Design Theory and Some Non-simple Forbidden Configurations

R.P. Anstee* and Farzin Barekat†
Mathematics Department
The University of British Columbia
Vancouver, B.C. Canada V6T 1Z2
anstee@math.ubc.ca

September 18, 2019

Abstract

Let $1_k0_l$ denote the $(k+l) \times 1$ column of $k$ 1’s above $l$ 0’s. Let $q \cdot (1_k0_l)$ denote the $(k+l) \times q$ matrix with $q$ copies of the column $1_k0_l$. A 2-design $S_\lambda(2, 3, v)$ can be defined as a $v \times \binom{\lambda}{3}$ $(0,1)$-matrix with all column sums equal 3 and with no submatrix $(\lambda+1) \cdot (1_20_0)$. Consider an $m \times n$ matrix $A$ with all column sums in $\{3, 4, \ldots, m-1\}$. Assume $m$ is sufficiently large (with respect to $\lambda$) and assume that $A$ has no submatrix which is a row permutation of $(\lambda+1) \cdot (1_20_0)$. Then we show the number of columns in $A$ is at most $\lambda^3 \binom{m}{3}$ with equality for $A$ being the columns of column sum 3 corresponding to the triples of a 2-design $S_\lambda(2, 3, m)$. A similar results holds for $(\lambda+1) \cdot (1_20_2)$.

Define a matrix to be simple if it is a $(0,1)$-matrix with no repeated columns. Given two matrices $A$, $F$, we define $A$ to have $F$ as a configuration if and only if some submatrix of $A$ is a row and column permutation of $F$. Given $m$, let $\text{forb}(m, q \cdot (1_k0_l))$ denote the maximum number of possible columns in a simple $m$-rowed matrix which has no configuration $q \cdot (1_k0_l)$. For $m$ sufficiently large with respect to $q$, we compute exact values for $\text{forb}(m, q \cdot (1_10_1))$, $\text{forb}(m, q \cdot (1_20_1))$, $\text{forb}(m, q \cdot (1_20_2))$. In the latter two cases, we use a construction of Dehon (1983) of simple triple systems $S_\lambda(2, 3, v)$ for $\lambda > 1$. Moreover for $l = 1, 2$, simple $m \times \text{forb}(m, q \cdot (1_20_l))$ matrices with no configuration $q \cdot (1_20_l)$ must arise from simple 2-designs $S_\lambda(2, 3, m)$ of appropriate $\lambda$.

The proofs derive a basic upper bound by a pigeonhole argument and then use careful counting and Turán’s bound, for large $m$, to reduce the bound. For

*Research supported in part by NSERC
†Research supported by NSERC of first author
1 Introduction

Some combinatorial objects can be defined by forbidden substructures. It is also true that most combinatorial objects can be encoded by a (0,1)-matrix. In this paper we consider submatrices of (0,1)-matrices as the substructures of interest.

Let \( \mathbf{1}_k \mathbf{0}_l \) denote the \( (k + l) \times 1 \) column consisting of \( k \) 1’s atop \( l \) 0’s. For any positive integer \( q \), let \( q \cdot (\mathbf{1}_k \mathbf{0}_l) \) denote the \( q \times (k + l) \) matrix of \( q \) copies of \( \mathbf{1}_k \mathbf{0}_l \). A 2-design \( S_\lambda(2,3,v) \) consists of \( \lambda \cdot \binom{v}{3} \) triples from \( \{1,2,\ldots,v\} \) such that for each pair \( i,j \in \{1,2,\ldots,v\} \), there are exactly \( \lambda \) triples containing \( i,j \). If we encode the triple system as a \( v \)-rowed (0,1)-matrix \( \mathbf{A} \) such that the columns are the incidence vectors of the triples, then \( \mathbf{A} \) has no submatrix \( (\lambda + 1) \cdot (\mathbf{1}_2 \mathbf{0}_0) \). In fact, if \( \mathbf{A} \) is a \( v \times n \) (0,1)-matrix with column sums 3 and \( \mathbf{A} \) has no submatrix \( (\lambda + 1) \cdot (\mathbf{1}_2 \mathbf{0}_0) \) then \( n \leq \frac{\lambda}{3} \binom{v}{3} \) with equality if and only if the columns of \( \mathbf{A} \) correspond to the triples of a 2-design \( S_\lambda(2,3,v) \). This can be shown by a pigeonhole counting argument.

The problem of forbidding a submatrix is usually extended to forbidding any row and column permutation of the submatrix. Let \( \mathbf{A}, \mathbf{F} \) be (0,1)-matrices. We say that \( \mathbf{A} \) has \( \mathbf{F} \) as a configuration if there is a submatrix of \( \mathbf{A} \) which is a row and column permutation of \( \mathbf{F} \). We extend the forbidden submatrix \( (\lambda + 1) \cdot (\mathbf{1}_2 \mathbf{0}_0) \) and obtain the following two design theory results.

**Theorem 1.1** Let \( \lambda \) and \( v \) be given integers. There exists an \( M \) so that for \( v > M \), if \( \mathbf{A} \) is an \( v \times n \) (0,1)-matrix with column sums in \( \{3,4,\ldots,v-1\} \) and \( \mathbf{A} \) has no configuration \( (\lambda + 1) \cdot (\mathbf{1}_2 \mathbf{0}_1) \) then

\[
n \leq \frac{\lambda}{3} \binom{v}{3}
\]

(1)

and we have equality if and only if the columns of \( \mathbf{A} \) correspond to the triples of a 2-design \( S_\lambda(2,3,v) \). □

When we extend the forbidden configuration to \( (\lambda + 1) \cdot (\mathbf{1}_2 \mathbf{0}_2) \) the case of equality becomes more difficult.

**Theorem 1.2** Let \( \lambda \) and \( v \) be given integers. There exists an \( M \) so that for \( v > M \), if \( \mathbf{A} \) is an \( v \times n \) (0,1)-matrix with column sums in \( \{3,4,\ldots,v-3\} \) and \( \mathbf{A} \) has no configuration \( (\lambda + 1) \cdot (\mathbf{1}_2 \mathbf{0}_2) \) then

\[
n \leq \frac{\lambda}{3} \binom{v}{3}
\]

(2)

and we have equality if and only if there are positive integers \( a,b \) satisfying \( a+b = \lambda \) and there are \( \frac{a}{3} \binom{v}{3} \) columns of \( \mathbf{A} \) of column sum 3 corresponding to the triples of a 2-design \( S_a(2,3,v) \) and there are \( \frac{b}{3} \binom{v}{3} \) columns of \( \mathbf{A} \) of column sum \( v-3 \) of \( v-3 \)-sets whose complements (in \( \{1,2,\ldots,v\} \)) corresponding to the triples of a 2-design \( S_b(2,3,v) \). □
Our first motivation for studying these problems came from extremal set theory. An \( m \times n \) \((0,1)\)-matrix \( A \) can be thought of a multiset of \( n \) subsets of \( \{1,2,\ldots,m\} \). Let \([m] = \{1,2,\ldots,m\}\). For an \( m \times 1 \) \((0,1)\)-column \( \alpha \), we define
\[
S(\alpha) = \{ i \in [m] : \alpha \text{ has } 1 \text{ in row } i \}. \tag{3}
\]

From this we define the natural multiset system \( A \) associated with the matrix \( A \):
\[
A = \{ S(\alpha_i) : \alpha_i \text{ is column } i \text{ of } A \}. \tag{4}
\]

Similarly, if we are given a multiset system \( A \), we can form a matrix \( A \), as long as we don’t care about column order. We define a simple matrix \( A \) as a \((0,1)\)-matrix with no repeated columns. In this case \( A \) yields as set system and it is in this setting that extremal set theory problems can be stated.

We define \( \text{forb}(m, F) \) as the smallest value (depending on \( m \) and \( F \)) so that if \( A \) is a simple \( m \times n \) matrix and \( A \) has no configuration \( F \) then \( n \leq \text{forb}(m, F) \). Alternatively \( \text{forb}(m, F) \) is the smallest value so that if \( A \) is an \( m \times (\text{forb}(m, F) + 1) \) simple matrix then \( A \) must have a configuration \( F \). A sampling of exact results for \( \text{forb}(m, F) \) are in [1], [2].

Let \( K_k \) denote the \( k \times 2^k \) simple matrix of all possible \((0,1)\)-columns on \( k \) rows and let \( K^s_k \) denote the \( k \times \binom{k}{s} \) simple matrix of all possible columns of column sum \( s \).

Many results have been obtained about \( \text{forb}(m, F) \). Exact results have been rare for non-simple configurations \( F \). We consider \( F = q \cdot (1_k0_l) \) for \((k,l) = (1,1), (2,1), (2,2)\).

In [1] we showed that
\[
\begin{aligned}
\left\lfloor \frac{q+1}{2m} \right\rfloor + 2 &\leq \text{forb}(m,q \cdot (1_10_1)) \\
&\leq \left\lfloor \frac{q+1}{2m} + \frac{(q-3)m}{2(m-2)} \right\rfloor + 2
\end{aligned}
\]

where the upper bound obtained by a pigeonhole argument is achieved for \( m = q - 1 \) by taking \( A = [K_0^m K_1^m K_2^m K_m^m - K_m^m] \). For \( m \) with \( m \geq \max\{3q + 2, 8q - 19\} \), we are able to show that the lower bound is correct and slice \( \frac{(q-3)m}{2(m-2)} \approx \frac{q-3}{2} \) off the pigeonhole bound. It is likely that our bound is valid for smaller \( m > q - 1 \). The case \( q = 4 \), is Lemma 3.1 in [2] and took a page to establish.

**Theorem 1.3** Let \( q \geq 3 \) be given. Then for \( m \geq \max\{3q + 2, 8q - 19\} \),
\[
\text{forb}(m,q \cdot (1_10_1)) = \left\lfloor \frac{q+1}{2m} \right\rfloor + 2. \tag{5}
\]

For \( m \) even or \( q - 3 \) even, let \( G \) be a (simple) graph on \( m \) vertices for which all the degrees are \( q - 3 \) and for \( m, q - 3 \) odd let \( G \) be a graph for which \( m - 1 \) vertices have degree \( q - 3 \) and one vertex has degree \( q - 4 \). Such graphs are easy to construct. Let \( H \) be the vertex-edge incidence matrix associated with \( G \), namely for each edge \( e = (i,j) \)
of $G$, we add a column to $H$ with 1’s in rows $i, j$ and 0’s in other rows. Thus $H$ is a simple $m$-rowed matrix with \( \left[ \frac{q-3m}{2} \right] \) columns each of column sum 2. The simple matrix $A = [K^0_m K^1_m H K^{m-1}_m K^m_m]$ has \( \left[ \frac{(q+1)m}{2} \right] + 2 \) columns and no configuration $q \cdot (1_1 0_1)$ which establishes $\text{forb}(m, q \cdot (1_1 0_1)) \geq \left[ \frac{(q+1)m}{2} \right] + 2$. We establish the upper bound in Section 2.

We are able to solve two more cases but need certain designs to achieve exact bounds. A 2-design $S_\lambda(2, 3, v)$ (or triple system) is defined to be simple if no triple is repeated. The associated $v \times \frac{1}{3}(v^2)$ matrix is a simple matrix. We need the following result.

