{j=1}^{k} (-1)^{k-j} \sum_{\sum k_i = k} p(a_1 + a_2 + \ldots + a_{k_1}) p(a_{k_1+1} + \ldots + a_{k_1+k_2}) \cdots \cdots p(a_{k_1+k_2} + \ldots + k_{j-1} + 1 + \ldots + a_k)$$

From this it is clear that

$$(5)\quad p(a_1, a_2, \ldots, a_k) = P(a_k, \ldots, a_2, a_1)$$

Equation (2) and induction on r also yields

$$(6)\quad p(a_1, \ldots a_r) p(a_{r+1}, \ldots, a_k) = p(a_1, \ldots, a_k) + p(a_1, \ldots, a_{r-1}, a_r + a_{r+1}, a_{r+2}, \ldots, a_k)$$

which holds for $k \geq 2$, for $r = 1, 2, \ldots, k - 1$, and for all $a_1, \ldots, a_k$.

---

[13] Up to a point, to be reached later, the domain of the mapping p could be any commutative ring, but such generality is not of interest here.

We shall use (6) to find a formula for $E_r(n)$. In doing so we will use the following symbols:

$$S_{k,r}^{(j)}(n) = \sum_{\Sigma a_i = n} \binom{a_i}{j} p(a_1, \ldots, a_k)$$

$$S_{k,r}(n) = \sum_{\Sigma a_i = n} a_r^r p(a_1, \ldots, a_k) = S_{k,r}^{(1)}(n)$$

$$S_k(n) = \sum_{\Sigma a_i = n} p(a_1, \ldots, a_k) = S_{k,r}^{(0)}(n)$$

$$S(n) = \sum_{k=1}^{n} \sum_{a_i = n} p(a_1, \ldots, a_k) = \sum_{k=1}^{n} S_k(n)$$

as well as

$$E_r^{(j)}(n) = \sum_{k=r}^{n} S_{k,r}^{(j)}(n).$$

For example,

$$E_r(n) = \sum_{k=r}^{n} S_{k,r}(n) = E_r^{(1)}(n)$$

If we multiply on both sides of equation (6) by $\binom{a_r}{j}$ and sum over all $a_1, \ldots, a_k$ whose sum is n, we get

$$(7) \qquad \sum_{t=r}^{n-k+r} S_{r,r}^{(j)}(t) S_{k-r}(n-t) = S_{k,r}^{(j)}(n) + S_{k-1,r}^{(j+1)}(n) - \delta_{0j} S_{k-1}(n), \quad n \geq k > r \geq 1$$

If we sum this equation over values with a fixed k + j and alternating signs we get

$$(8) \qquad S_{k,r}^{(j)}(n) = \sum_{i=0}^{k-r} (-1)^{k-r-i} \sum_{t=r}^{n-i} S_{r,r}^{(j+k-r-i)}(t) S_i(n-t) + \delta_{0j} S_{k-i}(n), \quad n \geq k \geq r \geq 1$$

where we have defined $S_0(n) = \delta_{0n}$.

Because $S_{1,1}^{(j)}(n) = \binom{n}{j} p(n)$, when we set r = 1 in (8) we get

$$(9) \qquad S_{k,l}^{(j)} = \sum_{i=0}^{k-l} (-1)^{k-l-i} \sum_{t=1}^{n-i} \binom{t}{j+k-i-1} p(t) S_i(n-t) + \delta_{0j} S_{k-l}(n)$$

For $j = 0$ this becomes

$$(10) \qquad S_k(n) = \sum_{i=0}^{k-l} (-1)^{k-l-i} \sum_{t=0}^{n-i} \binom{t}{k-l-i} p(t) S_i(n-t)$$

where we have defined $p(0) = 1$.

Let

$$P(x) = \sum_{n=0}^{\infty} p(n) x^n$$

Then (10) is equivalent to

$$(11) \qquad \sum_{n=0}^{\infty} \sum_{k=0}^{n} S_k(n) x^k y^n = \frac{1 - x}{1 - xP((1-x)y)}$$

For $x = 1$ this becomes

$$1 + \sum_{n=1}^{\infty} S(n) y^n = (1 - p_1 y)^{-1} \qquad p_1 = p(1)$$

In other words:

$$(12) \qquad S(n) = \sum_{k=1}^{n} \sum_{\Sigma a_i = n} p(a_1, \ldots, a_k) = p_1^n$$

From (11) we obtain

$$(13) \qquad \sum_{n=k}^{\infty} S_k(n) y^n = \sum_{n=0}^{k-l} \frac{(-y)^n}{n!} \frac{d^n}{dy^n} \left[\{P(y)\}^{k-l-n} \{P(y) - 1\} \right]$$

