The effect of edge weights on clique weights

Raphael Yuster *

Abstract

Suppose the edges of the complete r-graph on n vertices are weighted with real values. For $r \leq k \leq n$, the weight of a k-clique is the sum of the weights of its edges. Given the largest gap between the weights of two distinct edges, how small can the largest gap between the weights of two distinct k-cliques be? We answer this question precisely.

Keywords: inclusion matrices; weighted hypergraphs
MSC Codes: 05C50, 05C22

1 Introduction

All hypergraphs considered in this paper are uniform, i.e. they are r-graphs for some $r \geq 2$. The complete r-graph on n vertices is denoted by K_n^r. For an r-graph H, let $E(H)$ denote its edge set and $V(H)$ denote its vertex set. We assume that $V(K_n^r) = [n]$. Let $\binom{[n]}{k}$ denote the set of k-cliques of K_n^r where $r \leq k \leq n$. We are interested in weighings of the edges of K_n^r and their effect on the weights of larger cliques in K_n^r.

A *weighing* of K_n^r is a function $w : \binom{[n]}{r} \to \mathbb{R}$. Observe that any weighing of K_n^r induces a weighing of its subgraphs, where the weight of a subgraph is the sum of the weights of its edges. Trivially, if w is constant, then the weight of any two subgraphs with the same number of edges is the same. Now suppose that w is far from constant, what can be said about the weights of all subgraphs with the same number of edges and how far are they from being constant? In particular, what can be said about the weights of the k-cliques? We state this basic question more formally as follows.

Given $w : \binom{[n]}{r} \to \mathbb{R}$, and $r \leq k \leq n$, let

$$disc_k(w) = \max_{A,B \in \binom{[n]}{k}} |w(A) - w(B)| \ .$$

Notice that $disc(w) = disc_r(w)$ is just the maximum discrepancy between the weights of any two edges, i.e. the maximum gap between two values of w. The extremal question which emerges is to

*Department of Mathematics, University of Haifa, Haifa 31905, Israel. Email: raphy@math.haifa.ac.il

determine:

$$disc(r,k,n) = \min_w \frac{disc_k(w)}{disc(w)}$$

where $r < k \le n$ and the minimum is taken over all non-constant weighings w of K_n^r.

Our main motivation for this question (besides being natural on its own right) is that it is closely related to *inclusion matrices* and their generalized inverses. Inclusion matrices (see Section 2 for a definition) have been introduced by Gottlieb [3] and have since been well studied, mainly with respect to their rank, with applications in several areas such as quasi-randomness, see [1, 2, 4, 5, 8]. Our approach computes their generalized inverse, which, as it turns out, gives additional information and in particular assists in determining $disc(r,k,n)$.

One can easily determine $disc(r,k,n)$ when $k > n-r$. Indeed, trivially, $disc_n(w) = 0$ so $disc(r,n,n) = 0$. More generally, if $k > n-r$, then the number of elements in $\binom{[n]}{k}$ is smaller than the number of edges so the system of linear equations indexed by $\binom{[n]}{k}$ where each equation is just the sum of all variables corresponding to the edges contained in the k-set corresponding to that equation, has a nontrivial solution. Namely, we can have all weights of k-cliques 0 while w is not constant. Thus, $disc(r,k,n) = 0$ for $k > n-r$. This ceases to be the case when $k \le n-r$. Our main result determines $disc(r,k,n)$ for all relevant values of k.

To state our result, define for $0 \le t \le r < k$:

$$q(t,r,k) = (-1)^t \frac{\binom{k-r+t-1}{t}}{\binom{r}{t}\binom{k}{r}} \; .$$

Theorem 1 *For integers $2 \le r < k \le n-r$ we have $disc(r,k,n) = disc(r,k,k+r)$ and furthermore,*

$$disc(r,k,n) = \frac{2}{\max_{s=0}^{r}\left(\sum_{x=0}^{r}\sum_{y=0}^{r}\left(\sum_{j=0}^{\min\{x,y\}}\binom{s}{j}\binom{r-s}{x-j}\binom{r-s}{y-j}\binom{k-r+s}{r-y-x+j}\right)|q(x,r,k)-q(y,r,k)|\right)} \; .$$

For every fixed r, we have that for all k sufficiently large the maximum in the last equality is obtained for $s = 0$ hence for k sufficiently large

$$disc(r,k,n) = \frac{2}{\sum_{x=0}^{r}\sum_{y=0}^{r}\binom{r}{x}\binom{r}{y}\binom{k-r}{r-y-x}|q(x,r,k)-q(y,r,k)|} \; .$$

Furthermore,

$$disc(r,k,n) = \frac{1}{2^r-1} + o_k(1) \; .$$

We see from Theorem 1 that for every fixed r, for k sufficiently large $disc(r,k,n)$ is a rational function in k. For small r, it is simple enough to state the exact closed form expression.

