

Traces, Term Ranks, Widths and Heights*

Abstract: The notions of widths and heights of (0, 1)-matrices are discussed in the general setting of known results concerning traces and term ranks. Proofs are omitted throughout.

1. The (0, 1)-matrix

Let A be a matrix of m rows and n columns and let the entries of A be the integers 0 and 1. Such a matrix is called a (0, 1)-matrix of size m by n . For a specified size m by n there are 2^{mn} such matrices. This finite subset of the set of all m by n matrices with real elements is of fundamental importance in combinatorial investigations. One of the main reasons for this is the following. Let X_1, X_2, \dots, X_m be m subsets of a set X of n elements x_1, x_2, \dots, x_n . Let $a_{ij} = 1$ if x_j is a member of X_i and let $a_{ij} = 0$ if x_j is not a member of X_i . In this way we may define a (0, 1)-matrix $A = [a_{ij}]$ of size m by n . This matrix is called the *incidence matrix* for the subsets X_1, X_2, \dots, X_m of X . The 1's in row i of A designate the elements that occur in set X_i and the 1's in column j of A designate the sets that contain element x_j . Thus A characterizes these m subsets of X . A (0, 1)-matrix of size m by n may also be regarded as a punched card having m rows and n columns. The 1's in the matrix correspond to punches and the 0's to blanks, or vice versa. From this very concrete viewpoint it is evident that the (0, 1)-matrix is a convenient device for the systematic storage of information.

With each (0, 1)-matrix A , one may associate in various ways integers that give insight into the combinatorial structure of A . Examples of such integers are the trace and the term rank of A . The authors have described the *width* and the *height* of A [7]. The purpose of the present paper is to consider these two new concepts in the general setting

of the previously studied theory for traces and term ranks. We omit proofs but supply references to the literature.

2. Traces and term ranks

Let A be a (0, 1)-matrix of size m by n . The *trace* σ of A is defined by

$$\sigma = \sum_{i=1}^t a_{ii}, \quad t = \min(m, n). \quad (2.1)$$

This quantity is certainly elementary and causes no computational difficulties. If A is the incidence matrix for the subsets X_1, X_2, \dots, X_m of X , then the trace σ counts the number of times that x_i is a member of X_i ($i = 1, 2, \dots, t$).

The *term rank* ρ of A is the maximal trace obtained from A under arbitrary permutations of the rows and of the columns of A [18]. In other words, the term rank of A is the maximal number of 1's that may be chosen in A with no two in the same row or column. It is evident from the definition that the term rank of A is invariant under arbitrary permutations of the rows and of the columns of A . Combinatorially this means that for incidence matrices the term rank is independent of the particular labelling of elements x_1, x_2, \dots, x_n and subsets X_1, X_2, \dots, X_m of X . Indeed, the term rank is the maximal integer ρ for which there exists a labelling of elements x'_1, x'_2, \dots, x'_n and subsets X'_1, X'_2, \dots, X'_m such that x'_i is a member of X'_i for $i = 1, 2, \dots, \rho$. No discussion of term rank would be complete without mentioning the classical theorem in the subject [13]. It asserts that the term rank ρ equals the minimal number of rows and columns that

*The work of the second author was supported in part by the Office of Ordnance Research.

†Mathematics Division, The RAND Corporation.

‡Professor of Mathematics, Ohio State University.

contain all 1's of A . Thus if

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

then both quantities are equal to 3. The problem of evaluating the term rank ρ of a $(0, 1)$ -matrix A can be viewed either as a maximal network flow problem or an optimal assignment problem. Consequently efficient computational methods are available for determining term rank [1, 3, 14]. The special case that occurs when $\rho = m$ deserves mention. Then the subsets X_1, X_2, \dots, X_m are said to possess a *system of distinct representatives*. This topic has an extensive literature. See, for example, [4, 10, 11, 12, 15, 16, 17].

