



## **Optimal partial clique edge covering guided by potential energy minimization**

Downloaded from: <https://research.chalmers.se>, 2021-01-17 04:26 UTC

Citation for the original published paper (version of record):

Damaschke, P. (2019)

Optimal partial clique edge covering guided by potential energy minimization

Optimization Letters, 13(7): 1469-1481

<http://dx.doi.org/10.1007/s11590-019-01469-y>

N.B. When citing this work, cite the original published paper.



# Optimal partial clique edge covering guided by potential energy minimization

Peter Damaschke<sup>1</sup>

Received: 27 August 2018 / Accepted: 19 August 2019 / Published online: 22 August 2019  
© The Author(s) 2019

## Abstract

For given integers  $k, n, r$  we aim at families of  $k$  sub-cliques called blocks, of a clique with  $n$  vertices, such that every block has  $r$  vertices, and the blocks together cover a maximum number of edges. We demonstrate a combinatorial optimization method that generates such optimal partial clique edge coverings. It takes certain packages of columns (corresponding to vertices) in the incidence matrix of the blocks, considers the number of uncovered edges as an energy term that has to be minimized by transforming these packages. As a proof of concept we can completely solve the above maximization problem in the case of  $k \leq 4$  blocks and obtain optimal coverings for all integers  $n$  and  $r$  with  $r/n \geq 5/9$ . This generalizes known results for total coverings to partial coverings. The method as such is not restricted to  $k \leq 4$  blocks, but a challenge for further research (also on total coverings) is to limit the case distinctions when more blocks are involved.

**Keywords** Induced graph decomposition · Clique edge covering · Reaction network · Potential energy

## 1 Introduction

Covering designs are a classic subject in extremal combinatorics. Applications include the generation of efficient test cases that cover all (or many) conditions, the design of fault-tolerant systems, and collision avoidance. The general problem is: In a set  $V$  with  $|V| = n$ , place  $k$  subsets of size  $r$  called blocks, so as to cover a maximum number of subsets  $T \subset V$  with  $|T| = t$  (where  $T$  is said to be *covered* if  $T$  is a subset of a block). Unlike the trivial case  $t = 1$ , case  $t = 2$  is already subtle. It can be phrased as a graph problem:

---

✉ Peter Damaschke  
ptr@chalmers.se

<sup>1</sup> Department of Computer Science and Engineering, Chalmers University, 41296 Göteborg, Sweden

**Definition 1** Let  $n$  and  $r$  be any integers with  $0 < r < n$ . A *host clique* is a clique (complete graph) with  $n$  vertices. A *block* is any subgraph of the host clique, induced by a subset of  $r$  vertices (and hence a clique with  $r$  vertices).

As usual, let  $K_n$  denote a clique with  $n$  vertices. Notationally we do not always distinguish between several cliques on different vertex sets but of the same size, e.g., we simply speak of blocks  $K_r$ , if this causes no confusion.

**Definition 2** An optimal  $(n, r)$  *partial clique edge covering with  $k$  blocks* is a family of  $k$  blocks  $K_r$  that cover the largest possible number  $g$  of edges of the host clique  $K_n$ . If  $g = \binom{n}{2}$ , we speak of an  $(n, r)$  *clique edge covering*.

Clique edge coverings are also known as  $(n, r, 2)$ -covering designs or  $r$ -uniform clique coverings. They have been extensively studied, see, e.g., [1]. We have chosen the term *clique edge covering* for brevity, and to avoid confusion with the different *edge clique covers* in general graphs.

The smallest numbers  $k$  of blocks required for  $(n, r)$  clique edge coverings are completely known for  $r/n \geq 4/13$  and arbitrary  $n$ , thanks to [7,9]. (This is also mentioned in [1].) We express this knowledge as follows.

**Definition 3** Let  $\gamma_k$  denote the smallest possible ratio  $r/n$  such that an  $(n, r)$  clique edge covering with  $k$  blocks exists.

It is known that a clique edge covering always exists if  $r/n \geq \gamma_k$ . A counting argument trivially yields  $\gamma_k \geq 1/\sqrt{k}$ . However, in general one cannot exactly pack the edge sets of several  $K_r$  into  $K_n$ , hence the  $\gamma_k$  are larger. Precisely known values are  $\gamma_2 = 1$ ,  $\gamma_3 = 2/3$ ,  $\gamma_4 = 3/5$ ,  $\gamma_5 = 5/9$ ,  $\gamma_6 = 1/2$ . See, e.g., Theorem 8.21 in Chapter IV of [4]; here we have only adjusted the notation. Minimum  $k$  for all  $n \leq 32$  and  $r \leq 16$  can be found in [6].