Corollary 1.1 *For $r = 2$ (graphs), for all $n \ge k+2$ we have $disc(2,k,n) = \frac{2}{3}$ for $k = 3$ and $disc(2,k,n) = \frac{k-1}{3k-7}$ for $k \ge 4$.*

Corollary 1.2 *For $r = 3$ (3-graphs), for all $n \geq k+3$ we have $disc(3,k,n) = \frac{2}{5}$ for $k = 4$, $disc(3,k,n) = \frac{6}{17}$ for $k = 5$, $disc(3,k,n) = \frac{20}{67}$ for $k = 6$, and $disc(3,k,n) = \frac{(k-1)(k-2)}{7k^2-42k+65}$ for $k \geq 7$.*

The proof of Theorem 1 proceeds as follows. We first establish that $disc(r,k,k+r)$ is at least the value stated in the theorem. We then construct a weighing of K^r_{k+r} achieving this value thereby proving that $disc(r,k,k+r)$ is precisely the claimed value. This construction has the property that it can be extended to K^r_n for $n \geq k+r$ thereby proving that $disc(r,k,n)$ is at most the stated value. But since $disc(r,k,n) \geq disc(r,k,k+r)$ for $n \geq k+r$, equality holds. The fact that $disc(r,k,n) \geq disc(r,k,k+r)$ for $n \geq k+r$ is seen by any K^r_{k+r} subgraph of K^r_n that contains two edges with maximum discerpancy. As mentioned earlier, an important ingredient in our proof is the determination of the generalized inverses of inclusion matrices, which is done in Section 2. The cases of graphs and 3-graphs are given in Sections 3 and 4, respectively. Section 5 proves the general case of Theorem 1. In Section 6 we generalize our result in the graph-theoretic case to graphs that are not necessarily complete.

2 Inclusion matrices

For $2 \leq r \leq k \leq n$ consider the binary matrix $W = W(r,k,n)$ whose rows are indexed by $\binom{[n]}{k}$, columns are indexed by $\binom{[n]}{r}$, and $W[A,B] = 1$ if and only if $B \subseteq A$. $W(r,k,n)$ is called the *inclusion matrix* of r-sets in k-sets of $[n]$. Gottlieb [3] proved that if $k \leq n-r$, then W has full column rank. In particular, $W(r,k,r+k)$ is non-singular.

We provide here a proof of Gottlieb's result which is obtained by explicitly constructing a matrix $Q = Q(r,k,n)$ such that $QW = I$ where $W = W(r,k,n)$ and I is the identity matrix of order $\binom{n}{r}$. Recalling that for a matrix X with full column rank, a matrix Y such that $YX = I$ is called the *left generalized inverse* of X, we have that $Q(r,k,n)$ is the left generalized inverse of $W(r,k,n)$.

2.1 Constructing $Q(r,k,n)$

Throughout this subsection we assume that $W = W(r,k,n)$ and construct a matrix $Q = Q(r,k,n)$ such that $QW = I$. The rows of Q are indexed by $\binom{[n]}{r}$ while the columns of Q are indexed by $\binom{[n]}{k}$. We designate $r+1$ distinct values denoted by $q_0, \ldots, q_r$. Each entry of Q will be one of these values.

Let $A \in \binom{[n]}{r}$ be a row of Q and $B \in \binom{[n]}{k}$ be a column of Q. Then it will be the case that $Q[A,B] = q_{|A \cap B|}$. Notice that since $0 \leq |A \cap B| \leq r$ the indices are well-defined. It remains to choose values for the q_i so that indeed $QW = I$.

We start with q_r. Consider $A \in \binom{[n]}{r}$. Then the product of row A of Q and column A of W is just the sum over all $B \supseteq A$ of $Q[A,B]$, namely it is $q_r \cdot \binom{n-r}{n-k}$. As this product must be 1 it follows that one must choose

$$q_r = \frac{1}{\binom{n-r}{n-k}} \ .$$

Next consider q_{r-1}. Let $A, C \in \binom{[n]}{r}$ such that $|A \cap C| = r-1$. Then the product of row A of Q and column C of W is obtained as follows. Consider some $B \in \binom{[n]}{k}$ such that $B \supseteq C$. Then either $B \supseteq A$ or else $|B \cap A| = r-1$. In the former case, the value q_r contributes to the product and in the latter case the value q_{r-1} contributes to it. As the product must be zero we obtain

$$q_r \binom{n-r-1}{n-k} + q_{r-1} \binom{n-r-1}{n-k-1} = 0 \,.$$

Hence, using $q_r = \binom{n-r}{n-k}^{-1}$, we obtain that

$$q_{r-1} = -\frac{k-r}{\binom{n-r}{n-k}(n-k)} \,.$$

In general, consider q_{r-t} for $t = 0, \ldots, r$. Let $A, C \in \binom{[n]}{r}$ such that $|A \cap C| = r-t$. Then the product of row A of Q and column C of W is obtained as follows. Consider some $B \in \binom{[n]}{k}$ such that $B \supseteq C$. Then $|B \cap A| = r-j$ for some $j = 0, \ldots, t$. As the product must be zero we obtain

$$\sum_{j=0}^{t} q_{r-j} \binom{t}{j} \binom{n-r-t}{n-k-j} = 0 \,.$$

Specifically, using the already determined values of q_r and q_{r-1} we obtain that

$$q_{r-2} \binom{n-r}{n-k} = \frac{(k-r)(k-r+1)}{(n-k)(n-k-1)} \,.$$

Continuing in the same fashion,

$$q_{r-3} \binom{n-r}{n-k} = -\frac{(k-r)(k-r+1)(k-r+2)}{(n-k)(n-k-1)(n-k-2)}$$

and by induction substituting the previously determined values $q_r, \ldots, q_{r-t+1}$ we obtain that

$$q_{r-t} \binom{n-r}{n-k} = (-1)^t \prod_{j=0}^{t-1} \frac{k-r+j}{n-k-j} \,. \tag{1}$$

This proves that the required $q_0, \ldots, q_r$ exist and explicitly determines their value.

