TRACES OF A CLASS OF (0, 1)-MATRICES

DALE M. MESNER

1. Introduction. If $A = (a_{ij})$ is a matrix whose elements are 0's and 1's, more briefly a (0, 1)-matrix, the *trace* of A, defined as the sum of the diagonal elements a_{ii} and denoted by Tr A, is the number of 1's on the main diagonal. Matrices which can be obtained from A by permutation of its rows or of its columns, or both, may be expressed as PA, AQ, or PAQ, respectively, where P and Q are permutation matrices.

DEFINITION. *If* Tr $PAQ = t$ *for some permutation matrices* P *and* Q, *we say that trace* t *is possible for* A.

This paper discusses the set of trace values that are possible for a given (0, 1)-matrix A, i.e. the set of distinct traces of the class of matrices $\{PAQ\}$, where P and Q are arbitrary permutation matrices of appropriate orders. The largest and smallest possible trace values will be denoted by $t_{\max}(A)$ and $t_{\min}(A)$, or more briefly by $t_{\max}$ and $t_{\min}$. The main problem, solved in Section 2, is to determine possible intermediate values for Tr A, though $t_{\max}$ and $t_{\min}$ themselves are evaluated for some special matrices. Theorem 3 summarizes the principal results. We conclude with some discussion in Section 3.

Terminology and notation. J and Z will denote matrices all of whose elements are 1's and 0's respectively. The *complement* of a (0, 1)-matrix A is the matrix $J - A$. The *direct sum* of the square or rectangular matrices A_1, A_2, denoted by $A_1 \dotplus A_2$, is the matrix

$$\begin{bmatrix} A_1 & Z \\ Z & A_2 \end{bmatrix},$$

where the Z's are of appropriate orders. $A_1 \dotplus \ldots \dotplus A_k$ is similarly defined. $P(A_1 \dotplus \ldots \dotplus A_k)Q$ will be called a *rearranged direct sum* of $A_1, \ldots, A_k$. If A is an $m \times n$ matrix, $m \leqslant n$, the matrix obtained by permuting rows and carrying out the same permutation on the first m columns will be called a *simultaneous row and column permutation* of A and may be expressed as $PA(P^T \dotplus I)$, where P^T is the transpose of P and I is the identity matrix of order $n - m$, which does not occur if $m = n$. The *term rank* of a matrix A may be defined **(2, 4)** as the order of the greatest minor of A with a non-zero term in the expansion of its determinant. This integer is also equal to the minimal number of rows and columns which collectively contain all the non-zero elements of A **(1)**.

Received November 30, 1962.

https://doi.org/10.4153/CJM-1964-008-8 Published online by Cambridge University Press

2. Possible trace values. Let A be an $m \times n$ matrix of 0's and 1's. Since the same trace values are possible for A^T as for A, we assume without loss of generality that $m \leqslant n$. Simultaneous row and column permutation leaves the set of diagonal elements, and hence the trace, invariant and will be used to simplify matrices whenever convenient. We shall use the facts that $\mathrm{Tr}(J - A) = m - \mathrm{Tr}\, A$ and that trace t is possible for A if and only if trace $m - t$ is possible for $J - A$. It is also useful to note that

(i) $t_{\max}(A)$ is identical with the term rank of A;

(ii) $t_{\min}(A) = m - t_{\max}(J - A)$.

In general, not all trace values between $t_{\max}$ and $t_{\min}$ are possible. In the case of a permutation matrix the trace may be interpreted as the number of objects left fixed by the permutation. Since a permutation of m objects cannot leave $m - 1$ of them fixed without fixing the remaining one as well, trace $m - 1$ is not possible for a permutation matrix of order m, and in particular not for the identity matrix of order m. In the next lemma a similar result is proved for a direct sum of matrices of 1's, of which the identity matrix is a special case. Such direct sums play an exceptional role in this study.

LEMMA 1. *Let B be a $t \times t$ matrix which is a direct sum of square matrices of* 1*'s. Then trace $t - 1$ is not possible for B.*

Proof. Assume that some row and column permutation of B is a matrix $A = (a_{ij})$ with trace $t - 1$. We may suppose that $a_{11} = 0$ and that the other diagonal elements of A are 1's. Since each row of B contains 1's, row 1 of A contains 1's in exactly k columns for some integer $k \geqslant 1$. These 1's must be the elements of one row of a square matrix of 1's in the direct sum; hence there must be exactly $k - 1$ other rows of A which contain 1's in these columns. But, by assumption, the diagonal elements contained in these columns are 1's; hence there must be at least k other rows which contain 1's in these columns. This is a contradiction, proving the lemma.