In the present paper we consider the more general optimal *partial clique edge coverings*, and we introduce a method for constructing them. For fixed  $k$ , it can cope with arbitrarily large  $n$ . In [5] we had considered a continuous counterpart of the problem. The new combinatorial approach is able to compute exact edge numbers for the discrete problem, moreover, it gives an intuitive understanding of the structure of optimal coverings: The main idea is to interpret the vertices as columns of the  $k \times n$  incidence matrix, and the number of uncovered edges as potential energy between pairs of them. Then we transform packages of columns so as to decrease this energy, leading to an optimal solution composed of special packages of minimum energy. Amazingly, the potential energy view of graph problems has recently been proposed in [10], and here the analogy turns out to be directly fruitful for solving a concrete problem. As the idea looks natural and general, it may also apply to the construction of other optimal combinatorial designs.

Recall that optimal total clique edge coverings are known for all  $n, r$  with  $r/n \geq 4/13$ . Our work generalizes this type of results to partial coverings, and we manage all instances with  $k \leq 4$  and  $r/n \geq 5/9$ , as a proof of concept. But the method as such does not stop there. One might be surprised how complicated already the case of  $k = 4$  blocks is, however, even for the special case of total coverings the difficulty increases drastically as  $k$  grows.

The matter is also related to other known structures: A  $K_r$ -decomposition of  $K_n$  is an  $(n, r)$  clique edge covering whose blocks cover every edge *exactly* once. More generally, an *induced  $H$ -decomposition* of a graph  $G$  consists of induced subgraphs  $H_1, \dots, H_k$  of  $G$  such that every edge of  $G$  is in exactly one  $H_i$ , and all  $H_i$  are isomorphic to a fixed graph  $H$ . Elegant necessary and sufficient conditions for induced  $H$ -decompositions of  $K_n$  are given in [11]. Various cases of  $H$  and general  $G$  are studied in [2,3,8].

## 2 Further notation and preliminaries

**Definition 4** For a given number  $k$  of blocks we define:

- The *incidence matrix* of a family of  $k$  blocks in  $K_n$  is a binary  $k \times n$  matrix with a row for every block and a column for every vertex. A matrix entry equals 1 if the vertex belongs to the block, and 0 otherwise.
- For any set  $I \subseteq \{1, \dots, k\}$  of row indices, we use the shorthand “a column  $I$ ” to refer to any column whose set of row indices with matrix entry 1 is exactly  $I$ . We also write any column as the set  $I$  of the row indices where the matrix entry is 1.
- A *row sum* is the number of 1s in a row, hence it equals the size of the block represented by that row. A *column sum* is the number of 1s in a column. The column sum of a column  $I$  is denoted  $|I|$ .

That is, for convenience we treat a column both as a bit vector and as the set of positions of entries 1 interchangeably. Also, since the order of vertices is arbitrary, we need not distinguish between incidence matrices whose columns are permuted.

**Example 1** The matrix below is the incidence matrix of 3 blocks  $\{u, v, y\}, \{v, w, y\}, \{w, x, y\}$  in a  $K_5$  with vertex set  $\{u, v, w, x, y\}$  (ordered from left to right). All row sums are 3, and the columns can be written as  $\{1\}, \{1, 2\}, \{2, 3\}, \{3\}, \{1, 2, 3\}$ .

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

The covered edges in this example are  $uv, uy, vw, vy, wx, wy, xy$ . As we shall see later in Theorem 1, this example is not an optimal covering, since we can cover 8 edges by 3 blocks of size 3. In general, the connection to our problem is given by the following obvious fact:

**Proposition 1** *An edge is covered if and only if the two columns representing its two vertices intersect, i.e., they are columns  $I, J$  with  $I \cap J \neq \emptyset$ .*

**Proof** By Definition 4, a column  $I$  represents a vertex of  $K_n$  that belongs to exactly those blocks whose indices are in  $I$ . Thus, for any two vertices  $p$  and  $q$ , the following statements are equivalent: the edge  $pq$  is covered by some block; some block contains both  $p$  and  $q$ ; some row has entries 1 in the columns of  $p$  and  $q$ ; the columns of  $p$  and  $q$  (as sets of positions of the 1s) intersect. □

In [5] we found a property similar to the following one for the continuous counterpart of our problem. But the present lemma does not follow immediately, as there might be “discretization effects”.