2.2 The case of $Q(r, k, r+k)$

We will be particularly interested in the values of q_{r-t} determined in (1) in the case $n = r+k$. In this case, we define $q(t, r, k) = q_{r-t}$ so we have by (1) that

$$q(t, r, k) = (-1)^t \frac{\binom{k-r+t-1}{t}}{\binom{r}{t}\binom{k}{r}} \,.$$

3 Graphs

Suppose that $w : E(K_n) \to \mathbb{R}$ is a weighing of the complete graph K_n. One can consider w as a (column) vector indexed by $\binom{[n]}{2}$ with real entries. Thus, for the inclusion matrix $W = W(2,k,n)$ where $2 < k \le n-2$, we have that $v = Ww$ is a vector indexed by $\binom{[n]}{k}$ where for $X \in \binom{[n]}{k}$, v_X is the weight of the k-clique induced by X.

We consider first the case $n = k+2$. In this case, $W = W(2,k,k+2)$ is non-singular, and $W^{-1} = Q(2,k,k+2)$ where the entries of Q are explicitly determined in the previous section. So, assume that we are told the weight of each k-clique of K_{k+2}, and record these values in a vector v indexed by $\binom{[k+2]}{k}$ where v_X is the weight of the k-clique induced by X. Then we can recover uniquely the edge-weighing w of K_{k+2} giving rise to these weights of the k-cliques by computing $w = Qv$.

Recall that our goal is to estimate $disc_k(w)/disc(w)$. Now, $disc_k(w)$ is the maximum difference between two coordinates of v while $disc(w)$ is the maximum difference between two coordinates of w, i.e. two coordinates of Qv.

Suppose that $disk_k(w) = 2\delta$, and hence there exists $s \in \mathbb{R}$ such that each coordinate of v is in $[s-\delta, s+\delta]$. We may assume that $s = 0$ as subtracting the same weight from all edges does not change the discrepancies. Suppose that $disc(w)$ is realized by the two edges e and f so that $|w(e) - w(f)| = disc(w)$. Now, $w(e)$ is the product of row $Q[e]$ with v while $w(f)$ is the product of row $Q[f]$ with v. Thus,

$$disc(w) = |(Q[e] - Q[f])v| \ . \tag{2}$$

Entry $Q[e,Z]$ for $Z \in \binom{k+2}{k}$ is either q_0, q_1, or q_2 depending on $|e \cap Z| \in \{0,1,2\}$. Similarly $Q[f,Z]$ is either q_0, q_1, or q_2 depending on $|f \cap Z| \in \{0,1,2\}$. We recall from the previous section that

$$q_2 = q(0,2,k) = \frac{2}{k(k-1)} \ , \ q_1 = q(1,2,k) = -\frac{k-2}{k(k-1)} \ , \ q_0 = q(2,2,k) = \frac{k-2}{k} \ . \tag{3}$$

Now, clearly, $e \neq f$ as w is assumed to be non-constant. So, there are two cases to consider: either $|e \cap f| = 1$ or $e \cap f = \emptyset$. Consider first the case $e \cap f = \emptyset$. For how may distinct Z could we have that $Q[e,Z] = q_0$ and $Q[f,Z] = q_0$? Clearly, this can never happen. Similarly, it can never happen that $Q[e,Z] = q_0$ and $Q[f,Z] = q_1$. However, $Q[e,Z] = q_0$ and $Q[f,Z] = q_2$ happens exactly once, for $Z = [k+2] \setminus e$. Now, $Q[e,Z] = q_1$ and $Q[f,Z] = q_1$ occurs four times, as such a Z must contain all elements of $[k+2] - (e \cup f)$, and also contain one of the two endpoints of e and one of the two endpoints of f. Likewise, $Q[e,Z] = q_1$ and $Q[f,Z] = q_2$ occurs $2(k-2)$ times. Finally, $Q[e,Z] = q_2$ and $Q[f,Z] = q_2$ occurs $\binom{k-2}{2}$ times. The other possibilities are computed symmetrically by changing the roles of e and f. These values are summarized in the left part of Figure 1. It follows from this discussion, from (2), from (3), and from the fact that every coordinate of v has absolute value at most δ that

$$disc(w) \le \delta \cdot (2 \cdot |q_0 - q_2| + 4(k-2) \cdot |q_1 - q_2|) = \delta \frac{6k-14}{k-1} \ . \tag{4}$$

$Q[f,Z]$ vs. $Q[e,Z]$	q_0	q_1	q_2
q_0	0	0	1
q_1	0	4	$2k-4$
q_2	1	$2k-4$	$\binom{k-2}{2}$

$e \cap f = \emptyset$

$Q[f,Z]$ vs. $Q[e,Z]$	q_0	q_1	q_2
q_0	0	1	0
q_1	1	k	$k-1$
q_2	0	$k-1$	$\binom{k-1}{2}$

$|e \cap f| = 1$

Figure 1: The number of entries of Q for which $Q[e,Z] = q_i$ and $Q[f,Z] = q_j$ where Z ranges over all columns. The left table is for the case where $e \cap f = \emptyset$ and the right table is for the case $|e \cap f| = 1$.