3. Widths and heights

Let A be a $(0, 1)$ -matrix of size m by n . Let the sum of row i of A be denoted by r_i and let the sum of column j of A be denoted by s_j . We call $R = (r_1, r_2, \dots, r_m)$ the *row sum vector* and $S = (s_1, s_2, \dots, s_n)$ the *column sum vector* of A .

Suppose that

$$r_1 \geq r_2 \geq \dots \geq r_m > 0 \quad (3.1)$$

$$s_1 \geq s_2 \geq \dots \geq s_n > 0. \quad (3.2)$$

Then we call A *normalized*. Henceforth we take A normalized. This restriction is frequently a convenience rather than a necessity.

Let α be an integer in the interval

$$1 \leq \alpha \leq r_m \quad (3.3)$$

and let ϵ be an integer in the interval

$$1 \leq \epsilon \leq n. \quad (3.4)$$

Suppose that the normalized A has an m by ϵ submatrix E^* , each of whose row sums is at least α . Then the ϵ columns of E^* are said to form an α -set of *representatives* for the matrix A . Let $\epsilon(\alpha)$ be the minimal number of columns of A that form an α -set of representatives for A . Such a column set is called a *minimal α -set of representatives* for A and $\epsilon(\alpha)$ is called the α -width of A . The integer α and the matrix A uniquely determine $\epsilon(\alpha)$. We note that the α -width $\epsilon(\alpha)$ of A is invariant under arbitrary permutations of the rows and of the columns of A . However, the α -width of the transpose of A may differ drastically from that of A . This is not the case for the trace and term rank of A , both of which remain unchanged under transposition.

Let E^* be a submatrix of A of size m by $\epsilon(\alpha)$ that yields a minimal α -set of representatives for A . Let E be the submatrix of E^* composed of all of the rows of E^* that contain α 1's and $\epsilon(\alpha) - \alpha$ 0's. The matrix E is called a *critical α -submatrix* of A . Note that E cannot be empty since if all row sums of E^* exceed α , then deletion of a column of E^* yields an α -set of representatives for A , contradicting the minimality of $\epsilon(\alpha)$. It follows without difficulty that *the normalized matrix A has an α -width $\epsilon(\alpha)$ for each α in the interval $1 \leq \alpha \leq r_m$. A critical α -submatrix E of A contains no zero columns.*

Each of the critical α -submatrices E of A must contain $\epsilon(\alpha)$ columns. But the number of rows in the various critical α -submatrices need not be fixed. Let E be a critical α -submatrix containing the minimal number of rows $\delta(\alpha)$. The positive integer $\delta(\alpha)$ is called the α -height of A . Both $\epsilon(\alpha)$ and $\delta(\alpha)$ are basic invariants of A . Evidently

$$\epsilon(1) < \epsilon(2) < \dots < \epsilon(r_m) \quad (3.5)$$

and

$$\delta(1) \geq \epsilon(1). \quad (3.6)$$

Thus if

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix},$$

then $\epsilon(1) = 2$, $\epsilon(2) = 4$, $\delta(1) = 4$, $\delta(2) = 4$.

The preceding discussion has an important set theoretic interpretation. Let A be the incidence matrix for the subsets X_1, X_2, \dots, X_m of X . No loss is entailed by regarding A as normalized. A minimal α -set of representatives for A yields a subset X^* of $\epsilon(\alpha)$ elements of X . X^* has the property that each $X_i \cap X^*$ contains at least α elements ($i = 1, 2, \dots, m$). No subset of X containing fewer than $\epsilon(\alpha)$ elements possesses this property. At least $\delta(\alpha)$ of the sets $X_i \cap X^*$ contain exactly α elements. If $\alpha = 1$, then X^* has the property that each $X_i \cap X^*$ is nonempty, and no subset of X containing fewer than $\epsilon(1)$ elements possesses this property.