**Lemma 1** *For any integers  $k, n, r$  there exist  $k$  blocks of  $r$  vertices that cover a maximum number  $g$  of edges of  $K_n$  and obey the following property: For any two sets  $A, C$  of row indices such that  $A \subset C \subseteq \{1, \dots, k\}$  and  $|C| - |A| \geq 2$ , the incidence matrix does not contain both a column  $A$  and a column  $C$ . In particular, no column with only 0s exists if  $kr > n$ .*

**Proof** We show that, in any incidence matrix with  $k$  rows and with row sums  $r$ , we can get rid of the mentioned pairs of columns, by transformations that neither change the row sums nor decrease the number of covered edges.

Assume that  $(A, C)$  is any pair of columns as specified above. Consider two rows corresponding to some indices in  $C \setminus A$ . The crossing of the mentioned two columns and rows is a  $2 \times 2$  submatrix with two rows  $(0, 1)$ . Note that none of the other rows in the pair of columns  $(A, C)$  is  $(1, 0)$ , since  $A \subset C$ . We replace  $(A, C)$  with a new pair of columns  $(A', C')$ , by turning one row  $(0, 1)$  into  $(1, 0)$ :

$$\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Obviously, the row sums are preserved. Consider any further column  $B$ . If  $B$  intersects both  $A$  and  $C$ , then  $B$  also intersects both  $A'$  and  $C'$ . If  $B$  intersects only  $C$  but not  $A$ , then  $B$  still intersects at least one of  $A'$  and  $C'$ . From these two statements it follows that the number of covered edges does not decrease. Also note that  $|A'|$  and  $|C'|$  are strictly between  $|A|$  and  $|C|$ .

We repeat this step as long as two columns  $A$  and  $C$  as above exist. Specifically, we always pick a column  $C$  with maximum  $|C|$ . This decreases the number of columns with maximum column sum, and eventually it decreases the maximum column sum itself. Thus, the process does not run into a cycle and terminates with an incidence matrix satisfying the claimed property.

The last assertion follows since, by the pigeonhole principle, for  $kr > n$  some column must have at least two 1s.  $\square$

Henceforth it suffices to consider partial clique edge coverings that satisfy the property in Lemma 1. A first consequence are optimal coverings with  $k = 2$  blocks  $K_r$ : Since columns with two 0s or two 1s cannot coexist, the two blocks are either disjoint (if  $r \leq n/2$ ) or they together contain all  $n$  vertices (if  $r > n/2$ ).

### 3 Outline of the method

Next we introduce a novel concept that will allow us to structurally characterize optimal partial clique edge coverings.

**Definition 5** For incidence matrices with  $k$  rows we define:

- A *packet* with  $c$  columns is any binary  $k \times c$  matrix where all  $k$  row sums are equal. The *density* of a packet is the row sum divided by  $c$ , or equivalently, the number of 1s divided by  $kc$ .
- A *partitioning* of an incidence matrix divides the multiset of its columns into packets. A partitioning may contain arbitrarily many identical copies of every packet, however, for certain packets we allow only some fixed maximum number of copies. We refer to the latter packets as the *remainder*.
- The *energy* of an incidence matrix is the number of uncovered edges, i.e., of pairs of disjoint columns. The energy  $E(P, Q)$  between two packets  $P$  and  $Q$  is the number of pairs of disjoint columns, one being in  $P$  and one being in  $Q$ .
- By  $c_1 P_1 + \dots + c_l P_l$  we denote a submatrix consisting of  $c_i$  identical copies of packet  $P_i$ , for  $i = 1, \dots, l$ , where the sets of column indices of all these  $c_1 + \dots + c_l$  packets are pairwise disjoint. The expression  $d_1 Q_1 + \dots + d_m Q_m$  is similarly defined, and we assume that both submatrices have the same total number of columns and the same total row sums. The *interaction*  $c_1 P_1 + \dots + c_l P_l \rightarrow d_1 Q_1 + \dots + d_m Q_m$  replaces the submatrix  $c_1 P_1 + \dots + c_l P_l$  with the submatrix  $d_1 Q_1 + \dots + d_m Q_m$ . An interaction done within an incidence matrix is *valid* if it does not increase the energy of the incidence matrix.

The energy within a packet  $P$  is obviously  $E(P, P)/2$ . (If we take two copies of  $P$ , then every disjoint pair is counted twice, moreover, no column is disjoint to itself.) Also note that  $E(P, Q) = E(Q, P)$ .