We now consider the case $|e \cap f| = 1$. In a similar way, for every $i = 0,1,2$ and $j = 0,1,2$ we compute the number of columns Z for which $Q[e,Z] = q_i$ and $Q[f,Z] = q_k$. These values are summarized in the right part of Figure 1. Thus, in this case we have that

$$disc(w) \le \delta \cdot (2 \cdot |q_0 - q_1| + 2(k-1) \cdot |q_1 - q_2|) = \delta \frac{4k-6}{k-1} \ . \qquad (5)$$

Since $disk_k(w) = 2\delta$, it follows from (4) that in the case $e \cap f = \emptyset$ we have $disc_k(w)/disc(w) \ge (k-1)/(3k-7)$ and it follows from (5) that in the case $|e \cap f| = 1$ we have $disc_k(w)/disc(w) \ge (k-1)/(2k-3)$. We therefore have that for $k = 3$, $disc(2,3,5) \ge 2/3$ while for $k \ge 4$ we have $disc(2,k,2+k) \ge (k-1)/(3k-7)$. We now show that this lower bound is tight.

For the case $k = 3$, consider the following weighing of K_5: Assign weights $w(1,2) = 1$, $w(2,3) = 0$, $w(4,5) = 1/2$, $w(1,3) = 1/2$, the two remaining edges incident with 1 receive weight $1/3$, the two remaining edges incident with 2 receive weight $1/2$, and the two remaining edges incident with 3 receive weight $2/3$. It is easy to verify that the smallest weight of a triangle is $7/6$ while the largest wright of a triangle is $11/6$. So for this weighing we have $disc(w) = 1$ while $disc_3(w) = 2/3$. The construction proves that $disc(2,3,5) = 2/3$.

For the case $k \ge 4$ the construction proceeds as follows. We construct a vector v^* as above which corresponds to all the weights of the K_k in K_{k+2} where each coordinate of v^* is in $\{-\delta, 0, \delta\}$, at least one coordinate is $+\delta$ and at least one coordinate is $-\delta$. We need to show that v^* has the property that there are two coordinates of Qv^* that differ by $\delta\frac{6k-14}{k-1}$. Recall that we know that in this case our lower bound $(k-1)/(3k-7)$ is obtained in the case $e \cap f = \emptyset$. So, wlog, we will use $e = \{1,2\}$ and $f = \{3,4\}$. For the unique Z such that $Q[e,Z] = q_0$ and $Q[f,Z] = q_2$ (i.e. for $Z = \{3,4\ldots,k+2\}$), set $v^*_Z = \delta$. For the unique Z such that $Q[e,Z] = q_2$ and $Q[f,Z] = q_0$, set $v^*_Z = -\delta$. For all Z such that $Q[e,Z] = q_1$ and $Q[f,Z] = q_2$ (recall from Figure 1 that there are $2k-4$ such Z), set $v^*_Z = -\delta$ and for all Z such that $Q[e,Z] = q_2$ and $Q[f,Z] = q_1$, set $v^*_Z = \delta$. For all other Z, set v^*_Z to the same value, say 0. Now, since $q_0 > q_2 > q_1$, our construction shows that for this v^*, (4) is, in fact, an equality, and coordinate e of Qv^* is larger than coordinate f of Qv^* by precisely $\delta\frac{6k-14}{k-1}$. Let $w^* = Qv^*$ denote the resulting edge weighing of K_n. The construction proves that $disc(2,k,2+k) = (k-1)/(3k-7)$ for $k \geq 4$.

We show how to extend this construction of w^* from the case $n = k+2$ to all $n \geq k+2$. For an ordered pair of edges (e,f) of K_n, the *profile* of any edge $g \in E(K_n)$ is the ordered triple (p_1,p_2,p_3) where $p_1 = |e \cap g|$, $p_2 = |f \cap g|$ and $p_3 = |e \cap f \cap g|$. So, by the above construction of w^* (that is, by the construction of v^*), we see that if g and g' are two edges of K_{k+2} having the same profile, then $w^*(g_1) = w^*(g_2)$. Thus, we can extend the weighing w^* to all edges g of K_n by setting $w^*(g)$ equal to the weight of an edge of K_{k+2} with the same profile.