LEMMA 2. *Let B be a $t \times t$ matrix of* 0*'s and* 1*'s with trace t, and let trace $t - 1$ be impossible for B. Then some simultaneous row and column permutation of B is a direct sum of square matrices of* 1*'s.*

Proof. Neither B nor any matrix obtainable from B by simultaneous row and column permutation can contain a principal minor of either of the forms

$$(a)\ \begin{matrix} 1 & 0 \\ 1 & 1 \end{matrix}, \qquad (b)\ \begin{matrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{matrix}$$

for if it did, a cyclic permutation of the rows involved would reduce the trace by unity. The absence of (a) means that B is symmetric. The absence of (b) has the following interpretation for $1 \leqslant i, j$ and $k \leqslant t$. If $b_{ij} = 1$ and $b_{jk} = 1$, then $b_{ik} = 1$. Regarding "$b_{ij} = 1$" as a binary relation between i and j, the

https://doi.org/10.4153/CJM-1964-008-8 Published online by Cambridge University Press

two previous statements mean that the relation is symmetric and transitive. Since $b_{ii} = 1$, it is also reflexive. It is therefore an equivalence relation, dividing the set of indices $1, 2, \ldots, t$ into equivalence classes, with $b_{ij} = 1$ for i and j in the same class and $b_{ij} = 0$ otherwise. This implies the conclusion of the lemma.

COROLLARY. *If B is a square matrix of* 0'*s and* 1'*s with trace* 0, *and trace* 1 *is impossible for B, then some simultaneous row and column permutation of B is the complement of a direct sum of square matrices of* 1'*s.*

In Lemma 2 and Corollary we note in particular that matrix B is symmetric.

THEOREM 1. *Let A be an $m \times n$ matrix of* 0'*s and* 1'*s, $m \leqslant n$, such that neither A nor its complement is a rearranged direct sum of matrices of* 1'*s. Then for any integer t satisfying $t_{\min} \leqslant t \leqslant t_{\max}$, trace t is possible for A.*

Proof. Trace $t_{\min}$ is possible by definition. The proof will be by induction on t, using the induction hypothesis that if $t_{\min} \leqslant \operatorname{Tr} A = t < t_{\max}$, then trace $t + 1$ is possible. The induction hypothesis will be proved by adopting the assumption

$$(*) \quad \begin{cases} \textit{for some } t,\ t_{\min} \leqslant t \leqslant t_{\max}, \\ \operatorname{Tr} A = t, \textit{ but trace } t+1 \textit{ is impossible for } A, \end{cases}$$

analyzing the structure of A, and eventually reaching the conclusion $t = t_{\max}$.

If $\operatorname{Tr} A = t$, then after simultaneous row and column permutation A has the form

$$A = \begin{bmatrix} B & C & D \\ E & F & G \end{bmatrix}, \tag{1}$$

where B is an $(m - t) \times (m - t)$ matrix with trace zero and F is a $t \times t$ matrix with trace t. D and G have $n - m$ columns and do not occur if $m = n$.

Under assumption (*), D must be a matrix of 0's, for a 1 in D could replace a diagonal 0 of B under an exchange of the columns containing them, increasing the trace by unity. Under assumption (*), the corollary of Lemma 2 shows that B may be assumed to be the complement of a direct sum of square matrices of 1's. Two cases will be treated separately, according to the number of summands in the direct sum.

Case I. B is a matrix of 0's. In this case, assuming (*), a procedure will be described for selecting t rows and columns which collectively contain all the 1's of A. This will show that A has term rank at most t, which is sufficient to show that $t = t_{\max}$.

The row and column indices will be partitioned into disjoint sets $\mathfrak{S}_0, \mathfrak{S}_1, \ldots, \mathfrak{S}_r$. A_{ij} will denote the submatrix with row indices in $\mathfrak{S}_i$ and column indices in $\mathfrak{S}_j$. A_{00} is defined as the submatrix BD, determining $\mathfrak{S}_0$ as $\{1, \ldots, m - t, m + 1, \ldots, n\}$ and leaving t indices in the remaining sets. $\mathfrak{S}_1, \mathfrak{S}_2, \ldots$ are determined successively by rules to be stated in the next paragraph and illustrated by an example which follows. The motivation for the rules is that the ith selected set of rows (columns) disposes of as many as possible of the 1's in a previously unselected (typically the $(i-2)$nd) set of columns (rows). The rows with indices in $\mathfrak{S}_1, \ldots, \mathfrak{S}_{2j+1}, \ldots$ and the columns with indices in $\mathfrak{S}_2, \ldots, \mathfrak{S}_{2i}, \ldots$ will be the t rows and columns selected. The only elements of A not in selected rows or columns are those in the submatrices

$$(2) \qquad A_{2i,0} \text{ and } A_{2i,2j+1}, \qquad i, j \geqslant 0,$$

which will be shown to be matrices of 0's.