Theorem 1.4 Dehon[3] Let $v, \lambda$ be given. Then a simple 2-design $S_\lambda(2, 3, v)$ exists if and only if $v(v-1) \equiv 0(\text{mod } 6)$, $v-1 \equiv 0(\text{mod } 2)$ and $v \geq \lambda + 2$.

These designs are used in the constructions for the following two theorems in the following way. We form a simple $v \times \frac{1}{3}(v^2)$ matrix $T_{v, \lambda}$ whose columns correspond to the blocks of $S_\lambda(2, 3, v)$ so that if $B$ is a block then the corresponding column has a 1 in row $i$ if and only if $i \in B$. Note that $T_{v, \lambda}$ has no submatrix $(\lambda + 1) \cdot (1_2 0_0)$. Pigeonhole arguments will show that $\text{forb}(m, q \cdot (1_2 0_0)) \leq \binom{m}{0} + \binom{m}{1} + \frac{q+1}{3} \binom{m}{2}$ with equality, by Dehon’s Theorem [14] for $m \geq q$ and $m \equiv 1, 3(\text{mod } 6)$. The matrix achieving equality would be $[K^0_m K^1_m K^3_m T_{m,q-2}]$. Let $B^c$ denote the (0,1)-complement of a matrix $B$. Note that the $v \times \frac{q+1}{3} \binom{v}{2}$ simple matrix $[T_{v,a} T^c_{v,b}]$ has no submatrix $(a + b + 1) \cdot (1_2 0_0)$.

Theorem 1.5 Let $q > 2$ be given. There exists a constant $M = M(q)$ so that for $m > M$, 

$$\text{forb}(m, q \cdot (1_2 0_1)) \leq \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \end{bmatrix} \leq m + 2 + \frac{q+1}{3} \binom{m}{2} \quad (6)$$

with equality for $m \equiv 1, 3(\text{mod } 6)$. If $A$ is an $m \times \text{forb}(m, q \cdot (1_2 0_2))$ simple matrix with $m > M$ and $m \equiv 1, 3(\text{mod } 6)$, then $A$ consists of all possible columns of sum 0, 1, 2, $m$ and the columns of column sum 3 correspond to a simple triple system $T_{m,q-2}$ and $A$ has no further columns.

Theorem 1.6 Let $q > 2$ be given. There exists a constant $M = M(q)$ so that for $m > M$, 

$$\text{forb}(m, q \cdot (1_2 0_2)) \leq \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \end{bmatrix} \leq 2 + 2m + \frac{q+3}{3} \binom{m}{2}, \quad (7)$$

with equality for $m \equiv 1, 3(\text{mod } 6)$. If $A$ is an $m \times \text{forb}(m, q \cdot (1_2 0_2))$ simple matrix with $m > M$ and $m \equiv 1, 3(\text{mod } 6)$, then there exist positive integers $a, b$ with $a + b = q - 3$ so that $A$ consists of all possible columns of sum 0, 1, 2, $m - 2$, $m - 1$, $m$ and the columns
of column sum 3 correspond to a simple triple system $T_{m,a}$ and the columns of column sum $m - 3$ correspond to the complement of a simple triple system $T_{m,b}$ and $A$ has no further columns.

Thus the constructions for equality in Theorem 1.5 are $A = [K_0^0 K_1^1 K_2^2 T_{m,a} - 2 K_m^m]$ and the constructions for equality in Theorem 1.6 are found by selecting $a, b$ positive integers with $a + b = q - 3$ and using $A = [K_0^0 K_1^1 K_2^2 T_{m,a} T_{m,b} K_{m-2}^m K_{m-1}^m K_m^m]$. For $m = q + 1$, the construction $A = [K_0^0 K_1^1 K_2^2 K_3^3 K_m^m]$ avoids $q \cdot (1_2^0 0_1^1)$ and exceeds the bound (8) and for $m = q + 1$, the construction $A = [K_0^0 K_1^1 K_2^2 K_3^3 K_m^m - 2 K_{m-1}^m K_m^m]$ and avoids $q \cdot (1_2^0 0_1^1)$ and exceeds the bound (9) so our theorems need some condition on $m$.

To prove Theorem 1.1 and Theorem 1.5 we prove the following:

**Proposition 1.7** Let $m, q > 2$ be given. Let $A$ be an $m \times n$ $(0,1)$-matrix so that no column of sum 0,1,2, or $m$ is repeated. Assume $A$ has no configuration $q \cdot (1_2^0 0_1^1)$. Then there exists a constant $M$ so that for $m > M$,

$$n \leq m + 2 + \frac{q + 1}{3} \binom{m}{2}$$

with equality for $m \equiv 1, 3 \pmod{6}$. If $A$ is an $m \times \text{forb}(m, q \cdot (1_2^0 0_1^1))$ matrix with $m > M$ and $m \equiv 1, 3 \pmod{6}$, then $A$ consists of all possible columns of sum 0, 1, 2, $m$ once each and the columns of column sum 3 correspond to the triples of a 2-design $S_{q-2}(2,3,m)$ and $A$ has no further columns.

We see that Theorem 1.4 follows by taking a matrix $A$ of column sums in \{3, 4, \ldots, m - 1\} and with no configuration $(\lambda + 1) \cdot (1_2^0 0_1^1)$ and adding the $\binom{m}{2} + m + 2$ columns of column sum 0,1,2 and $m$ to obtain a matrix $A'$. Now $A'$ has no configuration $(\lambda + 2) \cdot (1_2^0 0_1^1)$ and satisfies the hypotheses of Proposition 1.7 with $q = \lambda + 2$. Applying Proposition 1.7 yields Theorem 1.1. The bound of Theorem 1.5 follows directly from Proposition 1.7. To prove Theorem 1.2 and Theorem 1.6 we prove the following:

**Proposition 1.8** Let $m, q > 2$ be given. Let $A$ be an $m \times n$ $(0,1)$-matrix so that no column of sum 0,1,2, $m - 2$, $m - 1$ or $m$ is repeated. Assume $A$ has no configuration $q \cdot (1_2^0 0_2^1)$. Then there exists a constant $M$ so that for $m > M$,

$$n \leq 2m + 2 + \frac{q + 3}{3} \binom{m}{2}$$

with equality for $m \equiv 1, 3 \pmod{6}$. If $A$ is an $m \times \text{forb}(m, q \cdot (1_2^0 0_2^1))$ matrix with $m > M$ and $m \equiv 1, 3 \pmod{6}$, then $A$ consists of all possible columns of sum 0, 1, 2, $m - 2$, $m - 1$ and $m$ once each and there are two positive integers $a, b$ satisfying $a + b = q - 3$ with the columns of column sum 3 correspond to the triples of a 2-design $S_{q-2}(2,3,m)$ and the columns of column sum $m - 3$ correspond to the complements in $[m]$ of the blocks of a 2-design $S_{q-2}(2,3,m)$ and $A$ has no further columns.
We see that Theorem 1.2 follows by taking a matrix $A$ of column sums in $\{3, 4, \ldots, m-3\}$ and with no configuration $(\lambda + 1) \cdot (1_2 0_2)$ and adding the $2\binom{m}{2} + 2m + 2$ columns of column sum $0, 1, 2, m - 2, m - 1$ and $m$ to obtain a matrix $A'$. Now $A'$ has no configuration $(\lambda + 3) \cdot (1_2 0_2)$ and satisfies the hypotheses of Proposition 1.8 with $q = \lambda + 3$. Applying Proposition 1.8 yields Theorem 1.2. The bound of Theorem 1.6 follows directly from Proposition 1.8.

We could give a simpler direct proof of Theorem 1.1 by using the proof of Proposition 1.8 and ignoring certain column sums. We were originally motivated by the forbidden configuration bounds of Theorems 1.5 and Theorem 1.6.

The proofs of Proposition 1.7 and Proposition 1.8 use Turán’s bound for the maximum number of edges in a graph with no complete graph of a certain size. We do not explicitly give values for $M$ since the values as given by the proofs are unlikely to be of value but our proof shows we may take $M$ to be $O(q^3)$. Proposition 1.7 for $q \cdot (1_1 0_1)$ is proven in Section 3 and Proposition 1.8 for $q \cdot (1_2 0_2)$ is proven in Section 4. The proofs are organized to highlight analogies with the proof of Theorem 1.3 but the details are different. We were surprised that exact bounds were obtained. We do not see how to extend our exact proofs to $F = t \cdot (1_k 0_k)$ with $k \geq 3$ and moreover do not have the analogue of Dehon’s lovely Theorem 1.4 to provide a construction of simple $k$-designs.

## 2 Exact Bound for $q \cdot (1_1 0_1)$

This section gives the proof of Theorem 1.3. We have broken it into lemmas. Assume $A$ is a simple $m$-rowed matrix with no configuration $q \cdot (1_1 0_1)$. Let $a_i$ denote the number of columns with either exactly $i$ 1’s or $i$ 0’s for $i = 0, 1, 2$ and let $a_3$ be the number of remaining columns. Without loss of generality, we may assume $a_0 = 2$ since the column of all 0’s and the column of all 1’s cannot contribute to $q \cdot (1_1 0_1)$. Thus $2 + a_1 + a_2 + a_3$ is the number of columns of $A$.

In [1], we establish that
\[
2 + a_1 + a_2 + a_3 \leq \left\lfloor \frac{(q + 1)m}{2} + \frac{(q - 3)m}{2(m - 2)} \right\rfloor + 2
\]
and as noted in the Introduction, we can achieve equality for some small $m$. We wish to show that these small values of $m$ are exceptional. We assume
\[
a_1 + a_2 + a_3 > \left\lfloor \frac{(q + 1)m}{2} \right\rfloor \tag{10}
\]
and seek a contradiction.

**Lemma 2.1** Let $A$ be an $m \times n$ simple matrix with no $q \cdot (1_1 0_1)$. Assume $m \geq 6$. Then
\[
(m - 1)a_1 + 2(m - 2)a_2 + 3(m - 3)a_3 \leq (2q - 2)\binom{m}{2} = (q - 1)m(m - 1). \tag{11}
\]
Assume \( n > \frac{q+1}{2} m + 2 \). Then
\[
2m - \frac{m(q - 3)}{m - 3} < a_1 \leq 2m, \tag{12}
\]
\[
a_3 < \frac{m(q - 3)}{m - 5}. \tag{13}
\]

**Proof:** A column of \( k \) 1’s contains \( \binom{k}{1} \binom{m-k}{1} \) configurations \( 1_10_1 \). Note that \( \binom{k}{1} \binom{m-k}{1} \geq \binom{3}{1} \binom{m-3}{1} \) for \( 3 \leq k \leq m - 3 \). By the pigeonhole argument, there are at most \( 2q-2 \binom{m}{2} \) configurations \( 1_10_1 \) in \( A \) else there will be \( 2q-1 \) in one of the \( \binom{m}{2} \) pairs of rows and hence at least \( q \) with the 1 of the \( 1_10_1 \) in the same row yielding the configuration \( q \cdot (1_10_1) \). This yields \((11)\). Given \( m \geq 6 \), we have \( m - 1 < 2(m - 2) < 3(m - 3) \). Substituting in \((11)\),
\[
a_1(m - 1) + 2(m - 2)(a_2 + a_3) \leq m(m - 1)(q - 1).
\]
Using \( a_2 + a_3 > \frac{q+1}{2} m - a_1 \) from \((10)\) we have
\[
a_1(m - 1) + 2(m - 2)\left(\frac{q+1}{2} m - a_1\right) < m(m - 1)(q - 1)
\]
and so
\[
2m^2 - mq - 3m < (m - 3)a_1
\]
from which we deduce the lower bound of \((12)\). The upper bound of \((12)\) follows from counting all possible columns.