**Example 2** The matrix below shows an incidence matrix (with  $k = 3, n = 8, r = 5$ , hence with density  $5/8$ ) partitioned into three packets two of which are identical. The energy is 0 within and between the first two packets, 1 within the last packet, and 1 between the last packet and each of the first two packets, resulting in the total energy 3.

$$\left( \begin{array}{ccc|ccc|cc} 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \end{array} \right)$$

In the rest of the paper we prove, for  $k \leq 4$  blocks, the existence of optimal clique edge coverings whose incidence matrices are composed of only two types of packets, subject to remainders. This finally allows to compute the amounts of these packets (and hence optimal coverings) uniquely from the given sizes  $n$  and  $r$ , by integer linear equations. The structure of the existence proofs is as follows. We start from an arbitrary incidence matrix that has row sums  $r$  and satisfies the property from Lemma 1, and show that it can be divided into a small number of different simple packets with many symmetries. With these packets we perform interactions reducing the energy, until the matrix is transformed into one consisting of the claimed packets. Every interaction replaces some packets with others, thereby preserving the number  $n$  of columns and the row sums  $r$ , that is, it changes the blocks and re-assigns some vertices but preserves the sizes of blocks.

The technical contribution is this proof method. We remark that it is not merely local search. Besides reducing the energy among the packets in an interaction, one must count in the energy between these packets and the rest of the matrix. It is especially

the limited number of different packets and their symmetries that will make it rather convenient to compute these energies.

We conclude this section with some simple but useful observations.

**Lemma 2** *For any column  $A$  with  $|A| = 1$ , any submatrix  $M$  of other columns, and any interaction applied to  $M$  that results in a submatrix  $M'$ , we have that  $E(M', \{A\}) = E(M, \{A\})$ .*

**Proof** Since  $E(M, \{A\})$  is the number of columns in  $M$  not intersecting  $A$ , it equals the number of 0s in  $M$  in the single row where  $A$  has a 1. This number is the same in  $M'$ , since an interaction preserves the row sums.  $\square$

A sequence of interactions may run into a cycle and never terminate. This cannot happen if they strictly decrease the energy. We give another simple sufficient condition:

**Lemma 3** *Let  $Y$  be some finite set of indices. Consider a set of interactions  $M_i \rightarrow M'_i$  ( $i \in Y$ ) that can be applied to an incidence matrix, each one transforming a submatrix (subset of columns) identical to  $M_i$  into a submatrix  $M'_i$ . Suppose that every  $M_i$  contains some column that does not appear in any  $M'_j$  ( $j \in Y$ ). Then any sequence of such interactions is finite. The conclusion also holds if all matrices  $M_i$  ( $i \in Y$ ) except one satisfy the above condition.*

**Proof** Trivially, an interaction that consumes some column that is not produced elsewhere can be applied only finitely often. If one  $M_i$  does not contain such a column, then still all other interactions can be applied only finitely often. But an infinite sequence of a single interaction is not possible either.  $\square$

**Lemma 4** *Suppose that an interaction turns a submatrix  $M$  into  $M'$ , and the rest of the incidence matrix is divided into submatrices  $P$ . If  $E(M, M) \geq E(M', M')$  and  $E(M, P) \geq E(M', P)$  for all  $P$ , then the interaction is valid.*

**Proof** This follows immediately from the definition of energy and the fact that the packets  $P$  are not changed by the interaction.  $\square$

## 4 Partial covering of a clique by three blocks

First we demonstrate the principle for  $k = 3$ . Since  $\gamma_3 = 2/3$ , three  $K_r$  can cover all edges of  $K_n$  if and only if  $r/n \geq 2/3$ . Now we will also yield optimal partial clique edge coverings with three blocks, for any  $n$  and  $r$ . Since the case  $r/n \leq 1/3$  is trivial, we assume  $r/n > 1/3$ .

**Theorem 1** *For  $1/3 < r/n < 2/3$ , there exist three  $K_r$  that cover a maximum number of edges of  $K_n$ , of the following form: Their incidence matrix can be partitioned into packets of the types<sup>1</sup>*

<sup>1</sup> The names of packets are arbitrary, but we tried to choose suggestive names. Think of the columns as vertices of a graph, and of the rows with two 1s as edges.

$$clique = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \quad anticlique = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and a remainder which is either empty or is one of the following packets (up to permutations of rows):

$$anti - edge = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \quad path = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

**Proof** Remember that it suffices to consider incidence matrices that satisfy the property in Lemma 1. If all three entries in some column are 1s, then all other columns have at least two 1s. But then any two columns intersect, thus all edges are covered. This in turn implies  $r/n \geq 2/3$  (by the known result  $\gamma_3 = 2/3$ ), contradicting our assumption  $r/n < 2/3$ . Hence only columns with one or two 1s are present, which are at most 6 different columns.