Note that $S_{n,r}(n) = p(1, \ldots, 1) = S_n(n) = E_n(n)$. In (7) set $j = 0$ and sum from $k = r + 1$ to $k = n$. We get (using (12)):

$$(14) \qquad E_r(n) = \sum_{t=r}^{n} S_r(t) p_1^{n-t}$$

Therefore

$$(15) \qquad \sum_{n=0}^{\infty} \sum_{r=0}^{n} E_r(n) x^r y^n = \frac{1}{1-p_1 y} \cdot \frac{1-x}{1-xP((1-x)y)}$$

Either this, or (13) and (14), can be considered the solution to our problem.

It is worth noting that for $j \geq 1$, if we sum (7) from $k = r + 1$ to $k = n$, we get

$$(16) \qquad E_r^{(j+1)}(n) + E_r^{(j)}(n) = \sum_{t=r}^{n} S_{r,r}^{(j)}(t) p_1^{n-t} \qquad j \geq 1$$

For example,

$$(17) \qquad \sum_{k=r}^{n} \sum_{\Sigma a_i = n} a_r^2 p(a_1, \ldots, a_k) = 2 \sum_{t=r}^{n} S_{r,r}^{(1)}(t) p_1^{n-t} - E_r(n)$$

Because of (5), we have $S_{r,r}^{(j)}(n) = S_{r,1}^{(j)}(n)$, so that we can compute $S_{r,r}^{(1)}$ from (9).

Note that the l. h. s. of (17) is the second moment M_2 of the random variable a_r; the first moment is $M_1 = E_r(n)$, so that the variance $M_2 - M_1^2$ is easily computable.

Now return to a consideration of (15). If $P(z)$ has a Taylor series expansion[14] at $z = p_1^{-1}$ then

$$E_r = \lim_{n \to \infty} E_r(n) p_1^{-n}$$

exists for all r, and

---

[14] At this point the domain of P must be specified.

$$(18)\qquad \sum_{r=0}^{\infty} E_r x^r = \frac{1-x}{1-xP\left(\frac{1-x}{p_1}\right)}$$

From this we can conclude that, if this function has no other singularities on the unit circle except at $x = 1$, and if $p_1^2 \neq p_2$, then

$$(19)\qquad \lim_{r\to\infty} E_r = \frac{p_1^2}{p_1^2-p_2},\quad \sum_{r=0}^{\infty}\left\{\frac{p_1^2}{p_1^2-p_2} - E_r\right\} = \frac{p_1(p_1p_2-p_3)}{(p_1^2-p_2)^2} \qquad \begin{matrix} p_2 = p(2) \\ p_3 = p(3) \end{matrix}$$

Before returning to our original question, we shall discuss some aspects of the question as to when $P(z)$ has a Taylor series expansion at $z = p_1^{-1}$. For the case of greatest interest, namely when all $p(a_1, \ldots, a_k) \geq 0$, it will turn out that, unless $p(n) = p_1^n$ for all n (in which case all other $p(a_1, \ldots, a_k) = 0$), $P(z)$ always has such an expansion.

Suppose $p(n) = ab^n$ for some constants a, b and all $n \geq 1$. From (2) we have $p(a_1, \ldots, a_k) = a(a-1)^{k-1}b^{a_1+\cdots+a_k}$; thus $p(n) = b^n$ implies $p(a_1, \ldots, a_k) = 0$ for $k > 1$. Clearly $a = p_1^2p_2^{-1}$, $b = p_2p_1^{-1}$. Therefore $P(z) = 1 + p_1z(1-p_1^{-1}p_2z)^{-1}$, and $P(z)$ has a Taylor series expansion at $z = p_1^{-1}$ if and only if $|a| = |p_1^2p_2^{-1}| > 1$ (in which case $E_r = a(a-1)^{-1}$).

We now show that if $p_1 > 0$ and

$$p(1,1,n) \geq 0,\ p(2,n) \geq 0 \qquad n = 1,2, \ldots$$

then either $p_1^2 = p_2$, in which case $p(n) = p_1^n$, or $p_1^2 > p_2 \geq 0$ and $P(z)$ has a Taylor expansion at $z = p_1^{-1}$.

From (6) we have

$$(20)\qquad p(1,1,n) = p_1p(1,n) - p(2,n)$$

$$(21)\qquad p(1,1,n) = p(1,1)p(n) - p(1,n+1)$$

$$(22)\qquad p(2,n) = p_2p(n) - p(n+2)$$

$$(23)\qquad p(1,n) = p_1p(n) - p(n+1)$$

Suppose $p(1,1) = p_1^2 - p_2 \neq 0$.