To show \( a_3 \) is small, use \((11)\) to obtain
\[
a_1(m - 1) + 2(m - 2)\left(\frac{q+1}{2} m - a_1 - a_3\right) + 3(m - 3)a_3 < m(m - 1)(q - 1).
\]
Rearranging yields
\[
(m - 5)a_3 < m(q - 2m + 3) + (m - 3)a_1.
\]
Substituting \( a_1 \leq 2m \), we obtain \((13)\). \( \blacksquare \)

Form two graphs \( G_0, G_1 \) from the columns of \( A \) where the vertex set for both graphs corresponds to the rows of \( A \). We form a graph \( G_0 \) from the columns of \( A \) of column sum \( m - 2 \) so that if there is a column of \( A \) with \( m - 2 \) 1’s and two 0’s on rows \( i, j \) we add an edge \((i, j)\) to \( G_0 \). Similarly we form a graph \( G_1 \) from the columns of \( A \) of column sum 2 so that if there is a column of \( A \) with \( m - 2 \) 0’s and two 1’s on rows \( i, j \), then \( G_1 \) has the edge \((i, j)\). Define \( d_0(i) \) and \( d_1(i) \) to be the degrees of \( i \) in \( G_0 \) and \( G_1 \) respectively. Hence
\[
a_2 = \frac{1}{2} \sum_{i=1}^{m} (d_0(i) + d_1(i)). \tag{14}
\]
Using (10), we obtain
\[ a_1 + \frac{1}{2} \sum_{i=1}^{m} (d_0(i) + d_1(i)) + a_3 > \frac{q + 1}{2} m \]
Multiplying by 2 and substituting the upper bounds (12) for \( a_1 \) and (13) for \( a_3 \), yields
\[ \sum_{i=1}^{m} (d_0(i) + d_1(i)) > (q + 1)m - 4m - \frac{2m(q - 3)}{m - 5} \]
\[ = m(q - 3) \left( 1 - \frac{2}{m - 5} \right). \quad (15) \]
Thus the average value of \( d_0(i) + d_1(i) \) is close to \( q - 3 \).

The possible columns of column sum 1 or \( m - 1 \) are as follows. Define \( e_i \) to be the \( m \)-rowed column with a 1 in row \( i \) and 0’s elsewhere and let \( e_i^c \) be the \((0,1)\)-complement of \( e_i \). Define
\[ E_1 = \{ i : 1 \leq i \leq m \text{ and } e_i \text{ is not in } A \} \]
\[ E_0 = \{ i : 1 \leq i \leq m \text{ and } e_i^c \text{ is not in } A \} \]
We have \( a_1 = 2m - |E_0| - |E_1| \) and so \( |E_1| + |E_0| < \frac{m(q - 3)}{m - 3} \) by (12). For convenience of counting define
\[ \epsilon(i) = \begin{cases} 
0 & \text{if } i \notin E_1 \cup E_0 \\
1 & \text{if } i \in E_1 \setminus E_0 \text{ or } i \in E_0 \setminus E_1 \\
2 & \text{if } i \in E_1 \cap E_0 
\end{cases} \]. \quad (16)
Thus \( \sum_{i=1}^{m} \epsilon(i) = |E_0| + |E_1| \).

Lemma 2.2 Assume \( m > 3q + 2 \). Then for all \( i = 1, 2, \ldots, m \), we have \( d_0(i) + d_1(i) \leq q - 3 + \epsilon(i) \).

Proof: Assume the contrary that \( k \) is an index with \( l = d_0(k) + d_1(k) \geq q - 2 + \epsilon(k) \). Let \( N_1 \) be the vertices/rows connected to \( k \) by no edges in either \( G_0 \) or \( G_1 \). Let \( N_2 \) be the number of vertices connected to \( k \) by an edge in \( G_0 \) or an edge in \( G_1 \) but not both. Let \( N_3 \) be the number of vertices connected to \( k \) by edges in both \( G_0 \) and \( G_1 \). We have
\[ |N_1| + |N_2| + |N_3| = m - 1, \quad |N_2| + 2|N_3| = d_0(k) + d_1(k) = l. \quad (17) \]
Consider a row \( i \neq k \). There are at most \( 2q - 2 \) configurations \( 1_10_1 \) contained in rows \( k, i \) of \( A \) and there are \( 4 - \epsilon(i) - \epsilon(k) \) configurations \( 1_10_1 \) contained in rows \( k, i \) of \( A \) in the columns of column sum 1 or \( m - 1 \) (corresponding to those columns \( e_k, e_k^c, e_i, e_i^c \) which are present in \( A \)). If \( i \in N_1 \) then each edge incident with either \( k \) or \( i \) in either \( G_0 \) or \( G_1 \) corresponds to a column of \( A \) of column sum 2 or \( m - 2 \) that has the configuration
\(1_10_1\) in rows \(k, i\) and hence we have \(d_1(k) + d_0(k) + d_1(i) + d_0(i)\) configurations \(1_10_1\) in these columns. Thus
\[
d_1(k) + d_0(k) + d_1(i) + d_0(i) + (4 - \epsilon(i) - \epsilon(k)) \leq 2q - 2 \text{ which yields}
\]
\[
d_1(i) + d_0(i) \leq 2q - 6 - l + \epsilon(i) + \epsilon(k).
\]

In the case \(i \in N_2\) then we note that an edge in say \(G_0\) joining \(k, i\) contributes 2 to \(d_0(i) + d_0(k)\) but the corresponding column does not contain the configuration \(1_10_1\) in rows \(i, k\). A similar argument holds for an edge \((k, i)\) in \(G_1\). By the above analysis we obtain
\[
d_1(i) + d_0(i) \leq 2q - 4 - l + \epsilon(i) + \epsilon(k).
\]

In the case \(i \in N_3\) then we note that the two edges in \(G_0\) and \(G_1\) joining \(k, i\) contributes 4 to \(d_0(i) + d_1(i) + d_0(k) + d_1(k)\) but correspond to only two columns neither of which contain the configuration \(1_10_1\). By the above analysis we obtain
\[
d_1(i) + d_0(i) \leq 2q - 2 - l + \epsilon(i) + \epsilon(k).
\]

Summarizing, we have for \(i \in N_j\) and \(j = 1, 2, 3\) that
\[
d_1(i) + d_0(i) \leq 2q - 6 + 2(j - 1) - l + \epsilon(i) + \epsilon(k). \quad (18)
\]

Now we sum our upper bounds on \(d_0(i) + d_1(i)\) over all rows \(i \in [m] = \{k\} \cup N_1 \cup N_2 \cup N_3\) and use (15) to obtain
\[
l + \sum_{j \in \{1, 2, 3\}} \sum_{i \in N_j} (2q - 6 + 2(j - 1) - l + \epsilon(i) + \epsilon(k))
\]
\[
\geq \sum_{i=1}^{m} (d_0(i) + d_1(i)) > m(q - 3) \left(1 - \frac{2}{m - 5}\right)
\]
This simplifies to
\[
l + (2q - 6)(|N_1| + |N_2| + |N_3|) + 2(|N_2| + 2|N_3|) - (m - 1)l +
\]
\[
+(|E_0| + |E_1| - \epsilon(k)) + (m - 1)\epsilon(k) > m(q - 3)(1 - \frac{2}{m - 5})
\]
Using \(|N_1| + |N_2| + |N_3| = m - 1, l = |N_2| + 2|N_3|\) from (17), and \(|E_0| + |E_1| \leq \frac{m(q - 3)}{m - 3}\) and rearranging yields
\[
(2q - 6)(m - 1) - (m - 4)l + (m - 4)\epsilon(k) + 2\epsilon(k) + \frac{m(q - 3)}{m - 3} > m(q - 3) - \frac{2m(q - 3)}{m - 5}
\]
Using \(-l + \epsilon(k) \leq -(q - 2)\) and \(\epsilon(k) \leq 2\) and rearranging we get
\[
\frac{m(q - 3)}{m - 3} + \frac{2m(q - 3)}{m - 5} > m - 2. \quad (19)
\]
We can rewrite (19) as $0 > m^3 - (3q + 2)m^2 + (11q - 2)m - 30$ which is impossible for $m > 3q + 2$. This contradiction establishes the lemma. □

Let

$$Y = \{i : d_0(i) + d_1(i) = q - 3 \text{ and } \epsilon(i) = 0\}$$

**Lemma 2.3** Assume $m > \max\{3q + 2, 8q - 19\}$. Then we may assume $|Y| \geq m/2$.

**Proof:** We consider $[m]$ divided into $Y$, $E_0 \cup E_1$, and $[m]\backslash (Y \cup E_0 \cup E_1)$. We use Lemma 2.2. We have

$$\sum_{i \in E_0 \cup E_1} d_0(i) + d_1(i) \leq \sum_{i \in E_0 \cup E_1} (q - 3) + |E_0| + |E_1| = |E_0 \cup E_1|(q - 3) + |E_0| + |E_1|$$

using $\sum_{i=1}^m \epsilon(i) = |E_0| + |E_1|$. We readily compute $\sum_{i \in Y} d_0(i) + d_1(i) = |Y|(q - 3)$ and

$$\sum_{i \in [m]\backslash (Y \cup E_0 \cup E_1)} d_0(i) + d_1(i) \leq \sum_{i \in [m]\backslash (Y \cup E_0 \cup E_1)} (q - 4) \leq (m - |Y| - |E_0 \cup E_1|)(q - 4)$$

Summing we obtain

$$\sum_{i \in [m]} d_0(i) + d_1(i) \leq m(q - 3) + |E_0| + |E_1| - m + |Y| + |E_0 \cup E_1|$$

Now using (15), we deduce

$$|E_0| + |E_1| + |E_0 \cup E_1| + \frac{2m(q - 3)}{m - 5} > m - |Y|$$

We use $|E_0 \cup E_1| \leq |E_0| + |E_1| < \frac{m(q - 3)}{m - 3}$ by (12) to obtain $\frac{2m(q - 3)}{m - 3} + \frac{2m(q - 3)}{m - 5} > m - |Y|$. Now for $m > 8q - 19$ (so that $m - 3 > m - 5 \geq 8(q - 3)$), we have $\frac{2m(q - 3)}{m - 3} + \frac{2m(q - 3)}{m - 5} \leq m/2$. Thus for $m > 8q - 19$, we may assume $|Y| \geq m/2$. □

Let $A_3$ denote the submatrix of $A$ formed by the columns of sum 3, 4, . . . , or $m - 3$. Then $A_3$ has $a_3$ columns. Let $A_3(Y)$ denote the submatrix of $A_3$ indexed by the rows of $Y$.