In fact, notice that in our construction for the case $k \geq 4$, we have used $e = \{1,2\}$ and $f = \{3,4\}$ so there are 6 distinct profiles, $(2,0,0),(0,2,0),(1,0,0),(0,1,0),(1,1,0),(0,0,0)$. In our construction for the case $k=3$ we have in fact used $e=(1,2)$ $f=(2,3)$ and there are 7 distinct profiles, $(2,1,1)$ which is the profile of $(1,2)$, $(1,2,1)$ which is the profile of $(2,3)$, $(1,1,0)$ which is the profile of $(1,3)$, $(1,0,0)$ which is the profile of $(1,4)$ and $(1,5)$, $(1,1,1)$ which is the profile of $(2,4)$ and $(2,5)$, $(0,1,0)$ which is the profile of $(3,4)$ and $(3,5)$, and $(0,0,0)$ which is the profile of $(4,5)$.

Now, suppose $n \geq k+2$. Define the profile of a K_k copy Z of K_n as the vector indexed by all possible profiles, where each coordinate is the number of edges of Z with the given profile. So, for example, consider the case $k=4$ and the extension of our constructed weighing w^* of K_6 to K_7. The profile of, say, $Z = \{2,3,4,7\}$ is as follows. There are no edges of Z with the profile $(2,0,0)$ since Z does not contain $e = \{1,2\}$. There is one edge of Z with the profile $(0,2,0)$ since Z contains $\{3,4\}$. There is one edge of Z with the profile $(1,0,0)$, namely $\{2,7\}$. There are two edges of Z with the profile $(0,1,0)$, namely $\{3,7\}$ and $\{4,7\}$. There are two edges of Z with the profile $(1,1,0)$, namely $\{2,3\}$ and $\{2,4\}$. There are no edges of Z with the profile $(0,0,0)$.

Since two K_k's with the same profile have the same weight, and for any Z there is a Z' already in K_{k+2} with the same profile, we obtain that $disc_k(w^*)$ and $disc(w^*)$ did not change after this extension. This proves that $disc(2,k,n) \leq disc(2,k,k{+}2)$. Since, trivially, $disc(2,k,n) \geq disc(2,k,k{+}2)$, equality holds. Summarizing, we have obtained that For all $n \geq k+2$ we have $disc(2,k,n) = \frac{2}{3}$ for $k=3$ and $disc(2,k,n) = \frac{k-1}{3k-7}$ for $k \geq 4$. Hence, we have proved Corollary 1.1.

4 3-graphs

The arguments given in the previous section for graphs can be extended to r-graphs for $r \geq 3$, but become more involved as there are more intersection types, more pairwise intersections, considerably more edge profiles, and even more clique profiles. Still, the case $r=3$ is simple enough to be explicitly given as a closed formula.

Suppose that $w : E(K_n^3) \to \mathbb{R}$ is a weighing of the complete 3-graph K_n^3. One can consider w as a (column) vector indexed by $\binom{[n]}{3}$ with real entries. Thus, for the inclusion matrix $W = W(3,k,n)$ where $3 < k \leq n-3$, we have that $v = Ww$ is a vector indexed by $\binom{[n]}{k}$ where for $X \in \binom{[n]}{k}$, v_X is the weight of the k-clique induced by X.

We consider first the case $n = k+3$. In this case, $W = W(3,k,k+3)$ is non-singular, and $W^{-1} = Q(3,k,k+3)$ where the entries of Q are explicitly determined in Section 2. So, assume that we are told the weight of each k-clique of K^3_{k+3}, and record these values in a vector v indexed by $\binom{[k+3]}{k}$ where v_X is the weight of the k-clique induced by X. Then we can recover uniquely the edge-weighing w of K_{k+3} giving rise to these weights of the k-cliques by computing $w = Qv$.

As in the previous section, suppose that $disk_k(w) = 2\delta$, and that each coordinate of v is in $[-\delta, \delta]$. Suppose that $disc(w)$ is realized by the two edges e and f so that $|w(e) - w(f)| = disc(w)$. In the graph theoretic case we had only two possibilities to consider, corresponding to $|e \cap f|$. Now we have three since $|e \cap f| \in \{0,1,2\}$.

Further, we recall from Section 2 that

$$q_3 = \frac{1}{\binom{k}{3}} \;,\; q_2 = -\frac{k-3}{3\binom{k}{3}} \;,\; q_1 = \frac{k-3}{k(k-1)} \;,\; q_0 = -\frac{k-3}{k} \;. \tag{6}$$

We compute tables analogous to the tables in Figure 1. We need a table for each possible value of $|e \cap f|$. These are given in Figure 2.