From (20) we have $p_1 p(1,n) \geqslant p(2,n) \geqslant 0$ so that $p(1,n) \geqslant 0$. From (21) we have $p(1,1)p(n) \geqslant p(1,n+1) \geqslant 0$ so that $p(n) \geqslant 0$. From (23) we have $p_1^2 \geqslant p_2$ and $p_1 p_2 \geqslant p_3$. From (22) we have $p_2 p(n) \geqslant p(n+2)$. It follows, by induction on n, that

$$p_1 p_2^{\frac{n-1}{2}} \geqslant p(n) \geqslant 0$$

Then

$$0 \leqslant \sum_{n=k}^{\infty} \binom{n}{k} p(n) p_1^{-n} \leqslant p_1^2 p_2^{\frac{k-1}{2}} (p_1 - p_2^{\frac{1}{2}})^{-k-1} \qquad k = 0,1, \ldots$$

The convergence of the sum in this inequality for all k is equivalent to the existance of a Taylor series for $P(z)$ at $z = p_1^{-1}$.

Now suppose $p(1,1) = p_1^2 - p_2 = 0$. From (21) we get $p(1,1,n) = -p(1,n+1)$. Since both $p(1,1,n)$ and $p(1,n+1)$ are non-negative both vanish. Then (23) implies $p(n+1) = p_1 p(n)$. It follows that $p(n) = p_1^n$ for all n, and our argument is complete.

We now return to our original problem in which $p(n) = 1/n!$ so that $P(x) = e^x$. Then (15), (18), and (19) become

$$(15') \qquad \sum_{m=0}^{\infty} \sum_{r=0}^{n} E_r(n) x^r y^n = \frac{1}{1-y} \cdot \frac{1-x}{1-xe^{(1-x)y}}$$

$$(18') \qquad \sum_{r=0}^{\infty} E_r x^r = \frac{1-x}{1-xe^{1-x}}$$

$$(19') \qquad \lim_{r \to \infty} E_r = 2, \quad \sum_{r=0}^{\infty} (2-E_r) = 4/3$$

From (15') we have

$$(24) \qquad E_r(n) = \sum_{k=r}^{n} \frac{1}{k!} \sum_{j=0}^{r} (-1)^j \binom{k+1}{j} (r-j)^k$$

From (18') we have

$$(25) \qquad E_r = r \sum_{k=0}^{r} \frac{(-1)^k}{k!} (r-k)^{k-1} e^{r-k}$$

Now consider the integral function $e^z - 1 - z$. By the Hadamard factorization theorem we can write

$$(26) \qquad e^z - 1 - z = \frac{1}{2} z^2 e^{az} \prod_{\rho} (1 - \frac{z}{\rho}) e^{\frac{z}{\rho}}$$

where the product is taken over all non-zero roots ρ of $e^z = 1 + z$.[15] We can write it in this form because (as it is easy to show) $\sum |\rho|^{-2}$ converges.

Taking the logarithmic derivatives of both sides of (26) we get

$$\frac{e^z - 1}{e^z - 1 - z} = \frac{2}{z} + a + \sum_{\rho} \frac{z}{\rho(z-\rho)} \qquad |z| < \min |\rho|$$

Thus $a = 1/3$ and

$$\frac{ze^z}{e^z - 1 - z} - \frac{2}{z} = \frac{1}{3} + \frac{z}{3} + (1+z) \sum_{\rho} \frac{z}{\rho(z-\rho)} \qquad |z| < \min |\rho|$$

Letting $z = x - 1$ we have

$$\frac{2}{1-x} - \frac{1-x}{1-xe^{1-x}} = 1 + \frac{x}{3} + x \sum_{\rho} \frac{(1-x)}{\rho(1+\rho-x)} \qquad |x| < \min |1+\rho|$$

From (18') we see that the l. h. s. is the generating function of $2 - E_r$. Computing the coefficients in the r. h. s. we have

$$(27) \qquad E_r - 2 = \sum_{\rho} (1+\rho)^{-r} = \sum_{\rho} e^{-r\rho} \qquad e^{\rho} = 1+\rho,\ \rho \neq 0;\ r = 2,3, \ldots$$

---

(15) This method is due to Professor L. Carlitz (correspondence).

which implies $E_r - 2 = O(|1+\rho_0|^{-r})$ as desired.