**Lemma 2.4** Assume $m > \max\{3q + 2, 8q - 19\}$. Then $A_3(Y)$ has no configuration $1_10_1$.

**Proof:** Assume there is a column $\alpha$ in $A_3$ which has both 0’s and 1’s in the rows indexed by $Y$. By taking the (0,1)-complement of $A$ if necessary, we may assume the number of 1’s in those rows is at least $|Y|/2 \geq m/4$. Consider a row $i \in Y$ where $\alpha$ has a 0. Then there exists a row $j \in Y$ where $\alpha$ has a 1 such that rows $i, j$ are not connected in $G_0$ or $G_1$, since $i$ is connected to at most $q - 3$ rows and $|Y|/2 \geq m/4 > q - 3$. Given $i, j \in Y$, we have $d_0(i) + d_1(i) = d_0(j) + d_1(j) = q - 3$ and $\epsilon(i) = \epsilon(j) = 0$. Given that $i, j$ are not
connected in \( G_0 \) or \( G_1 \), we have \( 2(q - 3) \) copies of the configuration \( 1_10_1 \) on rows \( i, j \) in the columns (of \( A \)) of exactly two 1’s or exactly two 0’s. Given \( Y \cap (E_1 \cup E_2) = \emptyset \), we have 4 copies of the configuration \( 1_10_1 \) in the columns (of \( A \)) of one 1 or one 0 and in rows \( i, j \). But \( \alpha \) has \( 1_10_1 \) in rows \( i, j \) and so we find \( q \cdot (1_10_1) \) in \( A \), a contradiction. This establishes the lemma. \( \blacksquare \)

**Proof of Theorem 1.3** We obtain a contradiction from assuming (10) and \( m > \max\{3q + 2.8q - 19\} \) and thus establish (3). By Lemma 2.4 each column of \( A_3 \) has either all 1’s or all 0’s on the rows of \( Y \). Considering column sums, every column in \( A_3 \) which has all 1’s on rows \( Y \), has at least three 0’s and every column in \( A_3 \) which has all 0’s on rows \( Y \), has at least three 1’s. For \( i \in [m] \setminus Y \), let \( t_0(i) \) denote the number of 0’s in columns of column sum in \( \{3, 4, \ldots, m - 3\} \) which are all 1’s on \( Y \) and let \( t_1(i) \) denote the number of 1’s in columns of column sum in \( \{3, 4, \ldots, m - 3\} \) which are all 0’s on \( Y \). Counting yields

\[
\sum_{i \in [m] \setminus Y} (t_0(i) + t_1(i)) \geq 3a_3. \tag{20}
\]

Let \( i \in [m] \setminus Y \) be given. We wish to establish

\[
d_0(i) + d_1(i) \leq q - 3 + \epsilon(i) - t_0(i) - t_1(i) \tag{21}
\]

We use a similar argument as Lemma 2.2. Consider a column of \( A_3 \) which is all 0’s on rows of \( Y \). Then the column has a configuration \( 1_10_1 \) in rows \( i,k \) for any choice of \( k \in Y \). A similar remarks holds for columns of \( A_3 \) which are all 1’s on rows of \( Y \). Let \( X \) denote all the neighbours of \( i \) in \( G_0 \) and in \( G_1 \). We have \( |X| \leq d_0(i) + d_1(i) \leq q - 1 \) using Lemma 2.2. Given \( |Y| > m/2 > q - 1 \), we can select a \( k \in Y \) with \( k \notin X \). Now the columns of sum 1 or \( m - 1 \) in \( A \) have \( 4 - \epsilon(i) - \epsilon(k) = 4 - \epsilon(i) \) configurations \( 1_10_1 \) in rows \( i,k \) (since \( \epsilon(k) = 0 \)). Given that \( k \notin X \), the columns of column sum 2 or \( m - 2 \) have \( d_0(i) + d_1(i) + d_0(k) + d_1(k) = d_0(i) + d_1(i) + q - 3 \) configurations \( 1_10_1 \) in rows \( i,k \). The columns of sum at least 3 and at most \( m - 3 \) have at least \( t_0(i) + t_1(i) \) configurations \( 1_10_1 \) in rows \( i,k \) for that choice of \( k \). Rows \( i,k \) of \( A \) have at most \( 2(q-1) \) such configurations and so we obtain (21).

Combining twice (10) and (14) we have

\[
2a_1 + \sum_{i=1}^{m} (d_0(i) + d_1(i)) + 2a_3 > m(q + 1).
\]

Using \( a_1 = 2m - |E_0| + |E_1| \), substituting \( d_0(i) + d_1(i) = q - 3 \) for \( i \in Y \) and using (21),

\[
\sum_{i \in Y} (q - 3) + \sum_{i \in [m] \setminus Y} (q - 3 + \epsilon(i) - t_0(i) - t_1(i)) + 2a_3 > m(q + 1) - 2(2m - |E_0| + |E_1|).
\]

Now using (20) and \( \sum_{i \in [m] \setminus Y} \epsilon(i) \leq |E_0| + |E_1| \),

\[
|Y|(q - 3) + (m - |Y|)(q - 3) + |E_0| + |E_1| - a_3 > m(q - 3) + 2(|E_0| + |E_1|)
\]

11
which yields the contradiction (even for $a_3 = 0$ and $|E_0| + |E_1| = 0$)

$$-a_3 > |E_0| + |E_1|.$$ 

This final contradiction establishes (5).

One could note that for a matrix $A$ to achieve equality, we would have $a_3 = 0$ and $|E_0| + |E_1| = 0$ and so $a_1 = 2m$. This suggests that $A$ would have to correspond to the construction given in the Introduction or its $(0,1)$-complement.

3 Exact Bound for $q \cdot (1_20_1)$

We are able to generalize the argument for Theorem 1.3 following a similar series of Lemmas to obtain Proposition 1.7. We do not explicitly calculate the smallest possible constant $M$ for our proof (following the argument yields that $M$ is $O(q^3)$), believing that our argument does not give a realistic values for $M$. Let $A$ be an $m \times n$ $(0,1)$-matrix with no configuration $q \cdot (1_20_1)$ so that there are no repeated columns of sum 0, 1, 2, $m$. We wish to ignore the $m+2$ possible columns of sum 0, 1, $m$ since they cannot contribute to a configuration $q \cdot (1_20_1)$. So assume $A$ has column sums between 2 and $m-1$, inclusive.

Assume $n > \frac{a_4+1}{3}(\binom{m}{3})$. We wish to arrive at a contradiction to prove (8).

For $i = 2, 3$, let $a_i$ denote the number of columns of column sum $i$ in $A$ and let $a_4$ denote the number of columns of column sum at least 4 in $A$. Note that the definition of $a_1, a_2, \ldots$ is different in this section from Sections 2 and 4. Note that we do not allow repeated columns of sum 2. We have by assumption that

$$a_2 + a_3 + a_4 > \frac{q+1}{3}\binom{m}{2}. \quad (22)$$

**Lemma 3.1** Let $m, q$ be given. Let $A$ be an $m \times n$ simple matrix with no $q \cdot (1_20_1)$. Assume $m \geq 6$ and (22). Then

$$\binom{2}{2}(m-2)\binom{1}{1}a_2 + \binom{3}{2}(m-3)\binom{1}{1}a_3 + \binom{4}{2}(m-4)\binom{1}{1}a_4 \leq \binom{m}{3}3(q-1). \quad (23)$$

There exists positive constants $c_1, c_2$ so that

$$\binom{m}{2} - c_1m \leq a_2 \leq \binom{m}{2}, \quad (24)$$

$$a_4 \leq c_2m. \quad (25)$$

**Proof:** We note that a column of column sum $k$ has $\binom{k}{2}\binom{m-k}{1}$ configurations $1_20_1$, and note that $\binom{2}{2}(m-k) \geq \binom{4}{2}(m-4)$ for $4 \leq k \leq m-1$. Counting the configurations $1_20_1$ and using the pigeonhole argument yields (23)
For $m \geq 6$ we have $\binom{3}{2} (m-3) < \binom{4}{2} (m-4)$. Hence

$$(m-2)a_2 + 3(m-3)(a_3 + a_4) \leq \binom{m}{3} 3(q - 1)$$

From (22), we have $a_3 + a_4 \geq \frac{q+1}{3} \binom{m}{2} - a_2$. We substitute and obtain

$$(m-3) \binom{m}{2} (q+1) - \binom{m}{3} 3(q - 1) \leq \left(3(m-3) - (m-2)\right) a_2$$

which simplifies as

$$\binom{m}{2} (2m - q - 5) \leq (2m - 7)a_2$$

from which we deduce that there is a constant $c_1$ (will depend on $q$) so that first half of (24) holds. The second half of (24) follows from the fact that no column of sum 2 is repeated.

In a similar way we have

$$(m-2)a_2 + 3(m-3) \left( \frac{q+1}{3} \binom{m}{2} - a_2 - a_4 \right) + 6(m-4)a_4 \leq \binom{m}{3} 3(q - 1)$$

and when we substitute the upper bound of (24), we deduce that there is a constant $c_2$ (will depend on $q$) so that (25) holds.

Partition $A$ into three parts: $A_2$ consists of the columns of column sum 2, $A_3$ is the columns of column sum 3 and $A_4$ is the columns of column sum greater or equal than 4. We will refer to $A_2$, $A_3$ using the notations of (3) and (4). Note that $A_3$ is a multiset and $A_2$ is a set given that there are no repeated columns of sum 2. Considering the columns of column sum 2, we adapt $\epsilon(i)$ of (10). Note that for convenience we represent every pair $\{i, j\}$ by $ij$ and so $ij \equiv ji$. We are not interested in ordered pairs in this context. Define

$$\epsilon(ij) = \begin{cases} 1 & \text{if } \{i, j\} \notin A_2 \\ 0 & \text{if } \{i, j\} \in A_2 \end{cases} \quad E = \{ij : \epsilon(ij) = 1\}.$$  

Thus

$$a_2 = \binom{m}{2} - \sum_{ij} \epsilon(ij) = \binom{m}{2} - |E|. \quad (26)$$

We deduce from (24) that $|E| \leq c_1 m$.

We adapt the definitions of the degrees $d_0$, $d_1$ of Section 2 by using a hypergraph degree definitions applied to the multiset $A_3 = \{B_1, B_2, \ldots\}$. Define

$$d(ij) = |\{s : B_s \in A_3 \text{ and } i, j \in B_s\}|$$

Then

$$3a_3 = \sum_{\{i,j\} \subseteq [m]} d(ij). \quad (27)$$
Let 
\[ U(pt) = \{ r : \{ p, t, r \} \in A_3 \}. \]

Since \( m > q + 2 \) and we are avoiding \( q \cdot (1_2 0_1) \) in \( A_3 \) then \( |U(pt)| < q \). Also let 
\[ T(r) = \{ pt : \{ p, t, r \} \in A_3 \}. \]

Since for every \( x \in [m] \) with \( x \neq r \), \( |U(rx)| < q \) we have \( |T(r)| < \frac{(m-1)q}{2} \). Note that \( U(pt) \) and \( T(r) \) are the generalizations of \( X \) (found after (21)) given in the proof of Theorem 1.3.