As in (4) we obtain from Figure 2 for the case $e \cap f = \emptyset$ that

$$disc(w) \le \delta \cdot \left(2 \cdot |q_3 - q_0| + 18 \cdot |q_1 - q_2| + 6(k-3)|q_1 - q_3| + 6\binom{k-3}{2}|q_3 - q_2|\right) . \tag{7}$$

For all $k \ge 5$ we have $q_1 \ge q_3 \ge q_2 \ge q_0$ so (6) and (7) imply for $k \ge 5$ that

$$disc(w) \ge \delta \frac{k-3}{k(k-1)(k-2)} \left(\frac{12}{k-3} + 4 - 42k + 14k^2\right) \tag{8}$$

while for $k = 4$

$$disc(w) \ge 5\delta \, . \tag{9}$$

For the case $|e \cap f| = 1$ we obtain from Figure 2 that

$$disc(w) \le \delta \cdot \left(2 \cdot |q_2 - q_0| + 4(k-1) \cdot |q_1 - q_2| + 2(k-2)|q_1 - q_3| + 4\binom{k-2}{2}|q_3 - q_2|\right) . \tag{10}$$

So for $k \ge 5$ we get from (6) and (10) that

$$disc(w) \ge \delta \frac{k-3}{k(k-1)(k-2)} \left(-\frac{12}{k-3} - 4 - 26k + 12k^2\right) \tag{11}$$

while for $k = 4$

$$disc(w) \ge \frac{13}{3}\delta \, . \tag{12}$$

For the case $|e \cap f| = 2$ we obtain from Figure 2 that

$$disc(w) \le \delta \cdot \left(2 \cdot |q_1 - q_0| + 4(k-1) \cdot |q_1 - q_2| + 2\binom{k-1}{2}|q_3 - q_2|\right) . \tag{13}$$

$Q[f,Z]$ vs. $Q[e,Z]$	q_0	q_1	q_2	q_3
q_0	0	0	0	1
q_1	0	0	9	$3(k-3)$
q_2	0	9	$9(k-3)$	$3\binom{k-3}{2}$
q_3	1	$3(k-3)$	$3\binom{k-3}{2}$	$\binom{k-3}{3}$

$e \cap f = \emptyset$

$Q[f,Z]$ vs. $Q[e,Z]$	q_0	q_1	q_2	q_3
q_0	0	0	1	0
q_1	0	4	$2(k-1)$	$k-2$
q_2	1	$2(k-1)$	$\binom{k-2}{2}+4(k-2)$	$2\binom{k-2}{2}$
q_3	0	$k-2$	$2\binom{k-2}{2}$	$\binom{k-2}{3}$

$|e \cap f| = 1$

$Q[f,Z]$ vs. $Q[e,Z]$	q_0	q_1	q_2	q_3
q_0	0	1	0	0
q_1	1	$k+1$	$2(k-1)$	0
q_2	0	$2(k-1)$	$(k-1)+2\binom{k-1}{2}$	$\binom{k-1}{2}$
q_3	0	0	$\binom{k-1}{2}$	$\binom{k-1}{3}$

$|e \cap f| = 2$

Figure 2: The number of entries of Q for which $Q[e,Z] = q_i$ and $Q[f,Z] = q_j$ where Z ranges over all columns in the various possibilities of $|e \cap f|$ for 3-graphs.

So for all $k \geq 4$ we get from (6) and (13) that

$$disc(w) \geq \delta \frac{k-3}{k(k-1)(k-2)} \left(\frac{12}{k-3} + 4 - 8k + 8k^2 \right) . \tag{14}$$

Recall that $disk_k(w) = 2\delta$. For $k = 4$, the maximum of (9),(12), (14) is obtained by (9) which implies that $disc_4(w)/disc(w) \geq 2\delta/(5\delta)$ for all possible w hence $disc(3, 4, 7) \geq 2/5$. For $k = 5$, the maximum of (8), (11), (14) is obtained by (14) and it is $17\delta/3$. Hence, $disc(3, 5, 8) \geq 6/17$. For $k = 6$, the maximum is obtained in (11) and it is $67/10$. Hence, $disc(3, 6, 9) \geq 20/67$. For all $k \geq 7$ the maximum of (8), (11), (14) is obtained by (14) and hence for all $k \geq 7$,

$$disc(3, k, k, k+3) \geq \frac{2\delta}{\delta \frac{k-3}{k(k-1)(k-2)} \left(\frac{12}{k-3} + 4 - 42k + 14k^2 \right)} = \frac{(k-1)(k-2)}{7k^2 - 42k + 65} = \frac{1}{7} + o_k(1) .$$

As in the proof for the graph theoretic case in the previous section, for each of the possible values of $|e \cap f|$ we can construct a vector v^* which corresponds to all the weights of the K_k in K_{k+3} where each coordinate of v^* is in $\{-\delta, 0, \delta\}$, and such that each of (7), (10), (13) is an equality, and the corresponding weighing is $Qv^* = w^*$. Hence all lower bounds for $disc(3, k, k+3)$ that have been computed in the previous paragraph are, indeed, equalities.

We show how to extend w^* from the case $n = k+3$ to all $n \geq k+3$. We use the exact same idea of profiles as in the previous section for the graph-theoretic case. For an ordered pair of edges (e, f) of K_n^3, the *profile* of any edge $g \in E(K_n^3)$ is the ordered triple (p_1, p_2, p_3) where $p_1 = |e \cap g|$, $p_2 = |f \cap g|$ and $p_3 = |e \cap f \cap g|$. So, if g and g' are two edges of K_{k+3}^3 having the same profile, then $w^*(g_1) = w^*(g_2)$. Thus, we can extend the weighing w^* to all edges g of K_n^3 by setting $w^*(g)$ equal to the weight of an edge of K_{k+3}^3 with the same profile. Now, suppose $n \geq k+3$. Define the profile of a K_k^3 copy Z of K_n^3 as the vector indexed by all possible profiles, where each coordinate is the number of edges of Z with the given profile. Since two K_k^3's with the same profile have the same weight, and for any Z there is a Z' already in K_{k+3}^3 with the same profile, we obtain that $disc_k(w^*)$ and $disc(w^*)$ did not change after this extension. This proves that $disc(3, k, n) \leq disc(3, k, k+3)$. Since, trivially, $disc(3, k, n) \geq disc(3, k, k+3)$, equality holds.