Another formulation of the solution in closed form obtainable from the generating function form is in terms of the coefficients of the powers of $P(y) - 1$:

$$\{P(y) - 1\}^m = \sum_{n=m}^{\infty} q_n^{(m)} y^n$$

In these terms (11) yields

$$(28) \qquad S_k(n) = \sum_{j=0}^{k} (-1)^{k-j} \binom{n-j}{k-j} q_n^{(j)}$$

Then (14) gives

$$(29) \qquad E_r(n) = \sum_{j=0}^{r} (-1)^{r-j} \sum_{t=j}^{n} \binom{t-j}{r-j} q_t^{(j)} p_1^{n-t},$$

from which we get

$$(30) \qquad E_r = \sum_{t=r}^{\infty} p_1^{-t} \sum_{j=0}^{r} (-1)^{r-j} \binom{t-j}{r-j} q_t^{(j)}$$

When $P(y) = e^y$ then $q_n^{(m)} = m! \mathcal{S}_n^{(m)}/n!$ where $\mathcal{S}_n^{(m)}$ is the Stirling number of the second kind.

Still another formulation is in terms of the Faber polynomials $\{u_n(t)\}$ generated by $Q(y) = P(y) - 1$. The generating function of $tu_n'(t)/n$ is $(1 - tQ(y))^{-1}$. Therefore

$$(31) \qquad \sum_{r=0}^{n} S_r(n) x^r = x(1-x)^{n-1} u_n'(x(1-x)^{-1})/n$$

However, the special character of e^y makes the more natural Faber polynomials, in the case $P(y) = e^y$, those generated by ye^{-y}: $\{v_n(t)\}$. In fact (25) is exactly the statement $E_r = v_r(e)$.

Another possible interpretation of the expected length of the r-th chain is the following: Assume that all sequences of "greater than" and "less than" relations are equally probable. In other words:

$$p(a_1, \ldots, a_k) = 2^{1-a_1- \cdots -a_k}$$

for all k, and all $a_1, \ldots, a_k$. It is easy to see that these values satisfy (2).

Then our formulas yield

$$E_r(n) = 2^{1-n} \sum_{k=r}^{n} \binom{n}{k} \qquad r \leqslant n = 1,2, \ldots$$

and

$$E_r = 2 \qquad r = 1,2, \ldots$$

This latter shows remarkable agreement with (27) despite the naive approach.

The reader may wish to prove for himself the following results:

1. $p(1^n) = \begin{vmatrix} p_1 & p_2 & p_3 & \cdots & p_n \\ 1 & p_1 & p_2 & \cdots & p_{n-1} \\ & 1 & p_1 & \cdots & \cdots \\ & & \ddots & \ddots & \ddots \\ & 0 & & \ddots & p_2 \\ & & & 1 & p_1 \end{vmatrix}$ where $p_n = p(n)$

and 1^n means 1 repeated n times.

2. When $p(n) = 1/n!$:

$$p(1^{a_1}, b_1+2, 1^{a_2}, b_2+2, \ldots, 1^{a_j}, b_j+2, 1^{a_{j+1}}) =$$
$$p(a_1+1, 1^{b_1}, a_2+2, 1^{b_2}, a_3+2, \ldots, a_j+2, 1^{b_j}, a_{j+1}+1)$$

for all $a_i, b_i \geqslant 0$.

3. When $\rho(n) = 1/n!$:

$$c_{n+1}(a_1, \ldots, a_k) = \sum_{a_i>1} c_n(a_1, \ldots, a_i-1, \ldots, a_k) +$$

$$\sum_{a_i>1} c_n(a_1, \ldots, a_i+a_{i+1}-1, a_{i+2}, \ldots, a_k) +$$

$$\delta_{1a_1} c_n(a_1+a_2-1, \ldots, a_k).$$

APPENDIX I

$p_n = p(n)$

n = 1

$p(1) = p_1$

n = 2

$p(2) = p_2$

$p(1,1) = p_1^2 - p_2$

n = 3

$p(3) = p_3$

$p(1,2) = p_1p_2 - p_3$

$p(1,1,1) = p_1^3 - 2p_1p_2 + p_3$

n = 4

$p(4) = p_4$

$p(1,3) = p_1p_3 - p_4$

$p(2,2) = p_2^2 - p_4$

$p(1,1,2) = p_1^2p_2 - p_2^2 - p_1p_3 + p_4$

$p(1,2,1) = p_1^2p_2 - 2p_1p_3 + p_4$

$p(1,1,1,1) = p_1^4 - 3p_1^2p_2 + p_2^2 + 2p_1p_3 - p_4$

n = 5

$p(5) = p_5$

$p(1,4) = p_1p_4 - p_5$

$p(2,3) = p_2p_3 - p_5$

$p(1,1,3) = p_1^2p_3 - p_1p_4 - p_2p_3 + p_5$

$p(1,3,1) = p_1^2p_3 - 2p_1p_4 + p_5$

$p(1,2,2) = p_1p_2^2 - p_1p_4 - p_2p_3 + p_5$

$p(2,1,2) = p_1p_2^2 - 2p_2p_3 + p_5$

$p(1,1,1,2) = p_1^3p_2 - p_1^2p_3 - 2p_1p_2^2 + 2p_2p_3 + p_1p_4 - p_5$

n = 5 (cont.)