We choose $\mathfrak{S}_1$ so that each row of A_{10} contains (one or more) 1's, and each A_{i0}, $i \geqslant 2$, is a matrix of 0's. In particular, the latter holds for the submatrices $A_{2i,0}$. We choose $\mathfrak{S}_{2i+2}$, $i \geqslant 0$, so that each column of $A_{2i,2i+2}$ contains 1's and each $A_{2i,k}$, $k \geqslant 2i+3$, is a matrix of 0's. In particular, the latter holds for each $A_{2i,2j+1}$ with $2j+1 \geqslant 2i+3$. We choose $\mathfrak{S}_{2j+3}$, $j \geqslant 0$, so that each row of $A_{2j+3,2j+1}$ contains 1's and each $A_{k,2j+1}$, $k \geqslant 2j+4$, is a matrix of 0's. In particular, the latter holds for each $A_{2i,2j+1}$ with $2j+1 \leqslant 2i-3$. Thus, by construction of the sets $\mathfrak{S}_i$, submatrices (2) are matrices of 0's with the exception of submatrices of the form

$$(3) \qquad A_{2i,2i\pm 1},$$

which have not yet been discussed.

We continue this procedure until we arrive at an empty set $\mathfrak{S}_\mu$. We then take $r = \mu + 1$ and choose $\mathfrak{S}_r$ as the set of all remaining indices. If it is non-empty, $\mathfrak{S}_r$ determines a submatrix with indices r and $r-2$ in some order which may violate the provisions of the preceding paragraph.

The situation after the choice of sets $\mathfrak{S}_0, \ldots, \mathfrak{S}_r$ is illustrated by the following schematic diagram and numerical example with $r = 6$. We assume simultaneous row and column permutation so that each $\mathfrak{S}_i$ contains consecutive indices, with the exception of column indices $> m$ in $\mathfrak{S}_0$.

$$(4) \qquad A = \left[\begin{array}{c|c|c} B & C & D \\ \hline E & F & G \end{array}\right]$$

$$= \left[\begin{array}{c|c:c:c:c:c|c} A_{00} & A_{01} & A_{02} & & Z_2 & & A_{00} \\ \hline A_{10} & \cdot & \cdot & \cdot & \cdot & \cdot & A_{10} \\ \hdashline & A_{21} & \cdot & A_{23} & A_{24} & Z_4 & \\ \cline{2-2} Z_1 & A_{31} & \cdot & \cdot & \cdot & \cdot & Z_1 \\ \cline{2-2} & Z_3 & \cdot & A_{43} & \cdot & A_{46} & \\ & & \cdot & Z_5 & \cdot & \cdot & \end{array}\right] \begin{array}{l} \\ \}\mathfrak{S}_1 \\ \\ \}\mathfrak{S}_3 \\ \\ \\ \end{array}$$

$$\begin{array}{ccc} \underbrace{\qquad}_{\mathfrak{S}_2} & \qquad & \underbrace{\qquad}_{\mathfrak{S}_4}\underbrace{\qquad}_{\mathfrak{S}_6} \end{array}$$

$$
= \begin{array}{c}
\left[\begin{array}{cc|c|cc|cccc|c}
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & (1) & 0 & 0 & 0 & 0 & 0 \\
\hline
(1) & 0 & 1 & . & . & . & . & . & . & 1 \\
\hline
0 & 0 & 0 & 1 & . & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & . & 1 & 0 & (w) & 0 & 0 & 0 \\
0 & 0 & 1 & . & . & 1 & . & . & . & 0 \\
0 & 0 & (1) & . & . & . & 1 & . & . & 0 \\
0 & 0 & 0 & . & . & 0 & 0 & 1 & . & 0 \\
0 & 0 & 0 & . & . & 0 & 0 & . & 1 & 0
\end{array}\right]
\begin{array}{l}
\\ \\ \}\mathfrak{S}_1 \\ \\ \\ \left.\begin{array}{l} \\ \\ \end{array}\right\} \mathfrak{S}_3 \\ \\ \\
\end{array}
\\
\qquad\qquad \underbrace{\qquad\quad}_{\mathfrak{S}_2} \qquad\qquad \underbrace{\quad}_{\mathfrak{S}_4}\underbrace{\quad}_{\mathfrak{S}_6}
\end{array}
$$