Summarizing, we have obtain Corollary 1.2: For all $n \geq k+3$ we have $disc(3, k, n) = \frac{2}{5}$ for $k = 4$, $disc(3, k, n) = \frac{6}{17}$ for $k = 5$, $disc(3, k, n) = \frac{20}{67}$ for $k = 6$, and $disc(3, k, n) = \frac{(k-1)(k-2)}{7k^2-42k+65}$ for $k \geq 7$.

5 r-graphs

As can be seen from the two previous sections, we can determine $disk(r, k, n)$ for all $2 \leq r < k \leq n$, but the precise closed formula becomes increasingly difficult to compute as r grows. The following describes the general procedure.

So, suppose r is fixed and $w : E(K_n^r) \to \mathbb{R}$ is a weighing of the complete r-graph K_n^r. One can consider w as a (column) vector indexed by $\binom{[n]}{r}$ with real entries. Thus, for the inclusion matrix

$W = W(r,k,n)$ where $r < k \le n-r$, we have that $v = Ww$ is a vector indexed by $\binom{[n]}{k}$ where for $X \in \binom{[n]}{k}$, v_X is the weight of the k-clique induced by X. As shown in the previous two sections, we need to consider the case $n = k+3$ where in this case, $W = W(r,k,k+r)$ is non-singular, and $W^{-1} = Q(r,k,k+r)$ where the entries of Q are explicitly determined in Section 2. So, assume that we are told the weight of each k-clique of K^r_{k+r}, and record these values in a vector v indexed by $\binom{[k+r]}{k}$ where v_X is the weight of the k-clique induced by X. Then we can recover uniquely the edge-weighing w of K_{k+r} giving rise to these weights of the k-cliques by computing $w = Qv$. As in the previous two sections, suppose that $disk_k(w) = 2\delta$, and that each coordinate of v is in $[-\delta,\delta]$. Suppose that $disc(w)$ is realized by the two edges e and f so that $|w(e)-w(f)| = disc(w)$. Notice that $|e \cap f| \in \{0,1,2,r-1\}$.

We define tables analogous to the tables in Figure 1 and Figure 2. We need a table for each possible value of $|e \cap f|$. Denote the tables by M^r_s for $s = 0,\ldots,r-1$. Each M^r_s is a symmetric matrix of order $r+1$ whose rows and columns are indexed by $q_0,\ldots,q_r$ where $q_{r-t} = q(t,r,k)$. The value of $M^r_s[q_i,q_j]$ equals the number of entries of $Q = Q(r,k,k+r)$ for which $Q[e,Z] = q_i$ and $Q[f,Z] = q_j$ where Z ranges over all columns, and $|e \cap f| = s$. So, for example M^3_1 is the middle table in Figure 2 and $M^3_1[q_2,q_3] = 2\binom{k-2}{2}$. We next determine $M^r_s[q_i,q_j]$ in general. Namely, we have that $|e \cap f| = s$ and wish to determine the number of subsets Z of $[k+r]$ of order k for which $|Z \cap e| = i$ and $|Z \cap f| = j$. Letting $i = r-x$ and $j = r-y$ we clearly have

$$M^r_s[q_{r-x},q_{r-y}] = \sum_{j=0}^{\min\{x,y\}} \binom{s}{j}\binom{r-s}{x-j}\binom{r-s}{y-j}\binom{k-r+s}{r-y-x+j}\,. \tag{15}$$

Let $disc(s,r,k,r+k)$ denote the minimum of $disc(w)$ where w ranges over all weighings $w : E(K^r_{k+r}) \to \mathbb{R}$ with $disc_k(w) = 2\delta$ and where the two edges e,f that realize $disc(w)$ (i.e. $|w(e)-w(f)| = disc(w)$) satisfy $|e \cap f| = s$. Generalizing (4), (5), (7),(10),(13), we have that

$$disc(s,r,k,r+k) \ge \delta\left(\sum_{i=0}^{r}\sum_{j=0}^{r} M^r_s[q_i,q_j]|q_i-q_j|\right)\,.$$

As shown in the previous two sections, equality can always be attained by choosing an appropriate vector v^* and hence we have

$$disc(r,k,r+k) = \min_{s=0}^{r}\frac{2\delta}{disc(s,r,k,r+k)} = \frac{2}{\max_{s=0}^{r}\left(\sum_{i=0}^{r}\sum_{j=0}^{r} M^r_s[q_i,q_j]|q_i-q_j|\right)}\,.$$