**Lemma 3.2** We have 
\[ d(ij) \leq (q - 2) + \epsilon(ij). \] (28)

**Proof:** Since \( m > q + 2 \geq |U(ij)| + 2 \), for every pair \( ij \), we can find row \( k \neq i, j \) so that \( k \notin U(ij) \). Now the number of submatrices
\[
\begin{bmatrix}
i \\
j \\
k
\end{bmatrix}
\]

in \( A_3 \) is \( d(ij) \) (since \( d(ij) \) is the number of triples \( i, j, l \) corresponding to columns in \( A_3 \) and each such column yields the submatrix since \( k \notin U(ij) \)) and the number of submatrices (29) in \( A_2 \) is \( 1 - \epsilon(ij) \). Thus
\[ d(ij) + 1 - \epsilon(ij) \leq q - 1 \]
and hence (28) holds. \( \blacksquare \)

Let
\[ Y = \{ ij : d(ij) = q - 2 \text{ and } \epsilon(ij) = 0 \} \]

**Lemma 3.3** There exists a constant \( c_3 \) so that
\[ |Y| \geq \binom{m}{2} - c_3 m \] (30)

**Proof:** We partition the \( \binom{m}{2} \) pairs \( ij \) into 3 parts: \( Y \), \( E \) and the rest. By Lemma 3.2, for each \( ij \in E \) we have \( d(ij) \leq (q - 2) + 1 \). Note that for \( ij \notin Y \cup E \), we have \( \epsilon(ij) = 0 \) and so \( d(ij) \leq (q - 2) - 1 \) else \( ij \in E \). Thus from (27)
\[
3a_3 = \sum_{ij} d(ij) \leq \left( (q - 2)|Y| + ((q - 2) + 1)|E| + ((q - 2) - 1)\left( \binom{m}{2} - |Y| - |E| \right) \right)
\]
Hence
\[ a_3 \leq \frac{1}{3} \left( (q - 2)\binom{m}{2} + |E| - \binom{m}{2} + |Y| + |E| \right) \] (31)
Substituting estimates of $a_2, a_3, a_4$ from (26), (31), (25) into (22), we have

$$\left(\frac{m}{2}\right) - |E| + \frac{1}{3} \left((q - 2)\left(\frac{m}{2}\right) + 2|E| - \left(\frac{m}{2}\right) + |Y|\right) + c_2 m > \frac{q + 1}{3} \left(\frac{m}{2}\right)$$

We deduce $-\frac{1}{3}|E| + \frac{1}{3}|Y| + c_2 m > \frac{1}{3} \left(\frac{m}{2}\right)$ and so there exists a constant $c_3 = 3c_2$ so that (30) holds.

Form a graph $G$ of $m$ vertices corresponding to the rows of $A$ and with edges $(i, j)$ if and only if $ij \in Y$. Thus by Lemma 3.3, the number of edges of $G$ is at least $\left(\frac{m}{2}\right) - c_3 m$.

By Turán’s Theorem [7], a graph with more than $\frac{m^2}{2} - \frac{m^2}{2(k-1)}$ edges has a clique of $k$ vertices. Thus $G$ has large cliques. Let $c_4$ be a constant chosen so that for any choices of $i, j, k$ the following three inequalities hold.

$$\left(\frac{c_4\sqrt{m}}{2}\right) > \frac{m-1}{2}q(> |T(k)|), \quad \frac{c_4\sqrt{m}}{2} > q(> |U(ij)|),$$

$$\left(\frac{c_4\sqrt{m}}{2}\right) > \frac{m-1}{2}q + 3m(\geq |T(k)| + |U(ij)|)$$

(32)

By Turán’s argument, there exists an $M$ so that for $m \geq M$, we can find a clique of $c_4\sqrt{m}$ vertices in $G$. Let the vertices in this clique be denoted $B$. Thus for $i, j \in B$ we have $d(ij) = q - 2$ and $\epsilon(ij) = 0$. Let $A_4(B)$ be the submatrix of $A_4$ of the rows indexed by $B$.

**Lemma 3.4** Assume $m > M$. Then $A_4(B)$ has no configuration $1_20_1$.

**Proof:** Consider a column $\alpha$ of $A_4$. We consider two cases based on whether there are more 1’s or more 0’s in the rows $B$. Assume $\alpha$ has at least $\frac{c_4\sqrt{m}}{2}$ 1’s in rows of $B$. Assume $\alpha$ has a 0 in row $k \in B$. Then by the first inequality (32), there is a pair $ij \notin T(k)$ with $i, j \in B$. Thus there are $q - 2$ columns of column sum 3 with the submatrix (29) using $d(ij) = q - 2$ and 1 column of column sum 2 with the submatrix (29) using $\epsilon(ij) = 0$ and column $\alpha$ has 1 further submatrix (29) which creates the configuration $q \cdot (1_20_1)$, a contradiction. So $\alpha$ has no configuration $1_20_1$.

Assume $\alpha$ of $A_4$ that has at least $\frac{c_4\sqrt{m}}{2}$ 0’s in the rows $B$. Assume $\alpha$ has 1’s in rows $i, j \in B$. Then there is a row $k \in B$ where $\alpha$ has a 0 in row $k$ and $k \notin U(ij)$ by the second inequality of (32). For that choice of $k$ and using $d(ij) = q - 2$, there are $q - 2$ columns of column sum 3 with the submatrix (29). There is one column of column sum 2 with the submatrix (29) using $\epsilon(ij) = 0$ and the $\alpha$ has one further submatrix (29) which creates the configuration $q \cdot (1_20_1)$, a contradiction. Thus $\alpha$ has no configuration $1_20_1$.

**Lemma 3.5** Assume $m > M$. Then the inequality (3) holds.
Proof: We obtain a contradiction from assuming \( m > M \) and (22) and thus establish (8). Our proof considers the \( a_4 \) columns of \( A_4 \) (which are the columns of column sum at least 4 and at most \( m - 1 \)).

From Lemma 3.4, each column in \( A_4 \) either has at most one 1 or has no 0’s in the rows of \( B \). Let \( A_0^4 \) be those columns of \( A_4 \) with at most one 1 in the rows of \( B \) and hence at least three 1’s in the rows \([m] \setminus B\). Let \( a_0^4 \) be the number of columns in \( A_0^4 \). Let \( A_1^4 \) be those columns of \( A_4 \) with no 0’s in the rows of \( B \) and hence at least one 0 in the rows \([m] \setminus B\). Let \( a_1^4 \) be the number of columns of \( A_1^4 \). We have \( a_0^4 + a_1^4 = a_4 \).

For a pair \( ij \) with \( i, j \in [m] \setminus B \), let \( t(ij) \) count the number of columns of \( A_0^4 \) with 1’s in both rows \( i \) and \( j \). Each column with at most one 1 in \( B \) has at least three 1’s in \([m] \setminus B\) and hence 1’s in at least \( \binom{3}{2} = 3 \) pairs \( ij \) with \( i, j \in [m] \setminus B \). We have verified that
\[
\sum_{ij : i,j \in [m] \setminus B} t(ij) \geq 3a_0^4. \tag{33}
\]

We must work harder to get an analog of (33) for \( A_1^4 \). Assume \( a_1^4 > 0 \). For a pair \( ij \) with \( i, j \in B \) and \( k \in [m] \setminus B \) with \( ij \notin T(k) \), let \( t(ij,k) \) denote the number of submatrices (29) in \( A_1^4 \). When \( ij \in T(k) \), set \( t(ij,k) = 0 \). For a pair \( ij \) with \( i, j \in B \), let
\[
t(ij) = \max_{k \in [m] \setminus B} t(ij,k) \tag{34}
\]

Each column \( \alpha \) in \( A_1^4 \) has at least one row, say \( l \in [m] \setminus B \) with a 0. For column \( \alpha \), we know \( |T(l)| < \frac{m-1}{2}q \) and at the same time there are \( \binom{c_4 \sqrt{m}}{2} \) pairs \( ij \) with \( i, j \in B \) and so there are at least \( \binom{c_4 \sqrt{m}}{2} - \frac{m-1}{2}q \) pairs \( ij \) with \( i, j \in B \) with \( ij \notin T(l) \). Thus by the third inequality of (32), column \( \alpha \) contributes at least \( 3m \) to the sum \( \sum_{ij \notin T(l)} t(ij,l) \) and so
\[
\sum_{l \in [m] \setminus B} \sum_{ij : i,j \in B} t(ij,l) > 3ma_1^4
\]
Thus by (34),
\[
(m - |B|) \cdot \sum_{ij : i,j \in B} t(ij) > \sum_{l \in [m] \setminus B} \sum_{ij : i,j \in B} t(ij,l)
\]
and so we deduce that
\[
\sum_{ij : i,j \in B} t(ij) > \frac{3ma_1^4}{m - |B|} > 3a_1^4. \tag{35}
\]

For a pair \( ij \) with \( i \in B \) and \( j \in [m] \setminus B \) or vice versa, let \( t(ij) = 0 \). We add (33) and (35) together to get
\[
\sum_{ij} t(ij) \geq 3a_4, \tag{36}
\]
with strict inequality if \( a_1^4 > 0 \).
We are able to extend Lemma 3.2 and establish

\[ d(ij) \leq q - 2 + \epsilon(ij) - t(ij) \quad (37) \]

By Lemma 3.2, we need only consider \( ij \) with \( t(ij) > 0 \). Given the definition of \( t(ij) \), we need only consider the two cases: \( i, j \in [m] \setminus B \) or \( i, j \in B \).

In the former case we note that each of the \( t(ij) \) columns of \( A_0^i \) with 1’s in both rows \( i \) and \( j \) have at most one 1 in rows of \( B \). With \( |B| > 2q \), by the second inequality of (32), we deduce that \( t(ij) < q \) else we will find the configuration \( q \cdot (1_2 0_1) \) in \( A_0^i \) in the rows \( i, j \) and a row of \( B \). Now in these \( t(ij) \) columns of \( A_0^i \), at least \( |B| - q + 1 \) rows of \( B \) are all 0’s. Again using the second inequality of (32) that \( |B| > 2q \), we can find some \( k \in B \) with \( k \notin U(ij) \) and all the \( t(ij) \) columns have 0’s in row \( k \). Now in these \( t(ij) \) columns of \( A_0^i \), at least \( |B| - q + 1 \) rows of \( B \) are all 0’s. Again using the second inequality of (32) that \( |B| > 2q \), we can find some \( k \in B \) with \( k \notin U(ij) \) and all the \( t(ij) \) columns have 0’s in row \( k \). Now in these \( t(ij) \) columns of \( A_0^i \), at least \( |B| - q + 1 \) rows of \( B \) are all 0’s.

In the latter case with \( i, j \in B \), we select \( k \) so that \( t(ij, k) = t(ij) \). Thus \( k \notin U(ij) \) and also there are at least \( t(ij) \) submatrices (29) in columns of \( A_1^i \). Thus we can now follow the same argument as in the former case to establish (37).

Now using (37) and (36),

\[ 3a_3 = \sum_{ij} d(ij) \leq \sum_{i,j} \left( q - 2 + \epsilon(ij) - t(ij) \right) = \left( q - 2 \right) \left( \binom{m}{2} \right) + |E| - 3a_4. \quad (38) \]

Substituting (26), (38), and (25) in (22) we obtain

\[ \left( \binom{m}{2} \right) - |E| + \frac{1}{3} \left( q - 2 \right) \left( \binom{m}{2} \right) + |E| - 3a_4 \] \[ + a_4 > \frac{q + 1}{3} \left( \binom{m}{2} \right). \]

Simplifying and rearranging,

\[ -\frac{2}{3} |E| > 0 \]

which is a contradiction (even for \(|E| = 0\)) and this establishes (8).