We partition our incidence matrix into packets. First we form cliques as long as possible. That is, we repeatedly take three columns  $\{1, 2\}$ ,  $\{1, 3\}$ ,  $\{2, 3\}$  and group them to a clique. Outside these cliques there remain at most two types of columns with two 1s, say,  $c$  further columns  $\{1, 2\}$  and  $c'$  further columns  $\{2, 3\}$ , but no further columns  $\{1, 3\}$ . (Other cases are symmetric.) Due to the equal row sums, there also exist  $c$  columns  $\{3\}$  and  $c'$  columns  $\{1\}$ .

Assume that  $c \geq 2$  (or  $c' \geq 2$ , but this case is symmetric). That is, we have two further columns  $\{1, 2\}$  and two further columns  $\{3\}$ . We take these four columns and apply the following interaction to them:

$$\begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

Its validity is seen by Lemma 4 as follows. Let  $M$  denote the submatrix displayed above, and  $N$  the rest of the incidence matrix. Note that  $E(M, N)$  strictly decreases. We claim that  $E(M, N)$  does not increase either: It suffices to verify that the energies between  $M$  and the packets (cliques) and further columns in  $N$  do not increase. For the cliques and for the columns  $\{1, 2\}$  and  $\{2, 3\}$  in  $N$  this is obvious, columns  $\{1, 3\}$  do not exist in  $N$ , and for columns in  $N$  with a single 1 this follows from Lemma 2.

We repeat the above steps (build further cliques and perform interactions) until  $c < 2$  and  $c' < 2$ . Hence there remains at most one anti-edge or path outside the cliques. Due to the equal row sums again, the numbers of further columns  $\{1\}$ ,  $\{2\}$ ,  $\{3\}$  are equal, hence we can group them to anticliques. Altogether, we always obtain one of the claimed partitionings. □

Knowing the structure of an optimal clique edge covering from Theorem 1, it is now straightforward to compute one, for a given  $n$  and  $r$ . The “algorithm” for that is described as follows. First we compute the number  $c$  and  $a$  of cliques and anticliques, respectively. Since these packets have 3 columns, Theorem 1 yields the following case distinction and formulas for calculating  $c$  and  $a$ .



- If  $n = 0 \pmod 3$  then the remainder is empty. Hence  $3c + 3a = n$ ,  $2c + a = r$ , which implies  $c = r - n/3$  and  $a = 2n/3 - r$ .
- If  $n = 1 \pmod 3$  then the remainder is a path. Hence  $3c + 3a = n - 4$ ,  $2c + a = r - 2$ , which yields  $c = r - n/3 - 2/3$  and  $a = 2n/3 - r - 2/3$ .
- If  $n = 2 \pmod 3$  then the remainder is an anti-edge. Hence  $3c + 3a = n - 2$ ,  $2c + a = r - 1$ , which yields  $c = r - n/3 - 1/3$  and  $a = 2n/3 - r - 1/3$ .

Finally, in either case we simply take  $c$  cliques and  $a$  anticliques and the respective remainder, and stack them together to an incidence matrix, which is our optimal clique edge cover.

Example 2 (in Sect. 3) actually shows an optimal incidence matrix for  $n = 8$  and  $r = 5$ , with an anti-edge and  $c = 2$  cliques, whereas  $a = 0$ .

It may be interesting to observe the number of covered edges. For example, for  $n = 0 \pmod 3$ , this number increases by exactly  $n$  whenever  $r$  is raised by 1. This is shown as follows. One anticlique is turned into a clique, and we have  $a - 1$  other anticliques and  $c$  other cliques. From the partitioning we see directly that the number of covered edges increases by  $3 + 3(a - 1) + 3c = 3a + 3c = n$ .

## 5 Partial covering of a clique by four blocks

Until now we can compute optimal partial clique edge coverings for all  $n, r$  with  $r/n \geq 3/5$ : Case  $k \leq 2$  was simple. If  $k = 3$  and  $r/n \geq 2/3$ , then all edges of  $K_n$  can be covered, due to  $\gamma_3 = 2/3$ . If  $k = 3$  and  $r/n < 2/3$ , then we use Theorem 1. If  $k \geq 4$  and  $r/n \geq 3/5$ , then all edges of  $K_n$  can be covered, due to  $\gamma_4 = 3/5$ .