Set $\mathfrak{S}_1$ (resp. $\mathfrak{S}_2$, $\mathfrak{S}_3$, $\mathfrak{S}_4$) is chosen so that there are 1's in each row of A_{10} (resp. column of A_{02}, row of A_{31}, column of A_{24}), but no 1's in Z_1 (resp. Z_2, Z_3, Z_4). The absence of 1's in Z_5 determines $\mathfrak{S}_5$ as the empty set. A_{46} may violate the condition on 1's in each column. Labels in the margins indicate the selected sets of rows and columns. Elements omitted from the numerical example are arbitrary, as are submatrices omitted from the schematic diagram (except that the diagonal elements of F are 1's). The numerical example contains an illustration of an argument which follows; if we take $h_{2p,2q+1} = h_{23} = w$, the elements in parentheses are suitable choices for the set (6) and the argument shows that $w = 0$.

Assumption (*) will be used to show that submatrices (3) are matrices of 0's. Let $A_{2p,2q+1}$ be any submatrix with $2q + 1 = 2p \pm 1$, and let an arbitrary element of this submatrix have row index $u_{2p} \in \mathfrak{S}_{2p}$ and column index $u_{2q+1} \in \mathfrak{S}_{2q+1}$. (Since $\mathfrak{S}_{r-1}$ is empty, submatrices with index $r - 1$ do not occur and we may assume that $2p$ and $2q + 1$ are $\leqslant r - 2$. As a result, the exceptional submatrix with indices r and $r - 2$ will not enter into the proof.) Additional indices $u_i \in \mathfrak{S}_i$ will next be determined for distinct $i < 2p$, $2q + 1$, and the matrix element $a_{u_i u_j}$ will be denoted by h_{ij}.

For $\sigma = p - 1, \ldots, 0$, successively determine $u_{2\sigma}$ so that $h_{2\sigma,2\sigma+2} = 1$; this is possible because each column of $A_{2i,2i+2}$ contains 1's. For $\tau = q - 1, \ldots, 0$, successively determine $u_{2\tau+1}$ so that $h_{2\tau+3,2\tau+1} = 1$; this is possible because each row of $A_{2j+3,2j+1}$ contains 1's. We also need to specify h_{10}; this we do by choosing a column index $v_0 \in \mathfrak{S}_0$ so that $a_{u_1 v_0} = 1$ (possible because each row of A_{10} contains 1's), then redefining $h_{10} = a_{u_1 v_0} = 1$ and $h_{00} = a_{u_0 v_0} = 0$. If $u_0 \neq v_0$ an interchange of columns u_0 and v_0, which in Case I will not alter $\operatorname{Tr} A$, is then used to place h_{00} on the main diagonal.

Let H be the principal minor of order $p + q + 2$ with diagonal elements

(5) $\quad h_{2\sigma,2\sigma}, h_{2\tau+1,2\tau+1}, h_{2p,2p}, h_{2q+1,2q+1}, \qquad \sigma = 0, \ldots, p - 1, \tau = 0, \ldots, q - 1.$

H has trace $p + q + 1$, since $h_{00} = 0$, while the other elements (5) are on the main diagonal of F and are hence 1's. The $p + q + 2$ elements

(6) $h_{10}, h_{2\sigma,2\sigma+2}, h_{2\tau+3,2\tau+1}, h_{2p,2q+1}, \qquad \sigma = 0, \ldots, p-1, \tau = 0, \ldots, q-1,$

fall in distinct rows and columns of H and may be placed on its main diagonal by a permutation of its columns. Each of elements (6) except $h_{2p,2q+1}$ was chosen equal to 1; we may therefore conclude that $h_{2p,2q+1} = 0$, for otherwise trace $p+q+2$ is possible for H, which violates assumption (*). But $h_{2p,2q+1} = a_{u_{2p}u_{2q+1}}$ is an arbitrary element of $A_{2p,2q+1}$; thus each $A_{2i,2i\pm1}$ is a matrix of 0's. This completes the proof that under assumption (*) and in Case I the t selected rows and columns of A contain all its 1's. It follows that A has term rank at most t and that $t = t_{\max}$. In Case I, if Tr $A = t$ for $t_{\min} \leqslant t < t_{\max}$, the negation of assumption (*) must hold, namely that trace $t+1$ is possible. But this is the induction hypothesis we need and the conclusion of the theorem holds in Case I.

Case II. B as defined in (1) is the complement of a direct sum of at least two square matrices of 1's. It has been shown that D is a matrix of 0's; thus any column of D can be exchanged with a column of B without altering Tr A. In Case II this permutation will destroy the symmetry of B and the corollary of Lemma 2 implies that trace 1 is possible for B. Then trace $t+1$ is possible for A and assumption (*) is violated. Therefore in Case II no columns of D can exist, and A is a square matrix.