**Proof of Proposition 1.7** Lemma 3.5 establishes most of Proposition 1.7 but we are also interested in cases when the bound is achieved. Assume \( m > M \) and \( m \equiv 1, 3 \text{ (mod 6)} \). We now consider an \( m \)-rowed simple matrix \( A \) which has no configuration \( q \cdot (1_2 0_1) \) and with \( \binom{m}{0} + \binom{m}{1} + \frac{q-1}{3} \binom{m}{2} + \binom{m}{m} \) columns. One repeats the previous lemmas and arguments replacing the inequality (22) with the equation

\[ a_2 + a_3 + a_4 = \frac{q + 1}{3} \binom{m}{2}. \quad (39) \]

We wish to show \( a_2 = \binom{m}{2}, a_4 = 0, a_3 = \frac{q-2}{3} \binom{m}{2} \). Now Lemma 3.1 holds with (22) as an equality. We deduce the same bounds for \( U(rx) \) and \( T(r) \). Lemma 3.2 still holds
since the final contradiction does not require the strict inequality of (22) merely the equality of (39). Lemma 3.3 holds and we can choose \( B \) as large as possible but at least satisfying the three inequalities (32). Lemma 3.4 continues to hold.

We use (39) and following the argument of Lemma 3.5 we deduce that \( E = \emptyset \) and so \( a_2 = \binom{m}{2} \). Also we deduce that

\[
\sum_{ij} t(ij) = 3a_4
\]

and as a result of the strict inequality in (35), we can deduce that \( a_1^4 = 0 \).

Assume \( a_4 = a_1^4 > 0 \) and consider \( \alpha \) in \( A_4 \) with column sum 4 and with 1’s in rows \( i, j, k, l \) where \( i \in B \) and \( j, k, l \in \{1, 2, \ldots, m\} \setminus B \). Choose \( r \in B \setminus i \) then \( \alpha \) has 1’s in rows \( i, j \) and 0’s in row \( r \). Using \( E = \emptyset \), we deduce that for this particular \( i, j \) we have \( d(ij) \leq (q - 2) - 1 \). This yields a slight variant of (38):

\[
3a_3 = \sum_{ij} d(ij) \leq \sum_{ij} ((q - 2) + \epsilon(ij) - t(ij)) - 1.
\]

The extra ‘-1’ is sufficient to obtain a contradiction when we substitute for \( a_2, a_3, a_4 \) in (39). We then deduce \( a_4 = 0 \).

With \( a_4 = 0 \) and \( a_2 = 2 \binom{m}{2} \), we deduce \( a_3 = \frac{q-3}{3} \binom{m}{2} \) using (39). Given that \( \epsilon(ij) = 0 \) for all \( ij \) and using Lemma 3.2, we deduce \( d(ij) = q - 2 \) for all pairs \( ij \) and so \( B = \{1, 2, \ldots, m\} \). From this we can readily conclude that the columns of column sum 3 correspond to a 2-design \( S_{q-2}(2, 3, m) \) and \( A \) has no further columns. \( \blacksquare \)

4 Exact Bound for \( q \cdot (1_20_2) \)

We generalize our proof of Proposition 1.7 given in Section 3 to prove Proposition 1.8. Again we do not explicitly calculate the smallest possible constant \( M \) but we note that we can take \( M = O(q^3) \).

Let \( A \) be a \( m \times n \) matrix with no \( q \cdot (1_20_2) \). Assume that there are no repeated columns of sums 0, 1, 2, \( m - 2, m - 1, m \). We will assume \( n > 2 + 2m + \left( \binom{m}{2} \frac{q+3}{3} \right) \). Let \( a_i \) denote the number of columns with either exactly \( i \) 1’s or \( i \) 0’s for \( i = 0, 1, 2, 3 \) and let \( a_4 \) be the number of remaining columns. We may assume \( a_0 = 2 \) and \( a_1 = 2m \) since all columns of column sum 0, 1, \( m - 1 \) or \( m \) do not contain the configuration \( 1_20_2 \). Thus

\[
a_2 + a_3 + a_4 > \binom{m}{2} \frac{q+3}{3}. \tag{40}
\]

Lemma 4.1 Assume \( A \) is an \( m \times n \) simple matrix with no configuration \( q \cdot (1_20_2) \) and (40) holds. Then there exists an \( m_0 \) so that for \( m > m_0 \),

\[
\binom{2}{2} \binom{m-2}{2} a_2 + \binom{3}{2} \binom{m-3}{2} a_3 + \binom{4}{2} \binom{m-4}{2} a_4 \leq 6 \binom{m}{4} (q - 1). \tag{41}
\]
Also there exist constants $c_1, c_2$ so that

$$2 \binom{m}{2} - c_1 m \leq a_2 \leq 2 \binom{m}{2}$$  \hfill (42)

$$a_4 \leq c_2 m$$  \hfill (43)

**Proof:** A column in $A$ of column sum $k$ has $\binom{k}{2}\binom{m-k}{2}$ configurations $1_20_2$. Note that $\binom{k}{2}\binom{m-k}{2} \geq \binom{4}{2}\binom{m-4}{2}$ for $4 \leq k \leq m - 4$. By the pigeonhole principle, there are at most $6(q-1)\binom{m}{2}$ configurations $1_20_2$ in $A$. We obtain (41). There exist an $m_0$, such that for $m > m_0$, $\binom{3}{2}\binom{m-3}{2} < \binom{4}{2}\binom{m-4}{2}$. Substituting in (41),

$$\left(\binom{m-2}{2}\right)a_2 + \left(\binom{m-3}{2}\right)(a_3 + a_4) \leq 6(q-1)\binom{m}{4}$$

which yields using $a_3 + a_4 > \binom{m}{2}\frac{q+3}{3} - a_2$ and rearranging

$$\left(\binom{m-2}{2}\right)a_2 + 3\left(\binom{m-3}{2}\right)\left[\binom{m}{2}\frac{q+3}{3} - a_2\right] < 6(q-1)\binom{m}{4}. \hfill (44)$$

Therefore,

$$\left(\binom{m-3}{2}\right)\left(\binom{m}{2}\right)(q+3) - 6(q-1)\binom{m}{4} < \left(3\left(\binom{m-3}{2}\right) - \left(\binom{m-2}{2}\right)\right)a_2. \hfill (44)$$

The leading term on the righthand side is exactly $m^4$ while the leading coefficient of $a_2$ on the lefthand side is exactly $m^2$. Thus (44) implies that there exists some constant $c_1$ so that the lower bound of (42) holds. The upper bound of (42) follows from the fact no column of sum 2 or $m - 2$ is repeated.

We can also bound $a_4$. From (40), we have $a_3 > \frac{q+3}{3}\binom{m}{2} - a_2 - a_4$. Using (41) we have

$$\left(\binom{m-2}{2}\right)a_2 + 3\left(\binom{m-3}{2}\right)\left[\frac{q+3}{3}\binom{m}{2} - a_2 - a_4\right] + 6\binom{m-4}{2}a_4 \leq 6(q-1)\binom{m}{4}.$$

Then

$$\left[6\binom{m-4}{2} - 3\binom{m-3}{2}\right]a_4 \leq 6(q-1)\binom{m}{4} - (q+3)\binom{m-3}{2}\binom{m}{2} + a_2 \left[3\binom{m-3}{2} - \binom{m-2}{2}\right].$$

Substituting $a_2 \leq 2\binom{m}{2}$ and rearranging we have

$$\left[6\binom{m-4}{2} - 3\binom{m-3}{2}\right]a_4 \leq \frac{m(m-1)(m-3)}{4}(2q-6). \hfill (45)$$
Then (43) implies that there exist some constant $c_2$ so that (43) holds.

We could have produced the bound $a_4 \leq (2q - 6)\frac{m}{6} + c'_2$ for some constant $c'_2$, but this is of little help. Now we form analogs of the degrees $d_0, d_1$ of Section 2 by defining $A_3$ as the submatrix of $A$ of the columns of column sum 3 and defining $A_{m-3}$ as the submatrix of $A$ of the columns of column sum $m - 3$. We refer to the multisets $A_3 = \{B_1, B_2, \ldots\}$, $A_{m-3} = \{C_1, C_2, \ldots\}$ using the notations of (3) and (4). Define

$$d_1(ij) = |\{s : B_s \in A_3 \text{ and } i, j \in B_s\}|, \quad d_0(ij) = |\{s : C_s \in A_{m-3} \text{ and } i, j \notin C_s\}|$$

Recalling $a_3 = |A_3| + |A_{m-3}|$, we note

$$3a_3 = \sum_{\{i,j\} \subseteq [m]} (d_0(ij) + d_1(ij)) \quad (46)$$

Define $e_{ij}$ to be the $m$-rowed column with 1 in rows $i$ and $j$ and 0’s elsewhere, and let $e'_{ij}$ be the (0,1)-complement of $e_{ij}$. These are the possible columns of column sum 2 or $m - 2$. Define

$$E_1 = \{ij : \{i, j\} \subseteq [m] \text{ and } e_{ij} \text{ is not in } A\} \quad E_0 = \{ij : \{i, j\} \subseteq [m] \text{ and } e'_{ij} \text{ is not in } A\}$$

For convenience of counting define

$$\epsilon(ij) = \begin{cases} 0 & \text{if } ij \notin E_1 \cup E_0 \\ 1 & \text{if } ij \in E_1 \setminus E_0 \text{ or } ij \in E_0 \setminus E_1 \\ 2 & \text{if } ij \in E_1 \cap E_0 \end{cases}. \quad (47)$$

Thus

$$a_2 = 2 \binom{m}{2} - \sum_{i,j \subseteq [m]} \epsilon(ij) = 2 \binom{m}{2} - (|E_1| + |E_0|), \quad (48)$$

and given (42) we have $|E_1| + |E_0| \leq c_1m$.