Now we turn to the case  $k = 4$  and  $r/n < 3/5$ , which is already intricate and shows the power of the packet approach. We continue on the lines of Theorem 1, now for the range  $1/2 < r/n < 3/5$ . Since  $\gamma_5 = 5/9$ , the following Theorem 2 enables us to compute optimal families of blocks for all  $n, r$  with  $r/n \geq 5/9$ . (Namely, for  $k \geq 5$  blocks it is known that all edges can be covered, and for  $k \leq 4$  blocks, the maximum number of covered edges is given by our results.) The final construction of an optimal covering is completely analogous to the case  $k = 3$ , only with different packets and numbers. The details are therefore omitted. Again, for each of the possible remainders,  $n$  and  $r$  uniquely determine the amount of cliques and stars, via two integer linear equations.

We come to the existence theorem. The basic ideas are the same as in Theorem 1, but the details are much more complex. The reader may first skip some case distinctions and verifications without losing track of the overall structure of the proof.

**Theorem 2** *For  $1/2 < r/n < 3/5$ , there exist four  $K_r$  that cover a maximum number of edges of  $K_n$ , of the following form: Their incidence matrix can be partitioned into cliques and stars, and one of these remainders:*

- at most two cycles and at most one pair,
- at most two diagonals.

*The mentioned packets are defined below (with the understanding that rows may be permuted simultaneously in all packets).*

$$\begin{aligned}
 \text{clique} &= \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix} & \text{star} &= \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \\
 \text{cycle} &= \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} & \text{pair} &= \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} & \text{diagonal} &= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}
 \end{aligned}$$

In the rest of this section we prove Theorem 2. First we define some other packets that will appear only in intermediate steps:

$$\text{hyperclique} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix} \quad \text{corner} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

**Assumption** Every column in the incidence matrix has at least two 1s. (Later we must drop this extra assumption and include also columns with one 1.)

Again we start from any incidence matrix that satisfies the property in Lemma 1, and we first build cliques as long as possible, from all six different columns with two 1s. After that, some of the columns with two 1s is no longer available outside the cliques. Specifically, we can assume (by permuting rows if necessary) that no further column {2, 4} exists.

Next we build cycles as long as possible, from the remaining columns. By definition they do not contain any columns {2, 4} and {1, 3}. After this phase, also some column from the cycle is no longer available outside the packets. Specifically, we can assume (by permuting rows if necessary) that no further column {2, 3} exists. *From now on the order of rows remains fixed.*

Next we also build stars as long as possible, from the remaining columns.

After this grouping of columns into cliques, cycles, and stars, we perform the following interactions, as long as possible and in any order.

$$\begin{aligned}
 \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} &\rightarrow \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} &\rightarrow \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \\
 \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix} &\rightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} & \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} &\rightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}
 \end{aligned}$$

This set fulfills the condition of Lemma 3, hence any sequence of these interactions terminates. To show that every interaction  $M \rightarrow M'$  is valid, we apply Lemma 4, where  $P$  is either a packet (clique, cycle, star) or consists of a single column outside

the packets. Verifying  $E(M, P) \geq E(M', P)$  is easy (just slightly tedious) in each case, recalling that  $P$  is neither  $\{2, 4\}$  nor  $\{2, 3\}$ .

After termination of the interactions we build further stars as long as possible, from columns that are not yet in packets. Assume that some column  $\{3, 4\}$  still remains outside the packets. Due to the equal row sums, there must exist a column with more 1s in the first two rows than in the last two rows. But neither  $\{1, 2, 3\}$  nor  $\{1, 2, 4\}$  is in the incidence matrix, since otherwise the above interactions would still apply. Thus, for every column  $\{3, 4\}$  there also exists a column  $\{1, 2\}$ , and we can form pairs of them.

At this stage, the only possible columns outside the packets (cliques, cycles, stars, pairs) are columns with three 1s, and  $\{1, 2\}$ ,  $\{1, 4\}$ ,  $\{1, 3\}$ . If some column  $\{1, 2\}$  or  $\{1, 4\}$  exists, then either this column can be turned into  $\{1, 3\}$  by some of the interactions above, or the partner column with three 1s required for the interaction does not exist. We consider the latter case now.

Suppose that there is some  $\{1, 2\}$  but no  $\{1, 3, 4\}$ . Again, since the row sums are equal, some other columns must have more 1s in the last two rows than in the first two rows. The only possibility for that is the presence of two columns  $\{2, 3, 4\}$ , one further column  $\{1, 3\}$  and one further column  $\{1, 4\}$ . From the aforementioned columns we can build another star.