For a concrete example, consider the problem of determining the fewest number of nodes or junction points in a network that touch all links of the network. Here we may regard X as the set of all nodes, a link as a subset of two nodes (its ends) of X . Then the problem is to find the fewest number of nodes that "1-represent" all the links, that is, the 1-width of the incidence matrix A of links vs. nodes. The famous "eight queens" chess board

problem is of this type. Here one forms a network by connecting two cells of the chess board if a queen can move from one to the other. Then the complement of a minimal system of cells that touch all links are positions in which the maximal number of queens can be placed so that no two attack each other.

Very little is known concerning good computational methods for determining widths and heights of (0, 1)-matrices. Efficient algorithms in this domain would be of great interest.

4. The class $\mathcal{Q}(R, S)$

Let A be a normalized (0, 1)-matrix of size m by n with row sum vector $R = (r_1, r_2, \dots, r_m)$ and column sum vector $S = (s_1, s_2, \dots, s_n)$. The vectors R and S determine a class

$$\mathcal{Q} = \mathcal{Q}(R, S) \quad (4.1)$$

consisting of all (0, 1)-matrices of size m by n , with row sum vector R and column sum vector S . Simple necessary and sufficient conditions on R and S are available in order that the class \mathcal{Q} be nonempty [8, 19]. We always take \mathcal{Q} nonempty and refer to \mathcal{Q} as the *normalized class* $\mathcal{Q}(R, S)$.

Let A be in the normalized class \mathcal{Q} and consider the 2 by 2 submatrices of A of the types

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

An *interchange* is a transformation of the elements of A that changes a minor of type A_1 into A_2 , or vice versa, and leaves all other elements of A unaltered. An interchange is in a sense the most elementary operation that may be applied to A to yield a new matrix within the class $\mathcal{Q}(R, S)$. The interchange theorem [19] asserts that if A and A' belong to $\mathcal{Q}(R, S)$, then A is transformable into A' by interchanges. This theorem is a very useful one for the study of the class $\mathcal{Q}(R, S)$.

Let A be in the normalized class $\mathcal{Q}(R, S)$ and write

$$A = \begin{bmatrix} W & X \\ Y & Z \end{bmatrix}, \quad (4.2)$$

where W is of size e by f ($0 \leq e \leq m$; $0 \leq f \leq n$). For an arbitrary (0, 1)-matrix Q , let $N_0(Q)$ denote the number of 0's in Q , and $N_1(Q)$ the number of 1's in Q . Let

$$t_{ef} = N_0(W) + N_1(Z) \quad (e = 0, 1, \dots, m; f = 0, 1, \dots, n) \quad (4.3)$$

and define

$$T = [t_{ef}] \quad (e = 0, 1, \dots, m; f = 0, 1, \dots, n). \quad (4.4)$$

T is called the *structure matrix* of the class \mathcal{Q} [21]. It follows at once from (4.3) that

$$t_{ef} = ef + (r_{e+1} + r_{e+2} + \dots + r_m) - (s_1 + s_2 + \dots + s_f) \quad (e = 0, 1, \dots, m; f = 0, 1, \dots, n). \quad (4.5)$$

Thus the structure matrix is independent of the particular choice of A in \mathcal{Q} .

The structure matrix contains a wealth of information concerning the class $\mathcal{Q}(R, S)$, as will be evident in succeeding sections. Here we mention only the following fact. If $r_1 \geq r_2 \geq \dots \geq r_m$, $s_1 \geq s_2 \geq \dots \geq s_n$, with $\sum_i r_i = \sum_j s_j$, and if t_{ef} is defined by (4.5), then a necessary and sufficient condition that \mathcal{Q} be nonempty is that $t_{ef} \geq 0$ for all e, f . This can either be seen directly from the max-flow min-cut theorem for network flows [3, 4] or can be deduced from the conditions stated in [8, 19]. Since we are dealing throughout with a nonempty class and have defined T by (4.3) rather than (4.5), its entries are of course nonnegative integers. It may also be seen that T satisfies the equation

$$E_m \begin{bmatrix} \tau & -s_1 & \dots & -s_n \\ -r_1 & & & \\ \vdots & & J & \\ -r_m & & & \end{bmatrix} E_n^T = T. \quad (4.6)$$

Here E_k is the triangular matrix of order $k + 1$ with 1's on and below the main diagonal. E_k^T is its transpose, J is the m by n matrix with all entries 1, and τ is the total number of 1's in a matrix A of the normalized $\mathcal{Q}(R, S)$ [21].