We have now proved that the theorem holds for a non-square matrix by the proof under Case I. Moreover, we have not yet applied the restriction on direct sum form of A, and we remark for later reference that if A is not square the theorem holds without this hypothesis.

Under this hypothesis we may assume that F has at least one row and column, since otherwise

$$A = \begin{bmatrix} B & C \\ E & F \end{bmatrix}$$

reduces to B and its complement has the forbidden direct sum form. In Case II, trace 2 is possible for B and trace $t-1$ is accordingly not possible for F under assumption (*). Lemma 2 then shows that F can be assumed to be a direct sum of square matrices of 1's. In particular, F must be symmetric.

Under assumption (*) an element of C and the symmetrically placed element of E cannot both be 1's, for if they were, an exchange of the columns containing them would increase the trace by unity.

In Case II we may partition A as follows:

$$A = \left[\begin{array}{c|c} B & C \\ \hline E & F \end{array}\right] = \left[\begin{array}{cc|c} B_1 & J_1 & C_1 \\ J_2 & B_2 & C_2 \\ \hline E_1 & E_2 & F \end{array}\right], \tag{7}$$

where B_1 and B_2 are square matrices and J_1 and J_2 are matrices of 1's. Let c_1 and c_2 be elements of C_1 and C_2 which are in the same column but are otherwise arbitrary, and let e_1 and e_2 be the symmetrically located elements of E_1 and E_2. Then the principal minor

$$\begin{matrix} 0 & 1 & c_1 \\ 1 & 0 & c_2 \\ e_1 & e_2 & 1 \end{matrix}$$

has trace 1; under assumption (*), trace 2 must be impossible for this minor, implying $c_1 + e_2 \neq 1$ and $c_2 + e_1 \neq 1$. These inequalities in elements equal to 0 or 1 imply $c_1 = e_2$ and $c_2 = e_1$. Therefore each element of any column of C_1 or C_2 is equal to every element of the symmetrically located row of E_2 or E_1 respectively.

If the complement of B is a direct sum of three or more matrices of 1's, partition (7) can be carried out in more than one way, and the conclusion is that all elements in the entire column of C and row of E have the same value, necessarily 0 because of the restriction against symmetrically located 1's. Then A has a principal minor of the form

$$\begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}.$$

But this is a matrix of trace 1 for which trace 2 is possible, leading to a violation of assumption (*). Therefore in Case II, B is the complement of a direct sum of exactly two matrices of 1's. In the partition (7) B_1 and B_2 are matrices of 0's, say of order β_1 and β_2 respectively.

In any column of C and the symmetrically located row of E, the β_1 elements in C_1 and the β_2 elements in E_2 have a common value x, while the β_2 elements in C_2 and the β_1 elements in E_1 have a common value y. The case $x = y = 1$ is impossible because of the restriction against symmetrically located 1's in C and E, but the other combinations of values of x and y correspond to three possible forms for columns of C and rows of E. After simultaneous row and column permutation, A has the form

$$A = \left[\begin{array}{c|c} B & C \\ \hline E & F \end{array}\right] = \left[\begin{array}{cc|ccc} B_1 & J_1 & J_3 & Z_1 & Z_2 \\ J_2 & B_2 & Z_3 & J_4 & Z_4 \\ \hline Z_5 & J_5 & F_{11} & F_{12} & F_{13} \\ J_6 & Z_6 & F_{21} & F_{22} & F_{23} \\ Z_7 & Z_8 & F_{31} & F_{32} & F_{33} \end{array}\right], \tag{8}$$

where J's and Z's represent matrices of 1's and of 0's. F_{11} (say of order t_1), F_{22} (say of order t_2), and F_{33} are principal minors of F with 1's on the main diagonal, and F is still symmetric. Now consider a column permutation which exchanges any column of B_1 or B_2 with any column of Z_1 or Z_3 respectively, or a row permutation which exchanges any row of B_1 or B_2 with any row of Z_5 or Z_6 respectively. Since these changes do not modify diagonal elements of A or the form of B, they leave A in the form of Case II, from which it follows that they must not destroy the symmetry of F. This shows that F_{11} and F_{22} are matrices of 1's and that F_{ij}, $i \neq j$, are matrices of 0's. F_{33} may be assumed to be a direct sum of square matrices of 1's.