We note for a quadruple of rows $p, t, r, s$ that are at most $2q - 2$

$$
\begin{bmatrix}
1 \\
1 \\
0 \\
0
\end{bmatrix}
\quad \text{or submatrices}
\begin{bmatrix}
1 \\
0 \\
1 \\
1
\end{bmatrix}
\quad (49)
$$

else $A$ has the configuration $q \cdot (1_20_2)$. For disjoint pairs $pt$ and $rs$ (i.e. $\{p, t\} \subseteq \{r, s\} = \emptyset$) we say pair $pt$ has triple overlapping $rs$ if and only if at least one of submatrices
Given $m > q$, column sum $3$ else we would have the configuration for columns with three $0$'s. Let $m - pt, rs$ appears in columns of column sum $3$ or at least one of submatrices note that $m$ exists a constant of column sum $3$ cannot have the submatrix $q$ $r, s$ or vice versa have $1$'s on rows $r, s$ yet no $1$'s on rows $p, t$ contribute to (49). Similarly for columns with three $0$'s. Let

\[
U(pt) = \{ij : \{i, j\} \subset [m], \text{pair } pt \text{ has triple overlapping } ij\},
\]

\[
T(pt) = \{ij : \{i, j\} \subset [m], \text{pair } ij \text{ has triple overlapping } pt\}.
\]

Given $m > q + 2$, we cannot have the submatrix $q \cdot (1_20_0)$ in rows $p, t$ in columns of column sum $3$ else we would have the configuration $q \cdot (1_20_2)$ (and so there are at most $q - 1$ columns of column sum $3$ with $1$'s in rows $p, t$). Similarly, we cannot have the submatrix $q \cdot (1_00_2)$ in rows $p, t$ in columns of column sum $m - 3$. To bound $U(pt)$, we note that $(m - 2 - (q - 1))$ counts the number of pairs $ij$ disjoint from $pt$ that avoids $q - 1$ further rows. Thus the number of pairs $ij$ where $pt$ overlaps $ij$ using a column of column sum $3$ is at most $(m - 2) - (m - 2 - (q - 1))$. Similarly, the number of pairs $ij$ where $pt$ overlaps $ij$ using a column of column sum $m - 3$ is at most $(m - 2) - (m - 2 - (q - 1))$. Thus there exists a constant $c_3$ depending only on $q$ so that

\[
|U(pt)| \leq 2\left(\frac{m - 2}{2} - \frac{m - 2 - (q - 1)}{2}\right) \leq c_3m. \quad (50)
\]

Given $m > q + 2$ and a fixed choice $x$ different from $p, t$, we note that the columns of column sum $3$ cannot have the submatrix $q \cdot (1_20_0)$ in rows $p, x$ nor the submatrix $q \cdot (1_20_0)$ in rows $t, x$ since either would produce the configuration $q \cdot (1_20_2)$. Thus for a fixed $x \neq p, t$ (of which there are $m - 2$ choices), there are at most $2(q - 1)$ choices for $j$ such that pair $xj$ has triple overlapping $pt$ in columns of column sum $3$. A similar argument applies to the columns of column sum $m - 3$. Thus there exists a constant $c_4 = 2(q - 1)$ so that

\[
|T(pt)| \leq 2\left(\frac{(m - 2)(2(q - 1))}{2}\right) \leq c_4m \quad (51)
\]

**Lemma 4.2** There exists a constant $m_1 \geq q + 4$ so that for $m > m_1$, we have for all $\{i, j\} \subset [m]$ that $d_0(ij) + d_1(ij) \leq q - 3 + \epsilon(ij)$. 

21
Thus contained in columns of column sum 2, \( m \) submatrices as in (49) contained in \( A \) else \( A \) has the configuration \( q \cdot (1_20_2) \). There are \( 4 - \epsilon(pt) - \epsilon(rs) \) submatrices (49) contained in columns of column sum 2, \( m - 2 \) and since \( rs \notin U(pt) \cup T(pt) \) there are \((d_1(pt) + d_0(rs)) + (d_0(pt) + d_1(rs)) \) submatrices (49) in columns of column sum 3, \( m - 3 \). Thus

\[
(d_1(pt) + d_0(rs)) + (d_0(pt) + d_1(rs)) + 4 - \epsilon(pt) - \epsilon(rs) \leq 2(q - 1)
\]

Substituting \( d_0(pt) + d_1(pt) \geq q - 3 + \epsilon(pt) + 1 \) and rearranging yields

\[
d_0(rs) + d_1(rs) \leq (q - 3) - 1 + \epsilon(rs).
\]  

(52)

We wish to bound \( a_3 \) using (46). We split all pairs \( ij \) into three sets: those with \( \{i, j\} \cap \{p, t\} = \emptyset \) and \( ij \notin U(pt) \cup T(pt) \), those with \( i, j \in U(pt) \cup T(pt) \) (which forces \( \{i, j\} \cap \{p, t\} = \emptyset \)) and those with \( \{i, j\} \cap \{p, t\} \neq \emptyset \). In the first case, we use (52).

\[
\sum_{\{i, j\} \subseteq [m] \atop i, j \notin U(pt) \cup T(pt) \atop \{i, j\} \cap \{p, t\} = \emptyset} d_0(ij) + d_1(ij) \leq (q - 3) - 1 + \epsilon(ij)
\]

In the latter cases, note that \( d_0(ij) \leq q - 1 \) and \( d_1(ij) \leq q - 1 \) else, since \( m \geq q + 4 \), we would find a copy of \( q \cdot (1_20_2) \).

\[
\sum_{\{i, j\} \subseteq [m] \atop i, j \notin U(pt) \cup T(pt)} d_0(ij) + d_1(ij) \leq (c_3 + c_4)m \cdot 2(q - 1)
\]

\[
\sum_{\{i, j\} \subseteq [m] \atop \{i, j\} \cap \{p, t\} \neq \emptyset} d_0(ij) + d_1(ij) \leq 2(m - 2) \cdot 2(q - 1)
\]

Let \( c_5 \) be a constant chosen so that \( c_5 > 2(c_3 + c_4)(q - 1) \). Combining yields

\[
\sum_{ij} (d_0(ij) + d_1(ij)) \leq \sum_{\{i, j\} \subseteq [m] \atop \{i, j\} \cap \{p, t\} = \emptyset} (q - 3 - 1 + \epsilon(ij)) + c_5m.
\]

Now using (40) and substituting for \( a_2 \) using (48) and substituting for \( a_3 \) using (46) and the above inequality with the estimate that there are at most \( \binom{m-2}{2} \) choices for pairs \( ij \) with \( \{i, j\} \cap \{p, t\} = \emptyset \) \( ij \notin U(ij) \cup T(ij) \) and substituting for \( a_4 \) using (43):

\[
2 \left( \binom{m}{2} - (|E_0| + |E_1|) + \frac{q - 3 - 1}{3} \left( \binom{m - 2}{2} \right) + \frac{1}{3}(|E_0| + |E_0|) + \frac{c_5}{3}m + c_2m > \left( \binom{m}{2} \right) \frac{q + 3}{3}
\]

22
The coefficient of $m^2$ on the left side of the above inequality is only $\frac{q+2}{6}$ while on the right side is $\frac{q+3}{6}$. Thus there exists a constant $m_1$ so that for $m > m_1$, we have a contradiction proving the claim.

Let

$$Y = \{ij : d_0(ij) + d_1(ij) = q - 3 \text{ and } \epsilon(ij) = 0\}$$

**Lemma 4.3** There exists a constant $c_6$ so that

$$|Y| > \left(\frac{m}{2}\right) - c_6m. \quad (53)$$

**Proof:** We partition the $\binom{m}{2}$ pairs $ij$ into 3 parts: $Y$, $E_0 \cup E_1$ and the rest. We note that for $ij \notin Y \cup E_0 \cup E_1$, we have $\epsilon(ij) = 0$ and $d_0(ij) + d_1(ij) \leq (q - 3) - 1$ by Lemma 4.2. Thus from (46) and using Lemma 4.2

$$a_3 = \frac{1}{3} \sum_{ij} d_0(ij) + d_1(ij) \leq \frac{1}{3} ((q - 3)|Y| + ((q - 3) + 2)|E_0 \cup E_1|$$

$$+ (q - 3) - 1)\left(\binom{m}{2} - |Y| - |E_0 \cup E_1|\right)$$

Thus

$$a_3 \leq \frac{1}{3} (q - 3)\binom{m}{2} + 3|E_0 \cup E_1| - \binom{m}{2} + |Y| \quad (54)$$

Using (48), (43), (54) in (40), we have

$$2\binom{m}{2} - (|E_0| + |E_1|) + \frac{1}{3} (q - 3)\binom{m}{2} + 3|E_0 \cup E_1| + \left(|Y| - \binom{m}{2}\right) + c_2m$$

$$> \frac{q + 3}{3}\binom{m}{2}.$$}

We deduce, noting that $|E_0| + |E_1| \geq |E_0 \cup E_1|$, that $\frac{1}{3} (|Y| - \binom{m}{2}) + c_2m > 0$ and so $|Y| > \binom{m}{2} - 3c_2m$. Thus (53) holds for $c_6 = 3c_2$. ■

Form a graph $G$ whose vertex set is the rows of the matrix $A$ with edges $ij$ for those $ij \in Y$. Thus $G$ has at least $\frac{m^2}{2} - c_6m$ edges. By Turán’s Theorem [7], a graph with more than $\frac{m^2}{2} - \frac{m^2}{2(2k-1)}$ edges has a clique of $k$ vertices. Choose a constant $c_7$ so that for any choices $i, j \in [m]$

$$\left(\frac{c_7\sqrt{m} - 2(q - 1)}{2}\right) > (c_3 + c_4)m(> |\mathcal{T}(ij)| + |\mathcal{U}(ij)|),$$

$$\frac{1}{2}\left(\frac{c_7\sqrt{m}}{2}\right) - 2m > (c_3 + c_4)m(> |\mathcal{T}(ij)| + |\mathcal{U}(ij)|),$$

23
\[
\left( \frac{c \sqrt{m}}{2} \right) > (c_3 + c_4)m \left( > |T(ij)| + |U(ij)| \right).
\] (55)

Then by Turán’s Theorem, there exists a \( M > m_0, m_1 \) (\( m_0 \) is from Lemma 4.1 and \( m_1 \) is from Lemma 4.2) so that for \( m > M \), graph \( G \) has a clique of \( c \sqrt{m} \) vertices.

Let \( B \) denote the set of the rows in this clique. Hence for every \( i, j \in B \) we have \( d_1(ij) + d_0(ij) = q - 3 \) and \( \epsilon(ij) = 0 \). Let \( A_4 \) denote the columns of \( A \) of column sum \( 4, 5, \ldots, m - 5 \) or \( m - 4 \). Let \( A_4(B) \) be the submatrix of \( A_4 \) of the rows indexed by \( B \).

**Lemma 4.4** Assume \( m > M \). Then \( A_4(B) \) has no configuration \( 1_20_2 \).

**Proof:** Assume there are rows \( i, j, k, l \in B \) and a column \( \alpha \) of \( A_4 \) with 0’s in rows \( i, j \) and 1’s in rows \( k, l \). Without loss of generality, we may assume that there are more 1’s than 0’s in \( \alpha \) in the rows of \( B \) so that the number of 1’s in the rows of \( B \) is more than \( c \sqrt{m}/2 \). Thus by the third inequality in (55), we can find a pair \( gh \) of rows with \( g, h \in B \), so that \( \alpha \) has 1’s in row \( g, h \) and \( gh \notin T(ij) \cup U(ij) \). We may now argue that for our choice of \( i, j, g, h \), we have \( (d_1(ij) + d_0(gh)) + (d_1(gh) + d_0(ij)) + 4 - \epsilon(ij) - \epsilon(gh) = 2(q - 1) \) submatrices

\[
\begin{bmatrix}
i & 1 \\
j & 1 \\
g & 0 \\
h & 0 \\
\end{bmatrix}
\quad \text{or} \quad
\begin{bmatrix}
i & 0 \\
j & 0 \\
g & 1 \\
h & 1 \\
\end{bmatrix}
\] (56)

in \( A \) in columns of column sum \( 2, 3, m - 3, m - 2 \). With another such submatrix in \( \alpha \) in \( A_4 \), we have \( 2(q - 1) + 1 \) such submatrices, for our chosen quadruple \( i, j, g, h \) and so \( A \) has the configuration \( q \cdot 1_20_2 \), a contradiction. \( \blacksquare \)

**Lemma 4.5** Assume \( m > M \). Then the inequality (9) holds.

**Proof:** Assume \( m > M \) and (40). Using Lemma 4.4 the columns of \( A_4 \) can be partitioned into two parts: \( Z \) the columns that have at most one 1 in the rows \( B \) and \( J \) the columns that have at most one 0 in the rows of section \( B \).