5. The fundamental formulas

We now discuss traces, term ranks, widths and heights for the matrices in the normalized class $\mathcal{Q}(R, S)$. We begin with the trace. Each matrix A in $\mathcal{Q}(R, S)$ has a trace σ . Let $\bar{\sigma}$ be the minimal and let $\bar{\sigma}$ be the maximal trace for the matrices in the normalized class $\mathcal{Q}(R, S)$. It is natural to attempt to determine the integers $\bar{\sigma}$ and $\bar{\sigma}$ explicitly. Unusually simple formulas are available in terms of the elements of the structure matrix T , namely

$$\bar{\sigma} = \max_{e,f} [\min(e, f) - t_{ef}] \quad (5.1)$$

$$\bar{\sigma} = \min_{e,f} [t_{ef} + \max(e, f)]$$

$$(e = 0, 1, \dots, m; f = 0, 1, \dots, n). \quad (5.2)$$

Formulas (5.1) and (5.2) are derived in [21].

Each matrix A in the normalized class $\mathcal{G}(R, S)$ has a term rank ρ . Let $\bar{\rho}$ be the minimal and let \bar{p} be the maximal term rank for the matrices in $\mathcal{G}(R, S)$. A remarkable formula is available for \bar{p} [20]

$$\bar{p} = \min_{e,f} [t_{ef} + (e + f)]$$

$$(e = 0, 1, \dots, m; f = 0, 1, \dots, n). \quad (5.3)$$

The derivation of (5.3) is not simple. However, \bar{p} appears to be an even more elusive and difficult quantity to handle. Haber has investigated \bar{p} carefully and devised an effective algorithm for its evaluation [9].

We next discuss the recent investigation by the authors involving widths and heights [7]. Let $1 \leq \alpha \leq r_m$. Then each A in the normalized class $\mathcal{G}(R, S)$ determines an α -width $\epsilon(\alpha)$ and an α -height $\delta(\alpha)$. For each α let the minimum of these $\epsilon(\alpha)$'s over all A in $\mathcal{G}(R, S)$ be denoted by

$$\bar{\epsilon} = \bar{\epsilon}(\alpha). \quad (5.4)$$

We call $\bar{\epsilon} = \bar{\epsilon}(\alpha)$ the *minimal α -width* of the class $\mathcal{G}(R, S)$. Let

$$\bar{\delta} = \bar{\delta}(\alpha) \quad (5.5)$$

equal the minimum of the α -heights $\delta(\alpha)$ over all matrices $A_{\bar{\epsilon}}$ of α -width $\bar{\epsilon}(\alpha)$ in $\mathcal{G}(R, S)$.

Now let

$$N(\epsilon, e, f) = t_{ef} + (s_1 + s_2 + \dots + s_e) - e\epsilon, \quad (5.6)$$

where ϵ, e, f are integer parameters such that

$$0 \leq \epsilon \leq n, \quad (5.7)$$

$$0 \leq e \leq m, \quad (5.8)$$

$$\epsilon \leq f \leq n. \quad (5.9)$$

One may deduce the following [7]:

Theorem 5.1

The minimal α -width $\bar{\epsilon}(\alpha)$ equals the first nonnegative integer ϵ such that

$$N(\epsilon, e, f) \geq \alpha(m - e) \quad (5.10)$$

for all integer parameters e and f restricted by $0 \leq e \leq m$ and $\epsilon \leq f \leq n$. Let

$$\bar{\gamma} = \min_{e,f} [N(\bar{\epsilon} - 1, e, f) + \alpha\bar{\epsilon}], \quad (5.11)$$

where $0 \leq e \leq m$ and $\bar{\epsilon} - 1 \leq f \leq n$. Then

$$\bar{\delta}(\alpha) = (\alpha + 1)m - \bar{\gamma} - s_{\bar{\epsilon}}. \quad (5.12)$$