$p(1,1,2,1) = p_1^3p_2 - 2p_1^2p_3 - p_1p_2^2 + p_2p_3 + 2p_1p_4 - p_5$

$p(1,1,1,1,1) = p_1^5 - 4p_1^3p_2 + 3p_1^2p_3 + 3p_1p_2^2 - 2p_2p_3 - 2p_1p_4 + p_5$

n = 6

$p(6) = p_6$

$p(1,5) = p_1p_5 - p_6$

$p(2,4) = p_2p_4 - p_6$

$p(3,3) = p_3^2 - p_6$

$p(1,1,4) = p_1^2p_4 - p_1p_5 - p_2p_4 + p_6$

$p(1,4,1) = p_1^2p_4 - 2p_1p_5 + p_6$

$p(1,2,3) = p_1p_2p_3 - p_1p_5 - p_3^2 + p_6$

$p(1,3,2) = p_1p_2p_3 - p_1p_5 - p_2p_4 + p_6$

$p(2,1,3) = p_1p_2p_3 - p_2p_4 - p_3^2 + p_6$

$p(2,2,2) = p_2^3 - 2p_2p_4 + p_6$

$p(1,1,1,3) = p_1^3p_3 - p_1^2p_4 - 2p_1p_2p_3 + p_1p_5 + p_2p_4 + p_3^2 - p_6$

$p(1,1,3,1) = p_1^3p_3 - 2p_1^2p_4 - p_1p_2p_3 + 2p_1p_5 + p_2p_4 - p_6$

$p(1,1,2,2) = p_1^2p_2^2 - p_1^2p_4 - p_1p_2p_3 - p_2^3 + p_1p_5 + 2p_2p_4 - p_6$

$p(1,2,1,2) = p_1^2p_2^2 - 3p_1p_2p_3 + p_1p_5 + p_2p_4 + p_3^2 - p_6$

$p(1,2,2,1) = p_1^2p_2^2 - p_1^2p_4 - 2p_1p_2p_3 + 2p_1p_5 + p_3^2 - p_6$

n = 6 (cont.)

$p(2,1,1,2) = p_1^2p_2^2-2p_1p_2p_3-p_2^3+2p_2p_4+p_3^2-p_6$

$p(1,1,1,1,2) = p_1^4p_2-p_1^3p_3-3p_1^2p_2^2+p_1^2p_4+4p_1p_2p_3+p_2^3-p_1p_5-2p_2p_4-p_3^2+p_6$

$p(1,1,1,2,1) = p_1^4p_2-2p_1^3p_3-2p_1^2p_2^2+2p_1^2p_4+4p_1p_2p_3-2p_1p_5-p_2p_4-p_3^2+p_6$

$p(1,1,2,1,1) = p_1^4p_2-2p_1^3p_3-2p_1^2p_2^2+3p_1^2p_4+2p_1p_2p_3+p_2^3-2p_1p_5-2p_2p_4+p_6$

$p(1,1,1,1,1,1) = p_1^6-5p_1^4p_2+4p_1^3p_3+6p_1^2p_2^2-3p_1^2p_4-6p_1p_2p_3-p_2^3+2p_1p_5+2p_2p_4+p_3^2-p_6$

APPENDIX II

c ≡ c_2

c(2) = 1

c(1,1) = 1

c ≡ c_3

c(3) = c(1,1,1) = 1

c(1,2) = c(2,1) = 2

c ≡ c_4

c(4) = c(1,1,1,1) = 1

c(3,1) = c(1,3) = c(2,1,1) = c(1,1,2)=3

c(2,2) = c(1,2,1) = 5

c ≡ c_5

c(5) = c(1,1,1,1,1) = 1

c(4,1)=c(1,4)=c(2,1,1,1)=c(1,1,1,2)=4

c(3,1,1)=c(1,1,3)=6

c(3,2)=c(2,3)=c(1,2,1,1)=c(1,1,2,1)=9

c(1,3,1)=c(2,1,2)=11

c(2,2,1)=c(1,2,2)=16

c ≡ c_6

c(6) = c(1,1,1,1,1,1) = 1

c ≡ c_6 (cont.)