Submatrices

$$\begin{bmatrix} J_1 & J_3 \\ J_5 & F_{11} \end{bmatrix} \tag{9}$$

of order $(\beta_1 + t_1) \times (\beta_2 + t_1)$ and

$$\begin{bmatrix} J_2 & J_4 \\ J_6 & F_{22} \end{bmatrix} \tag{10}$$

of order $(\beta_2 + t_2) \times (\beta_1 + t_2)$ are matrices of 1's, and A is a rearranged direct sum of these matrices and of matrices of 1's occurring in F_{33}; this violates a hypothesis of the theorem and shows that Case II is impossible under assumption (*). Thus assumption (*) has led to a contradiction. We conclude in Case II as in Case I that the induction hypothesis holds, and the proof of Theorem I is complete.

We turn now to the cases in which A is a rearranged direct sum of matrices of 1's, or the complement of such a matrix. Since the possible trace values for the complement $J - A$ are easily deduced from those for a matrix A, we may confine ourselves to the first of the two cases. Portions of the proof of Theorem I are applicable and will be used in Lemma 4 and the subsequent paragraph. The special case in which A is a rearranged direct sum of exactly two matrices of 1's is exceptional in that $J - A$ then has the same form, and will not be covered fully until Lemma 5.

LEMMA 4. *If A is an $m \times m$ matrix of* 0'*s and* 1'*s and statements* 1–4 *are true, then statements* 5 *and* 6 *are true.*

1. *Assumption* (*): $\operatorname{Tr} A = t$ *but trace* $t + 1$ *is impossible.*
2. *A has the form of Case* II.
3. *$J - A$ is not a direct sum of matrices of* 1'*s.*
4. *A is a rearranged direct sum of matrices of* 1'*s.*
5. *If $t = m - 2$, each direct summand of A is square.*
6. *If $t \leqslant m - 3$, A has exactly two direct summands.*

Proof. In the proof of Case II, statements 1–3 were shown to imply statement 4. It will next be shown that they imply statements 5 and 6 as well. A fortiori, statements 1–4 will imply statements 5 and 6. If $t = m - 2$, then $\beta_1 = \beta_2 = 1$ and submatrices (9) and (10) are square. The direct summands in F_{33} are also square and statement 5 is proved. If $t \leqslant m - 3$ and F_{33} has at least one row and column, then A may be assumed to have a principal minor of the form

$$\begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{matrix},$$

https://doi.org/10.4153/CJM-1964-008-8 Published online by Cambridge University Press

which has trace 1 and for which trace 2 is possible, leading to a violation of (*). Therefore if $t \leqslant m - 3$, submatrix F_{33} does not occur and A reduces to a rearranged direct sum of submatrices (9) and (10), proving statement 6.

For A a rearranged direct sum of matrices of 1's, we next determine those values of t for which assumption (*) is true, since the induction proof of Theorem 1 is valid for trace values between $t_{\min}$ and the smallest such t. The conclusion of that theorem holds if and only if there is no such $t < t_{\max}$; from the discussion following Case I, it holds if A is not square and in Case I if A is square. Thus we may restrict attention to a square matrix A which has the form of Case II. At this point we have adopted all of statements 1–4 of Lemma 4 as hypotheses and can conclude that A satisfies statements 5 and 6. Assumption (*) for A is true with $t \leqslant m - 3$ only in the special case of a rearranged direct sum of exactly two matrices of 1's. If A is a rearranged direct sum of three or more matrices of 1's, then $t = m - 2$ is the only trace value $< t_{\max}$ such that (*) is true for A, and then only if all direct summands are square. In this case $t = m - 1$ is the only trace value between $t_{\min}$ and $t_{\max}$ which is impossible for A; that it actually is impossible follows from Lemma 1. If A is a rearranged direct sum of three or more matrices of 1's which are not all square, then (*) is not true for A for any $t < t_{\max}$ and the conclusion of Theorem 1 holds. These results for A of this form, along with the corresponding statements for the complement, will be included in Theorem 3 after the remaining special case is disposed of in Lemma 5.

THEOREM 2. *If* $A = J_1 \dot{+} J_2 \dot{+} \ldots$, *where* J_1 *is an* $m_i \times n_i$ *matrix of* 1's, *and* $\sum_i m_i = m \leqslant \sum_i n_i = n$, *then*

$$t_{\max}(A) = \sum_i \min(m_i, n_i), \tag{11}$$

$$t_{\min}(A) = \max\Big(0, -n + \max_i (m_i + n_i)\Big). \tag{12}$$

Proof. $t_{\max}(A)$ is equal to the term rank of A, and (11) is merely an expression for the minimal number of rows and columns which contain all the 1's of A.