For each pair \( i, j \in [m] \setminus B \), let \( t(ij) \) count the sum of the number of columns in \( Z \) with 1’s in both rows \( i, j \) as well as the number of columns in \( J \) with 0’s in both rows \( i, j \). For all other pairs \( ij \), let \( t(ij) = 0 \). Given the column sums in \( A_4 \), every column in \( Z \) has at least three 1’s in rows \( [m] \setminus B \) and every column in \( J \) has at least three 0’s in rows \( [m] \setminus B \). We have

\[
\sum_{ij} t(ij) \geq 3a_4
\] (57)

Moreover, we find that \( t(ij) \leq 2(q - 1) \): Given a choice for \( i, j \), if we have \( q \) columns in \( Z \) with 1’s in rows \( i, j \) then there are at most \( q \) rows of \( B \) containing 1’s for these \( q \) columns (since each column of \( Z \) has at most one 1 in the rows of \( B \)). But then if we choose two rows of \( B \) from the remaining \( \geq |B| - q \) rows in conjunction with \( i, j \) then
we have a copy of the configuration \( q \cdot (1_2 0_2) \). Similarly, there cannot be \( q \) columns of \( J \) with 0’s on rows \( i, j \). We conclude \( t(ij) \leq 2(q - 1) \).

For a given pair \( i, j \in [m] \setminus B \), consider the \( t(ij) \) columns contributing to \( t(ij) \). By the first inequality in (55), we can find a pair of rows \( gh \) \((g, h \in B)\) so that \( gh \notin T(ij) \cup U(ij) \) and in addition \( g, h \) are not chosen from the up to 2\((q - 1) \) rows of \( B \) which are given as follows: the \( \leq q - 1 \) rows of \( B \) which have 1’s in the columns of \( Z \) having 1’s in both rows \( i, j \) and the \( \leq q - 1 \) rows of \( B \) which have 0’s in the columns of \( J \) having 0’s in both rows \( i, j \). Thus if \( \alpha \) is a column of \( Z \) with 1’s in rows \( i, j \) then \( \alpha \) has 0’s in rows \( g, h \) and if \( \alpha \) is a column of \( J \) with 0’s in rows \( i, j \) then \( \alpha \) has 1’s in rows \( g, h \). There will be \( 4 - \epsilon(ij) - \epsilon(gh) \) submatrices as in (56) in the columns of column sum 2 or \( m - 2 \).

Neither pair \( ij \) has triple overlapping \( gh \) nor pair \( gh \) has triple overlapping \( ij \) and so there will be \((d_1(ij) + d_0(gh)) + (d_0(ij) + d_1(gh))\) submatrices as in (56) in the columns of column sum 3 or \( m - 3 \). By our choice of \( g, h \), a column \( \alpha \) in \( Z \) with 1’s in rows \( i, j \) will have 0’s on rows \( g, h \). A column \( \beta \) in \( J \) with 0’s in rows \( i, j \) will have 1’s on rows \( g, h \). Thus in \( A_1 \) we can find \( t(ij) \) submatrices as in (56). In the matrix \( A \), an ordered quadruple of rows \( i, j, g, h \) has at most 2\((q - 1) \) submatrices as given in (56) else \( A \) would have the configuration \( q \cdot (1_2 0_2) \). Thus

\[
(d_1(ij) + d_0(gh)) + (d_0(ij) + d_1(gh)) + 4 - \epsilon(ij) - \epsilon(gh) + t(ij) \leq 2(q - 1).
\]

Substituting \( d_0(gh) + d_1(gh) = q - 3 \) and \( \epsilon(gh) = 0 \) and rearranging we have

\[
d_1(ij) + d_0(ij) \leq (q - 3) + \epsilon(ij) - t(ij) . \tag{58}
\]

This inequality is true for other \( i, j \) using Lemma 4.2 when \( t(ij) = 0 \). Thus

\[
\sum_{ij} (d_0(ij) + d_1(ij)) \leq \sum_{ij} (q - 3 + \epsilon(ij) - t(ij)) \tag{59}
\]

Taking (40) with \( a_2 \) from (48) and with \( a_3 \) from (46) using (58) we obtain

\[
2 \binom{m}{2} - |E_0| - |E_1| + \frac{1}{3} \sum_{ij} (q - 3 + \epsilon(ij) - t(ij)) + a_4 > \frac{q + 3}{3} \binom{m}{2}
\]

Simplifying and using \( \sum_{ij} \epsilon(ij) = |E_0| + |E_1| \) and (57) we obtain

\[
-\frac{2}{3} (|E_0| + |E_1|) > 0
\]

which is a contradiction (even for \(|E_0| + |E_1| = 0 \)). This establishes (9).

**Proof of Proposition 1.8** Lemma 4.5 establishes most of Proposition 1.8 but we are also interested in cases when the bound is achieved. Assume \( m > M \) and \( m \equiv 1, 3 (mod 6) \). We now consider an \( m \)-rowed simple matrix \( A \) which has no configuration
\( q \cdot (1_20_2) \) and with \( \binom{m}{0} + \binom{m}{1} + \frac{q+3}{3}\binom{m}{2} + \binom{m}{m-1} + \binom{m}{m} \) columns. One repeats the previous lemmas and arguments replacing the inequality (40) with the equation

\[
a_2 + a_3 + a_4 = \frac{q+3}{3}\binom{m}{2}.
\]

We wish to show \( a_2 = 2\binom{m}{2}, a_4 = 0, a_3 = \frac{2-q}{3}\binom{m}{2} \) and there exists positive integers \( a, b, a+b = q-3 \) so that for all pairs \( ij \), \( d_0(ij) = a \) and \( d_1(ij) = b \). Now Lemma 4.1 holds with (40) as an equality. We deduce the same bounds for \( U \).

\[|U| \leq \frac{2}{3}|B| - 2|B|\]

Now using the second inequality of (55) with \( m \) we deduce that \( B \) is merely the equality of (60). Lemma 4.3 holds and we can choose \( B \) as large as possible but at least satisfying the inequalities (55). Lemma 4.4 continues to hold.

Assume that not all pairs \( pt \) with \( p, t \in B \) have the same value for \( d_0(pt) \). We can choose \( ij \) with \( i, j \in B \) so that at least \( \frac{1}{2}\binom{|B|}{2} \) pairs \( pt \) of \( B \) have \( d_0(ij) \neq d_0(pt) \). Then the number of pairs \( pt \) of \( B \) in \( \binom{|B|}{2} \) with \( d_0(ij) \neq d_0(pt) \) is at least \( \frac{1}{2}\binom{|B|}{2} - 2|B| \). Now using the second inequality of (55) with \( |U(ij)| + |T(ij)| \leq (c_3+c_4)m \) and \( |B| \leq m \), we can find a pair \( k, l \in B \setminus \{i, j\} \) with \( d_0(ij) \neq d_0(kl), kl \notin U(ij) \cup T(ij) \). By definition of \( B \),

\[ \sum_{ij} x_{ij} = q - 3, \quad d_0(kl) + d_1(kl) = q - 3. \]

We may assume without loss of generality that \( d_0(kl) < d_0(ij), d_1(kl) > d_1(ij) \) and then

\[ d_0(ij) + d_1(kl) \geq q - 2. \]

We also have \( \epsilon(ij) = \epsilon(kl) = 0 \). Then \( A \) has a column of column sum 2 and a column of column sum \( m-2 \) both with 1’s in rows \( k, l \) and 0’s in rows \( i, j \). Also we have \( d_0(ij) + d_1(kl) \) columns with 1’s in rows \( k, l \) and 0’s in rows \( i, j \). But then \( A \) has \( q \cdot (1_20_2) \), a contradiction. We conclude that all pairs \( pt \) with \( p, t \in B \) have the same value for \( d_0(pt) \).

We follow our proof of Lemma 4.5 using (60) and deduce that \( E_0 \cup E_1 = \emptyset \) and so \( a_2 = 2\binom{m}{2} \). Also we deduce that

\[ \sum_{ij} t(ij) = 3a_4 \]

and as a result we can deduce that any column \( \alpha \) in \( A_4 \) either has column sum 4 with exactly one 1 in a row of \( B \) or has column sum \( m-4 \) with exactly one 0 in a row of \( B \).

Assume \( a_4 > 0 \) and consider \( \alpha \) in \( A_4 \), say with column sum 4 and with 1’s in rows \( i, j, k, l \) where \( i \in B \) and \( j, k, l \in \{1, 2, \ldots, m\} \setminus B \). Choose \( r, s \in B \setminus i \) so that \( d_0(rs) + d_1(rs) = q - 3 \) and with \( rs \notin T(ij) \cup U(ij) \) (using first inequality of (55)). Column \( \alpha \) has 1’s in rows \( i, j \) and 0’s in row \( r, s \). Using \( E_0 \cup E_1 = \emptyset \), we deduce that \( d_1(ij) + d_0(rs) \leq q - 3 - 1 \) and \( d_0(ij) + d_1(rs) \leq q - 3 \) else if either inequality is violated we create \( q \cdot (1_20_2) \). We deduce \( d_0(ij) + d_1(ij) \leq (q-3) - 1 \). This yields a slight variant of (59):

\[ \sum_{ij} (d_0(ij) + d_1(ij)) \leq \sum_{ij} ((q-3 + \epsilon(ij) - t(ij)) - 1. \]
The extra ‘-1’ is sufficient to obtain a contradiction when we substitute for \(a_2, a_3, a_4\) in (60). We then deduce \(a_4 = 0\).

With \(a_4 = 0\) and \(a_2 = 2{m \choose 2}\), we deduce \(a_3 = \frac{2-3}{3}{m \choose 2}\) using (60). Given that \(\epsilon(ij) = 0\) for all \(ij\) and using Lemma 4.2, we deduce \(d_0(ij) + d_1(ij) = q - 3\) for all pairs \(ij\) and so \(B = \{1, 2, \ldots, m\}\). Our above arguments tell us \(d_0(pt)\) is the same for every choice \(p, t \in B\), allowing us to conclude that there exists positive integers \(a, b, a + b = q - 3\) so that for all pairs \(ij\), \(d_0(ij) = a\) and \(d_1(ij) = b\). From this we can readily conclude that the columns of column sum 3 correspond to a 2-design \(S_a(2, 3, m)\) and the columns of column sum \(m - 3\) correspond to the \((0,1)\)-complement of a 2-design \(S_b(2, 3, m)\).

References

[1] R.P. Anstee, R. Ferguson, A. Sali, Small Forbidden Configurations II, *Electronic J. Combin.* 8(2001), R4 (25pp)

[2] R.P. Anstee, N. Kamoosi, Small Forbidden Configurations III, *Electronic J. Combin.* 14(2007), R79 (34pp)

[3] Michel Dehon, On the existence of 2-designs \(S_\lambda(2, 3, v)\) without repeated blocks, *Discrete Math.* 43(1983), 155-171.

[4] Z. Füredi, 1983, private communication.

[5] N. Sauer, On the density of families of sets, *J. Combin. Th. Ser A* 13(1972), 145-147.

[6] S. Shelah, A combinatorial problem: Stability and order for models and theories in infinitary languages, *Pac. J. Math.* 4(1972), 247-261.

[7] P. Turán, Eine Extremalaufgabe aus der Graphentheorie, *Math. Fiz Lapok.* 48(1941), 436-452.

[8] V.N. Vapnik and A.Ya. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, *Th. Prob. and Applics.* 16(1971), 264-280.