c(5,1)=c(1,5)=c(2,1,1,1,1)=c(1,1,1,1,2)=5

c(4,1,1)=c(1,1,4)=c(3,1,1,1)=c(1,1,1,3)=10

c(4,2)=c(2,4)=c(1,2,1,1,1)=c(1,1,1,2,1)=14

c(3,3)=c(1,4,1)=c(2,1,1,2)=c(1,1,2,1,1)=19

c(2,1,3)=c(3,1,2)=c(1,3,1,1)=c(1,1,3,1)=26

c(1,2,3)=c(3,2,1)=c(2,2,1,1)=c(1,1,2,2)=35

c(1,3,2)=c(2,3,1)=c(2,1,2,1)=c(1,2,1,2)=40

c(2,2,2)=c(1,2,2,1)=61

c ≡ c_7

c(7)=c(1,1,1,1,1,1,1)=1

c(6,1)=c(6,1)=c(2,1,1,1,1,1)=c(1,1,1,1,1,2) =6

c(5,1,1)=c(1,1,5)=c(3,1,1,1,1)=c(1,1,1,1,3) =15

c(4,1,1,1)=c(1,1,1,4)=c(2,5)=c(5,2)= c(1,2,1,1,1,1)=c(1,1,1,1,2,1)=20

c(1,5,1)=c(2,1,1,1,2)=29

c(4,3)=c(3,4)=c(1,1,2,1,1,1)=c(1,1,1,2,1,1) =34

c(4,1,2)=c(2,1,4)=c(1,3,1,1,1)=c(1,1,1,3,1) =50

$c \equiv c_7$ (cont.)

c(1,4,1,1)=c(1,1,4,1)=c(3,1,1,2)=c(2,1,1,3)=55

c(1,2,4)=c(4,2,1)=c(2,2,1,1,1)=c(1,1,1,2,2)=64

c(3,1,3)=c(1,1,3,1,1)=71

c(1,4,2)=c(2,4,1)=c(2,1,1,2,1)=c(1,2,1,1,2)=78

c(3,2,1,1)=c(1,1,2,3)=90

c(3,3,1)=c(1,3,3)=c(2,1,2,1,1)=c(1,1,2,1,2)=99

c(3,1,2,1)=c(1,2,1,3)=c(2,3,1,1)=c(1,1,3,2)=111

c(2,1,3,1)=c(1,3,1,2)=132

c(3,2,2)=c(2,2,3)=c(1,2,2,1,1)=c(1,1,2,2,1)=155

c(2,3,2)=c(1,2,1,2,1)=169

c(2,2,1,2)=c(2,1,2,2)=c(1,2,3,1)=c(1,3,2,1)=181

c(2,2,2,1)=c(1,2,2,2)=272

8. APPROXIMATIONS OF FINITE SET STRUCTURES[16]

If a symmetric relation on a set of elements is also transitive, it is called an equivalence relation. (We assume all relations discussed are reflexive.) An equivalence relation determines a partitioning of the set into equivalence classes such that two elements are related if and only if they are in the same class. We are interested here in how a general symmetric relation (i.e. not necessarily transitive) can be used to define a partition into approximate equivalence classes, in some suitable sense. That is, we seek a partition corresponding to an equivalence relation which "approximates" the given relation.

Given any symmetric relation R (which can be represented as a linear graph with no circuits of one or two edges), and also a partition P of the set on which R is defined (i.e., the vertex-set of the linear graph), let D_P be the number of pairs of R-related pairs of elements which lie in different sets of P, and let S_P be the number of R-unrelated pairs of elements which lie in the same set of P. An approximate partition, A, for R would be one in which D_A and S_A are both small. In fact, $D_A = S_A = 0$ if and only if R is an equivalence relation and A is the partition into the equivalence classes of R. Hence a natural definition of an approximate equivalence partition for R, is a partition P which minimizes

---

[16] The preliminary study reported here, due to John Edmonds, has been superseded by the author's later work.

$M(D_P, S_P)$, where $M(D, S)$ is a decreasing function of D and S which vanishes only when $D = S = 0$. $M = D + S$ is of course the simplest.