We argue similarly if some  $\{1, 4\}$  but no  $\{1, 2, 3\}$  exists. It follows that all remaining columns with two 1s outside the packets are  $\{1, 3\}$ , that is, all others have three 1s. Using again the fact that all row sums are equal, we conclude that all remaining columns can finally be grouped to diagonals and hypercliques. Finally we have managed to put all columns in packets.

With these packets we do another set of interactions which are again applied exhaustively and in any order. (See Definition 5 for the notation.)

- (1) 2 pair  $\rightarrow$  cycle
- (2) clique + hyperclique  $\rightarrow$  2 star
- (3) cycle + 2 diagonal  $\rightarrow$  2 star
- (4) cycle + hyperclique  $\rightarrow$  star + diagonal
- (5) pair + diagonal  $\rightarrow$  star

**Table 1** Energies between the packets (see Definition 5) and their inner energies  $E(P, P)/2$

$E(P,Q)$	Clique	Star	Cycle	Pair	Diagonal	Hyperclique
Clique	6	3	4	2	1	0
Star	3	0	2	1	0	0
Cycle	4	2	4	2	0	0
Pair	2	1	2	2	0	0
Diagonal	1	0	0	0	0	0
Hyperclique	0	0	0	0	0	0
$E(P,P)/2$	3	0	2	1	0	0

Due to Lemma 3, these interactions terminate. To verify their validity we use Lemma 4 and Table 1 of the pairwise and inner energies of packets. Among all interaction products, only the cycle has a positive energy, and it is only produced in interaction (1). But since  $2 + 2 \cdot 1 \geq 2$ , interaction (1) does not increase the energy of interacting packets either. It remains to compare the energies, before and after an interaction, between the interacting packets and any other packet in the partitioning. For that, we only need to compare the corresponding multiples of rows in Table 1 component-wise. For the five listed interactions this just means to confirm the following inequalities.

$$\begin{aligned}
 2 \cdot (2, 1, 2, 2, 0, 0) &= (4, 2, 4, 4, 0, 0) \geq (4, 2, 4, 2, 0, 0) \\
 (6, 3, 4, 2, 1, 0) + (0, 0, 0, 0, 0, 0) &\geq 2 \cdot (3, 0, 2, 1, 0, 0) = (6, 0, 4, 2, 0, 0) \\
 (4, 2, 4, 2, 0, 0) + 2 \cdot (1, 0, 0, 0, 0, 0) &= (6, 2, 4, 2, 0, 0) \geq 2 \cdot (3, 0, 2, 1, 0, 0) \\
 &= (6, 0, 4, 2, 0, 0) \\
 (4, 2, 4, 2, 0, 0) + (0, 0, 0, 0, 0, 0) &\geq (3, 0, 2, 1, 0, 0) + (1, 0, 0, 0, 0, 0) \\
 &= (4, 0, 2, 1, 0, 0) \\
 (2, 1, 2, 2, 0, 0) + (1, 0, 0, 0, 0, 0) &= (3, 1, 2, 2, 0, 0) \geq (3, 0, 2, 1, 0, 0)
 \end{aligned}$$

Now, only the following packets can still co-exist in a partitioning. The five cases are the maximal multisets of packets that do not contain the antecedents of the interactions done so far. Numbers in parentheses indicate the maximum number of copies of the packets; all other packets may appear arbitrarily often.

1. clique, cycle, star, pair(1)
2. clique, cycle, star, diagonal(1)
3. clique, star, diagonal
4. star, diagonal, hyperclique
5. star, hyperclique, pair(1)

Recall that  $r/n < 3/5$ . But in case 4, all packets have density at least  $3/5$ , a contradiction. In case 5, star and hyperclique have densities at least  $3/5$ . Hence the pair must be present. But a pair and a hyperclique together still have density  $2/3$ . Thus, only stars and one pair remain, which is already subsumed under case 1. Therefore we can also disregard case 5. In each of the cases 1–3, the absence of certain packets enables further interactions as shown below. (Again, Lemma 4 and Table 1 allow to verify their validity one by one.)

In case 1, the interaction “3 cycle  $\rightarrow$  2 clique” is valid since no diagonal is present. Hence at most two cycles remain.

In case 2, of course, we can also apply “3 cycle  $\rightarrow$  2 clique” if the diagonal does not exist. If the diagonal does exist, and a cycle is present as well, then the interaction “cycle + diagonal  $\rightarrow$  star + pair” applies, and it is valid since no pair is present before the interaction. Hence we get rid of either the diagonal (leading back to case 1) or all cycles (leading to case 3).