$m - t_{\min}(A)$ is equal to the term rank of the complement of A, or to the minimal number of rows and columns which contain all the 0's of A. Let a set of s rows and columns contain all the 0's of A. Then $t_{\min}(A) = m - \min(s)$. If the set consists of all rows of A, then $s = m$. If at least one row is missing from the set, say a row of J_i, then s can be reduced by omitting all m_i rows of J_i, but all columns not in J_i must be included and $s = m - m_i + n - n_i$. s cannot be further reduced by omitting rows of more than one J_i, for then all columns of A must be included and $s \geqslant n \geqslant m$. Therefore

$$\min(s) = \min\Big(m, \min_i (m - m_i + n - n_i)\Big),$$

leading to the expression in (12) for $t_{\min}(A)$.

COROLLARY. $t_{\max}(A) = m$ *if and only if* $m_i \leqslant n_i$, $i = 1, 2, \ldots$. *If* $m = n$, *then* $t_{\max}(A) = m$ *if and only if* $m_i = n_i$, $i = 1, 2, \ldots$. $t_{\min}(A) = 0$ *if and only if*

$$\max_i (m_i + n_i) \leqslant n.$$

LEMMA 5. *If a square matrix* A *is a direct sum of two matrices of* 1'*s, then trace* t *is possible for* A *if and only if*

$$t_{\min} \leqslant t \leqslant t_{\max} \quad \textit{and} \quad t \equiv t_{\max} \pmod 2.$$

Proof. The $r \times s$ matrix of 1's will be denoted by $J_{r\times s}$. Let $A = J_{m_1\times n_1} \dotplus J_{m_2\times n_2}$, where $m_1 + m_2 = n_1 + n_2 = m$. Also choose notation, and transpose A if necessary, so that

$$(13) \qquad m_1 + n_1 \geqslant m_2 + n_2, \qquad m_1 \leqslant n_1.$$

It follows from Theorem 2 that $t_{\max} = m_1 + n_2$ and $t_{\min} = m_1 - n_2$. Then trace zero is possible for A if and only if

$$(14) \qquad m_1 = n_2.$$

Note that (14) is equivalent to $n_1 = m_2$ and, by symmetry, is independent of inequalities (13).

Consider a row and column permutation of A which places on the main diagonal exactly t_1 elements of $J_{m_1\times n_1}$ and exactly t_2 elements of $J_{m_2\times n_2}$. Such a permutation is possible if and only if when the $t_1 + t_2$ rows and $t_1 + t_2$ columns of A containing these elements are deleted, the remaining submatrix is one for which trace zero is possible. But this submatrix is

$$J_{(m_1-t_1)\times(n_1-t_1)} \dotplus J_{(m_2-t_2)\times(n_2-t_2)},$$

and from (14) the condition for trace zero is

$$(15) \qquad m_1 - t_1 = n_2 - t_2.$$

That is, trace t is possible for A if and only if there are integers t_1, t_2 satisfying (15) and

$$\begin{aligned} &0 \leqslant t_1 \leqslant \min(m_1, n_1) = m_1,\\ &0 \leqslant t_2 \leqslant \min(m_2, n_2) = n_2,\\ &t_1 + t_2 = t. \end{aligned}$$

These conditions reduce to $t = m_1 - n_2 + 2t_2$, $0 \leqslant t_2 \leqslant n_2$, holding for precisely the values of t stated in the lemma.

The following theorem is a collection of results that have been proved in Theorem 1, Lemmas 4 and 5, and the accompanying discussion.

THEOREM 3. *Let* A *be an* $m \times n$ *matrix of* 0'*s and* 1'*s*, $m \leqslant n$, *with maximum trace* $t_{\max}$ *and minimum trace* $t_{\min}$. *Then*

(1) *if* A *is a rearranged direct sum of three or more square matrices of* 1'*s, trace* $m - 1$ *is impossible*; $t_{\min} < m - 1 < t_{\max} = m$;

(2) *if the complement of A is a rearranged direct sum of three or more square matrices of* 1's, *trace* 1 *is impossible*;

$$0 = t_{\min} < 1 < t_{\max};$$

(3) *if A is square and is a rearranged direct sum of two matrices of* 1*'s, trace t is impossible if $t_{\max} - t \equiv 1 \pmod 2$*;

$$0 < t_{\max} - t_{\min} \equiv 0 \pmod 2;$$

(4) *in every other case, if $t_{\min} \leqslant t \leqslant t_{\max}$, trace t is possible.*

3. Discussion. If trace t is possible for A, then it is possible under row permutation alone or column permutation alone. To see this in the case $m \leqslant n$, let PAQ have trace t and multiply on the left and right by P^T and by $P \dotplus I$, a simultaneous row and column permutation which preserves the trace. But $P^T PAQ(P \dotplus I) = AR$, where $R = Q(P \dotplus I)$ is a permutation matrix. If $m = n$, then also Tr $RA = t$.