To minimize $M_P = D_P + S_P$ one can begin with any partition and move one element after another from one set to another (possibly empty) so that each move decreases M (or least does not increase M). Eventually one reaches a relative minimum beyond which one cannot go by such moves. Unfortunately, however, this relative minimum is not necessarily an absolute minimum. For an example of this phenomenon, consider a set with 3n elements:

$$\{a_1, b_1, c_1, a_2, b_2, c_2, \ldots, a_n, b_n, c_n\} .$$

A symmetric relation R on this set is obtained by requiring that a_i and a_j be R-related for $1 \leq i, j \leq n$, and that each two of a_i, b_i, c_i be R-related for $1 \leq i \leq n$. First partition the set into the subsets $\{a_i, b_i, c_i\}$. Call this partition P; we have

$$D_P = \tfrac{1}{2}n(n-1), \quad S_P = 0, \quad M_P = \tfrac{1}{2}n(n-1).$$

Second, partition the set into $\{a_i, \ldots, a_n\}$ and the subsets $\{b_i, c_i\}$; calling this partition Q, we have

$$D_Q = 2n, \quad S_Q = 0, \quad M_Q = 2n$$

Thus $M_P > M_Q$, so that P does not yield an absolute minimum, for $n > 5$. P does however yield a relative minimum in the sense described above. To see this, observe that (beginning with P and) moving any b_i or c_i to another set of the partition would increase D_P and not decrease S_P, while moving any a_i to another set decreases D_P by no more than 1 but increases S_P by 2.

It would be interesting to know whether the following stronger types of moves would always enable one to reach an absolute minimum:

1. Move a completely R-related set[17] of elements to one of the partition sets so that M is not increased.

2. Move a completely R-unrelated set of elements from one partition set to another, so that M is not increased.

One can similarly define approximating partitions of the elements for a general class C of sets of elements. Let M_{ij} be the number of sets of C containing both elements $\underline{i}$ and $\underline{j}$, and let N_{ij} be the number of sets of C containing one but not both of these elements. For a partition P of the elements we write "(i, j) in P" for distinct elements $\underline{i}$ and $\underline{j}$ if and only if they are in the same set of P. Then one kind of approximating partition for C is a partition P which minimizes

$$M_P^* = \sum\{M_{ij}:(i, j)\text{not in P}\} + \sum\{N_{ij}:(i, j)\text{ in P}\},$$

i.e. which (loosely speaking) avoids putting together elements which are less frequently together in sets of C, and avoids separating elements which are more frequently together in C. Note that $M_P^* = 0$ only when C = P (so that C is a partition).

A second kind of approximating partition for C is a partition P which maximizes

---

[17] A set is called completely R-related if each two of its elements are R-related.

$$M_P^{**} = \Sigma\{M_{ij}:(i, j) \text{ in } P\} + \Sigma\{N_{ij}:(i, j) \text{ not in } P\},$$

i.e. which (loosely speaking) puts together elements which are more frequently together in C, and separates elements which are less frequently together in C. The partitions which minimize M_P^* are in general not the same as those which maximize M_P^{**}. As a compromise, one might seek to minimize $M_P^* - M_P^{**}$, or M_P^*/M_P^{**}.

Approximating a relation by a partition, as discussed above, is <u>not</u> a special case of approximating a class of sets by a partition. A symmetric relation R does of course give rise to a class of sets, namely the class C(R)of R-related pairs. The aspect of R which we approximate (in the previous discussion involving M_P) is however not the same as that which we treat in approximating C(R). Hence we should not expect M_P^* and M_P^{**} to be generalizations of $M_P = D_P + S_P$; the latter <u>does</u> generalize to

$$M_P^{***} = \Sigma\{M_{ij}:(i, j) \text{ not in } P\} + \Sigma\{1 - M_{ij}:(i, j) \text{ in } P\},$$

whose minimization by suitable P defines a third kind of "approximate partition" for a class C of sets.

We now describe two structures quite different from the relations and classes already discussed. Because of their frequent use in information storage, it might be of interest to find approximations of one by the other more restricted one.

The first structure we call a catalog because it is an abstraction of a library subject catalog. A catalog consists of sets (corresponding to subjects) and elements (corresponding to books) such that (1) any set may contain as members other sets as well as elements and such that (2) every element X contained in a set S is also contained in every set member of S. Of course to avoid an infinite sequence of inclusions there must be sets containing no sets. A set member R of a set S corresponds to a sub-subject R of a subject S. An element of set S with set members S_i corresponds to a book X which treats to some extent the entire subject S and not simply a portion of the sub-subjects S_i.

By disregarding in a catalog every occurrence of an element X in a set contained in a set containing X we obtain a structure which determines uniquely the catalog from which it is derived. Structures like this are such that (2a) no element X is a member of a member of ...a member of a set containing X. Any such structure determines uniquely a catalog from which it is uniquely derived as above. We call these structures, satisfying (1) and (2a), derived catalogs.