In case 3, the interaction “1 clique + 3 diagonal  $\rightarrow$  3 star” is valid since no cycle or pair is present. Hence we get rid of either the cliques (leading to the impossible case 4) or all diagonals but at most 2.

**Table 2** Energies being relevant for the final interactions

E(P,Q)	Clique	Star	Cycle	Pair	Corner
Clique	6	3	4	2	3
Star	3	0	2	1	2
Cycle	4	2	4	2	2
Pair	2	1	2	2	1
Corner	3	2	2	1	2
E(P,P)/2	3	0	2	1	1

Only two cases remain, which are those claimed in the Theorem.

1. clique, star, cycle(2), pair(1)
2. clique, star, diagonal(2)

*General situation* Finally we must also permit columns with a single 1.

Consider any column  $S$  with exactly one 1. Since  $r/n > 1/2$ , some column has three 1s. Due to Lemma 1, every such column is the complement of  $S$ . Thus, all columns with one 1 are equal to  $S$ . Furthermore, diagonals and hypercliques cannot exist, since they contain different columns with three 1s.

Precisely as before we assign the columns with two or three 1s to packets of these types, as long as possible: clique, cycle, star, pair. Now the small Lemma 2 turns out to be very useful: The interactions we had applied earlier are still valid, since the energy terms of all further columns with one 1 are not changed, due to Lemma 2. By the equal row sums and the absence of diagonals and hypercliques, all remaining columns form corners: In fact, the columns with a single 1 must be equal to  $\{1\}$ : If stars exist, then this claim follows from Lemma 1, and otherwise we can permute the rows.

Finally we erase the corners by further interactions. Their validity is checked as before, using Lemma 4 and Table 2. First we observe that the interaction “corner + pair  $\rightarrow$  cycle” is valid. If no pair exists, then “corner  $\rightarrow$  pair” is valid, too, hence we can produce a pair and do the former interaction. The interactions “2 pair  $\rightarrow$  cycle” and “3 cycle  $\rightarrow$  2 clique” remain valid. After their exhaustive application, we reach the case “clique, star, cycle(2), pair(1)”. This completes the proof.

## 6 Conclusions and further research

We have constructed optimal clique edge coverings with  $k \leq 4$  blocks, by inventing the method of interactions between packets of columns of incidence matrices that guides the search. Only the interaction sequences are laborious, but the final solutions have a nice and simple structure. We conjecture that, likewise, for every fixed  $k$ , optimal coverings with  $k$  blocks consist of only two types of packets and remainders of constant size. Probably, further ideas that reduce the amount of case distinctions would be needed to attack larger  $k$ . The method might also be suited for obtaining approximate solutions.

**Acknowledgements** Open access funding provided by Chalmers University of Technology.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Bailey, R.F., Burgess, A.C., Cavers, M.S., Meagher, K.: Generalized covering designs and clique coverings. *J. Comb. Des.* **19**, 378–406 (2011)
2. Bondy, J.A., Szwarcfiter, J.L.: Induced decompositions of graphs. *J. Graph Theory* **72**, 462–477 (2013)
3. Brouwer, A.E.: Optimal packings of  $K_4$ 's into a  $K_n$ . *J. Comb. Theory A* **26**, 278–297 (1979)
4. Colbourn, C.J., Dinitz, J.H.: *The CRC Handbook of Combinatorial Designs*. CRC Press, Boca Raton (1996)
5. Damaschke, P.: Calculating approximation guarantees for partial set cover of pairs. *Optim. Lett.* **11**, 1293–1302 (2017)
6. Gordon, D.M., Kuperberg, G., Patashnik, O.: New constructions for covering designs. *J. Comb. Des.* **3**, 269–284 (1995)
7. Greig, M., Li, P.C., van Rees, G.H.J.: Covering designs on 13 blocks revisited. *Util. Math.* **70**, 221–261 (2006)
8. Spencer, J.: Maximal consistent families of triples. *J. Comb. Theory* **5**, 1–8 (1968)
9. Mills, W.H.: Covering designs. I. Coverings by a small number of subsets. *Ars Combin.* **8**, 199–315 (1979)
10. Veremyev, A., Boginski, V., Pasiliao, E.L.: Potential energy principles in networked systems and their connections to optimization problems on graphs. *Optim. Lett.* **9**, 585–600 (2015)
11. Wilson, R.M.: Decompositions of complete graphs into subgraphs isomorphic to a given graph. *Congressus Numerantium* **XV**, 647–659 (1976)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.