For each α let the maximum of the α -widths $\epsilon(\alpha)$ over all matrices A in $\mathcal{G}(R, S)$ be denoted by

$$\bar{\epsilon} = \bar{\epsilon}(\alpha). \quad (5.13)$$

We call $\bar{\epsilon} = \bar{\epsilon}(\alpha)$ the *maximal α -width* of the class $\mathcal{G}(R, S)$. Almost nothing is known about $\bar{\epsilon}(\alpha)$ but it seems certain that its behavior is decidedly more intricate than that of $\bar{\epsilon}(\alpha)$.

A direct application of the interchange theorem allows us to prove that if ϵ is an integer in the interval

$$\bar{\epsilon}(\alpha) \leq \epsilon \leq \bar{\epsilon}(\alpha), \quad (5.14)$$

then there exists a matrix A_{ϵ} in $\mathcal{G}(R, S)$ of α -width ϵ [7]. The analogous result for term ranks is also valid [20]. Traces for matrices in the normalized class $\mathcal{G}(R, S)$ usually take on all integer values in the interval $\bar{\sigma} \leq \sigma \leq \bar{\sigma}$, but certain classes exclude $\bar{\sigma} + 1$ and others exclude $\bar{\sigma} - 1$ [21].

Concluding remarks

In this section we give a brief discussion concerning the proofs of the combinatorial formulas described in Section 5. Formulas (5.1), (5.2), (5.3), and Theorem 5.1 have all been derived by first constructing a kind of canonical matrix in the class $\mathcal{G}(R, S)$. By this we merely mean that the existence of a matrix A is established with certain very special properties. These properties are such that they make the combinatorial formula more or less apparent. The canonical matrix is constructed by whatever techniques are available. Here the efficient use of interchanges is frequently a very powerful tool. We illustrate by stating the following theorem [7]:

Theorem 6.1

Let $\bar{\epsilon} = \bar{\epsilon}(\alpha)$ be the minimal α -width of the normalized class $\mathcal{G}(R, S)$ and let $\bar{\delta} = \bar{\delta}(\alpha)$ be the minimum of the α -heights $\delta(\alpha)$ over all matrices $A_{\bar{\epsilon}}$ of α -width $\bar{\epsilon}$ in $\mathcal{G}(R, S)$. Then there exists a matrix $A_{\bar{\epsilon}}$ of α -width $\bar{\epsilon}$ in $\mathcal{G}(R, S)$ of the form

$$A_{\bar{\epsilon}} = \left[\begin{array}{c|c|c} M & J & * \\ \hline F & * & 0 \\ \hline E & & \end{array} \right]. \quad (6.1)$$

Here E is a critical submatrix of $A_{\bar{\epsilon}}$ of size $\bar{\delta}$ by $\bar{\epsilon}$. M is a matrix of size e by $\bar{\epsilon}$ with $\alpha + 1$ or more 1's in each row. F is a matrix of size $m - (e + \bar{\delta})$ by $\bar{\epsilon}$ with exactly $\alpha + 1$ 1's in each row. J is a matrix of size e by $f - \bar{\epsilon}$ consisting entirely of 1's, and 0 is a zero matrix. Each of the first $\bar{\epsilon}$ columns of $A_{\bar{\epsilon}}$ contains

more than $m - \delta$ 1's. The degenerate cases $e = 0$, $e + \delta = m$, $f = \epsilon$, and $f = n$ are not excluded.

Theorem 6.1 makes the derivation of Theorem 5.1 a relatively easy task, once one has succeeded in "guessing" the appropriate formula (5.10). Of course the existence of canonical matrices of the form (6.1) is of interest in its own right and gives us considerable insight into the structure of the class $\alpha(R, S)$.