Theorems 1 and 3 lead to convenient sufficient conditions for arbitrary trace, $0 \leqslant t \leqslant m$, to be possible for an $m \times n$ (0, 1)-matrix, $m \leqslant n$. For instance, it is sufficient for each of A and $J - A$ to have rank m and be different from the $m \times m$ identity matrix.

As a slight generalization of the idea of a direct sum of matrices, we may consider a $0 \times k$ summand with no rows or a $k \times 0$ summand with no columns as contributing k columns or rows respectively, of 0's to the direct sum matrix. We remark that Theorem 2 with a slightly modified proof is still valid if such summands are admitted.

A necessary and sufficient condition for trace t to be possible for the matrix of Lemma 5 was shown to be the existence of a partition $t = t_1 + t_2$ such that t_i rows and t_i columns can be deleted from the ith direct summand, $i = 1, 2$, to leave a matrix for which trace zero is possible. Although we did not need it, there is a natural generalization of this condition to the direct sum of any number of matrices of 1's.

Since the trace of a (0, 1)-matrix can be regarded as an enumeration of the 1's on a particular diagonal, our results have an application to any rectangular array of objects some of which are distinguished by a specified attribute. Given the location in the array of the specified objects, our theorems indicate how many of them can be placed on a particular diagonal by permutation of rows and columns of the array.

Ryser **(5)** determines the set of distinct trace values of the class of all (0, 1)-matrices having a specified ordered set of row totals and set of column totals. We have made the same determination for the classes $\{PA\}$, $\{AQ\}$, $\{PAQ\}$ of matrices which can be obtained from a given matrix A by row or column permutation or both. In the special case of an $m \times n$ matrix, $m \leqslant n$, with equal column totals, $\{AQ\}$ is a subclass of Ryser's class (similarly $\{PA\}$ if $m \geqslant n$ with equal row totals) and the present results are a little stronger. In general Ryser's class neither includes nor is included by the present ones and the results are independent.

Row and column permutation preserve some algebraic properties of a matrix. If X is a solution of the matrix equation $X^TX = C$ for a given matrix C, then any row permutation gives a matrix $Y = PX$ which is also a solution, and Theorem 3 shows in many instances that if a solution exists in (0, 1)-matrices, it exists with arbitrary trace. An important example is a v, k, λ matrix, defined as a (0, 1)-matrix N of order v satisfying the equation

$$NN^T = N^TN = \lambda J + (k - \lambda)I. \tag{16}$$

Necessary conditions are known for the set of integers v, k, λ, and solutions are known for infinitely many sets, but necessary and sufficient conditions for the existence of a solution are not yet known. The matrices I, J, Z, and $J - I$ are trivial solutions; for other solutions it is easy to show that N and $J - N$ are non-singular. It follows that if a solution exists, it exists with arbitrary trace $0 \leqslant t \leqslant v$.

The trace of a square matrix is equal to the sum of its characteristic roots, while the determinant is equal to their product, except possibly for sign; both trace and determinant occur as coefficients in the characteristic equation of the matrix. Row and column permutation may modify the characteristic roots, but in such a way that their product is invariant up to sign. For a (0, 1)-matrix, Theorem 3 casts some light on the nature of the modification of the roots by indicating possible values of their sum. In the case of a v, k, λ matrix the roots are known **(3)** to be k, $\rho_i \sqrt{(k - \lambda)}$, $i = 1, \ldots, v - 1$, where each ρ_i is of absolute value unity. Row and column permutation results in a matrix which still satisfies (16) and whose characteristic roots still have the special form just mentioned. It is natural to ask whether this regularity is reflected in restrictions on the trace of a v, k, λ matrix; in fact, it was this question in a conversation between A. J. Goldman and the author which led to this investigation. The negative answer seems somewhat surprising.

References

1. Denes König, *Theorie der endlichen und unendlichen Graphen* (New York, 1950).

2. Oystein Ore, *Graphs and matching theorems*, Duke Math. J., *22* (1955), 625–639.

3. H. J. Ryser, *Geometries and incidence matrices*, Slaught Memorial Papers No. 4, Math. Assoc. Amer. (1955), pp. 25–31.

4. ——— *The term rank of a matrix*, Can. J. Math., *10* (1958), 57–65.

5. ——— *Traces of matrices of zeros and ones*, Can. J. Math., *12* (1960), 463–476.

Purdue University,
Lafayette, Indiana