A derived catalog such that every element is contained in exactly one set is called a file because of its correspondence with the manner of actually grouping the books on shelves, there being only one copy of each book.

A method often used by a library is to have a fairly natural card cataloging by subjects, each element occurrence of which gives the location of the corresponding book in the fairly artificial filing of

the books. (A library also orders the elements in every set of a catalog or file.) However, it is convenient for the browser if the filing approximates the (derived) catalog. In most catalogs in actual use, every element has a most important set occurrence and these occurrences determine the approximating file. This is the method used for filing patents in the Patent Office search room. However in the absence of a ranking of occurrences, we might still ask how might we approximate a catalog of elements by a file of these elements.

USCOMM-NBS-DC

U. S. DEPARTMENT OF COMMERCE
Luther H. Hodges, *Secretary*

NATIONAL BUREAU OF STANDARDS
A. V. Astin, *Director*

THE NATIONAL BUREAU OF STANDARDS

The scope of activities of the National Bureau of Standards at its major laboratories in Washington, D.C., and Boulder, Colorado, is suggested in the following listing of the divisions and sections engaged in technical work. In general, each section carries out specialized research, development, and engineering in the field indicated by its title. A brief description of the activities, and of the resultant publications, appears on the inside of the front cover.

WASHINGTON, D.C.

Electricity. Resistance and Reactance. Electrochemistry. Electrical Instruments. Magnetic Measurements. Dielectrics.

Metrology. Photometry and Colorimetry. Refractometry. Photographic Research. Length. Engineering Metrology. Mass and Scale. Volumetry and Densimetry.

Heat. Temperature Physics. Heat Measurements. Cryogenic Physics. Equation of State. Statistical Physics.

Radiation Physics. X-ray. Radioactivity. Radiation Theory. High Energy Radiation. Radiological Equipment. Nucleonic Instrumentation. Neutron Physics.

Analytical and Inorganic Chemistry. Pure Substances. Spectrochemistry. Solution Chemistry. Standard Reference Materials. Applied Analytical Research.

Mechanics. Sound. Pressure and Vacuum. Fluid Mechanics. Engineering Mechanics. Rheology. Combustion Controls.

Organic and Fibrous Materials. Rubber. Textiles. Paper. Leather. Testing and Specifications. Polymer Structure. Plastics. Dental Research.

Metallurgy. Thermal Metallurgy. Chemical Metallurgy. Mechanical Metallurgy. Corrosion. Metal Physics. Electrolysis and Metal Deposition.

Mineral Products. Engineering Ceramics. Glass. Refractories. Enameled Metals. Crystal Growth. Physical Properties. Constitution and Microstructure.

Building Research. Structural Engineering. Fire Research. Mechanical Systems. Organic Building Materials. Codes and Safety Standards. Heat Transfer. Inorganic Building Materials.

Applied Mathematics. Numerical Analysis. Computation. Statistical Engineering. Mathematical Physics. Operations Research.

Data Processing Systems. Components and Techniques. Digital Circuitry. Digital Systems. Analog Systems. Applications Engineering.

Atomic Physics. Spectroscopy. Infrared Spectroscopy. Solid State Physics. Electron Physics. Atomic Physics.

Instrumentation. Engineering Electronics. Electron Devices. Electronic Instrumentation. Mechanical Instruments. Basic Instrumentation.

Physical Chemistry. Thermochemistry. Surface Chemistry. Organic Chemistry. Molecular Spectroscopy. Molecular Kinetics. Mass Spectrometry.

Office of Weights and Measures.

BOULDER, COLO.

Cryogenic Engineering. Cryogenic Equipment. Cryogenic Processes. Properties of Materials. Cryogenic Technical Services.

Ionosphere Research and Propagation. Low Frequency and Very Low Frequency Research. Ionosphere Research. Prediction Services. Sun-Earth Relationships. Field Engineering. Radio Warning Services.

Radio Propagation Engineering. Data Reduction Instrumentation. Radio Noise. Tropospheric Measurements. Tropospheric Analysis. Propagation-Terrain Effects. Radio-Meteorology. Lower Atmosphere Physics.

Radio Standards. High Frequency Electrical Standards. Radio Broadcast Service. Radio and Microwave Materials. Atomic Frequency and Time Interval Standards. Electronic Calibration Center. Millimeter-Wave Research. Microwave Circuit Standards.

Radio Systems. High Frequency and Very High Frequency Research. Modulation Research. Antenna Research. Navigation Systems.

Upper Atmosphere and Space Physics. Upper Atmosphere and Plasma Physics. Ionosphere and Exosphere Scatter. Airglow and Aurora. Ionospheric Radio Astronomy.


NBS