Recently certain combinatorial results have been derived by the use of network flows [3, 4, 5, 6, 8, 12]. This approach has been on the whole very successful. Network flows are effective in dealing with problems involving the trace [6]. Theorem 5.1 may also be derived by flow theory. But the use of flows appears

to be ineffective in dealing with maximal and minimal term rank, or maximal width. For example, no flow derivation has been obtained for the \bar{p} formula (5.3) although several attempts have been made in this direction. Perhaps the most attractive feature about using flows to derive combinatorial results is that, if one is successful in obtaining a flow formulation of the problem, usually little subsequent guesswork is involved. For instance, a flow formulation of the minimal width problem leads one rather directly to the conclusion that (5.10) is the appropriate formula for $\bar{\epsilon}(\alpha)$. At the present, however, there appears to be no truly systematic way of dealing with combinatorial problems of the kind we have discussed.

References

1. G. B. Dantzig, "Application of the Simplex Method to a Transportation Problem," *Cowles Commission Monograph*, 13, 359-373 (1951).
2. A. L. Dulmage, and N. S. Mendelsohn, "The Term and Stochastic Ranks of a Matrix," *Can. J. Math.*, 11, 269-279 (1959).
3. L. R. Ford, Jr., and D. R. Fulkerson, "A Simple Algorithm for Finding Maximal Network Flows and an Application to the Hitchcock Problem," *Can. J. Math.*, 9, 210-218 (1957).
4. L. R. Ford, Jr., and D. R. Fulkerson, "Network Flow and Systems of Representatives," *Can. J. Math.*, 10, 78-85 (1958).
5. D. R. Fulkerson, "A Network Flow Feasibility Theorem and Combinatorial Applications," *Can. J. Math.*, 11, 440-451 (1959).
6. D. R. Fulkerson, "Zero-One Matrices with Zero Trace," to appear in *Pac. J. Math.*
7. D. R. Fulkerson, and H. J. Ryser, "Widths and Heights of (0, 1)-Matrices," to appear in *Can. J. Math.*
8. David Gale, "A Theorem on Flows in Networks," *Pac. J. Math.*, 7, 1073-1082 (1957).
9. R. M. Haber, "Term Rank of 0, 1 Matrices," *Rend. Sem. Mat. Padova*, 30, 24-51 (1960).
10. P. Hall, "On Representatives of Subsets," *J. Lond. Math. Soc.*, 10, 26-30 (1935).
11. P. J. Higgins, "Disjoint Transversals of Subsets," *Can. J. Math.*, 11, 280-285 (1959).
12. Alan J. Hoffman, "Some Recent Applications of the Theory of Linear Inequalities to Extremal Combinatorial Analysis," *Proceedings of Symposia in Applied Mathematics* 10, 113 (1960), American Mathematical Society.
13. Dénes König, *Theorie der endlichen und unendlichen Graphen*, Chelsea Publishing Company, New York, 1950.
14. H. W. Kuhn, "The Hungarian Method for the Assignment Problem," *Naval Res. Logist. Quart.*, 2, 83-97 (1955).
15. H. W. Kuhn, and A. W. Tucker (ed.), "Linear Inequalities and Related Systems," *Annals of Math. Study* 38 (1956).
16. H. B. Mann, and H. J. Ryser, "Systems of Distinct Representatives," *Amer. Math. Monthly*, 60, 397-401 (1953).
17. N. S. Mendelsohn, and A. L. Dulmage, "Some Generalizations of the Problem of Distinct Representatives," *Can. J. Math.*, 10, 230-241 (1958).
18. Oystein Ore, "Graphs and Matching Theorems," *Duke Math. J.*, 22, 625-639 (1955).
19. H. J. Ryser, "Combinatorial Properties of Matrices of Zeros and Ones," *Can. J. Math.*, 9, 371-377 (1957).
20. H. J. Ryser, "The Term Rank of a Matrix," *Can. J. Math.*, 10, 57-65 (1958).
21. H. J. Ryser, "Traces of Matrices of Zeros and Ones," to appear in *Can. J. Math.*

Received February 29, 1960