Using the argument of edge profiles and clique profiles given in the previous two sections we obtain that for all $n \ge r+k$,

$$disc(r,k,n) = disc(r,k,r+k) = \frac{2}{\max_{s=0}^{r}\left(\sum_{i=0}^{r}\sum_{j=0}^{r} M^r_s[q_i,q_j]|q_i-q_j|\right)}\,.$$

Plugging in (15) we get that

$$disc(r,k,n) = \frac{2}{\max_{s=0}^{r}\left(\sum_{x=0}^{r}\sum_{y=0}^{r}\left(\sum_{j=0}^{\min\{x,y\}}\binom{s}{j}\binom{r-s}{x-j}\binom{r-s}{y-j}\binom{k-r+s}{r-y-x+j}\right)|q_{r-x}-q_{r-y}|\right)} \tag{16}$$

where for any fixed $r \geq 2$, $q_{r-x} = q(x,r,k)$ and $q_{r-y} = q(y,r,k)$ and thus

$$|q_{r-x}-q_{r-y}| = \left|(-1)^x\frac{\binom{k-r+x-1}{x}}{\binom{r}{x}\binom{k}{r}} - (-1)^y\frac{\binom{k-r+y-1}{y}}{\binom{r}{y}\binom{k}{r}}\right| .$$

This proves the first part of Theorem 1.

Notice that for any fixed r we have that $disc(r,k,n)$ is a rational function in k. We have determined it precisely for $r = 2,3$ in the previous two sections. Although a closed formula for numerator and denominator coefficients of this rational function for arbitrary r seems hopeless, we would still like to determine its asymptotic value, and thus complete the proof of Theorem 1.

So, we now fix $r \geq 2$ and wish to determine the asymptotic value of $disc(r,k,n) = disc(r,k,r+k)$ as a function of k. We start with q_{r-x} and q_{r-y}. Notice that

$$|q_{r-t}| = |q(t,r,k)| = \frac{\binom{k-r+t-1}{t}}{\binom{r}{t}\binom{k}{r}} = \frac{(r-t)!}{k^{r-t}+o(k^{r-t})} .$$

Thus, if $t = \max\{x,y\}$ and furthermore $x \neq y$, then

$$|q_{r-x}-q_{r-y}| = \frac{(r-t)!}{k^{r-t}+o(k^{r-t})} . \tag{17}$$

We next analyze the denominator of (16) for each s separately. Thus let

$$g(s,r,k) = \sum_{x=0}^{r}\sum_{y=0}^{r}\left(\sum_{j=0}^{\min\{x,y\}}\binom{s}{j}\binom{r-s}{x-j}\binom{r-s}{y-j}\binom{k-r+s}{r-y-x+j}\right)|q_{r-x}-q_{r-y}| .$$

Consider first the case $s = 0$. Here we have

$$g(0,r,k) = \sum_{x=0}^{r}\sum_{y=0}^{r}\binom{r}{x}\binom{r}{y}\binom{k-r}{r-y-x}|q_{r-x}-q_{r-y}| .$$

By (17) all terms in the sum defining $g(0,r,k)$ are $o_k(1)$ unless precisely one of x or y is 0. If, say, $x = 0$ and $y > 0$ the corresponding term is $\binom{r}{y}\binom{k-r}{r-y}\frac{(r-y)!}{k^{r-y}+o(k^{r-y})} = \binom{r}{y} + 0_k(1)$. Thus,

$$g(0,r,k) = \left(\sum_{y=1}^{r}\binom{r}{y} + 0_k(1)\right) + \left(\sum_{x=1}^{r}\binom{r}{x} + 0_k(1)\right) = 2^{r+1} - 2 + o_k(1) .$$

Now consider the case $s > 0$. Here we have by (17) that whenever, say, $y > x$, the terms in the inner summation on j in the definition of $g(s,r,k)$ are all $o_k(1)$ unless $j = x$. Hence the inner term amounts to

$$\binom{s}{x}\binom{r-s}{y-x} + o_k(1) .$$

Thus,

$$g(s,r,k) = 2\left(\sum_{x=0}^{r-1}\sum_{y=x+1}^{r}\binom{s}{x}\binom{r-s}{y-x}\right) + 0_k(1) = 2^{r+1} - 2^{s+1} + o_k(1)\ .$$

Since for all k sufficiently large $2^{r+1} - 2^{s+1} + o_k(1)$ is maximized when $s = 0$, we have that for all k sufficiently large, the denominator of (16) is maximized when $s = 0$ and hence

$$disc(r,k,n) = \frac{2}{\sum_{x=0}^{r}\sum_{y=0}^{r}\binom{r}{x}\binom{r}{y}\binom{k-r}{r-y-x}|q(x,r,k) - q(y,r,k)|}\ .$$

and furthermore,

$$disc(r,k,n) = \frac{1}{2^r - 1} + o_k(1)$$

completing the proof of Theorem 1. ■

6 Dense graphs

We can generalize the definition of $disc_k(w)$ from complete r-graphs to general r graphs. Let G be an r-graph and let $w : E(G) \to \mathbb{R}$. Then, $disc_k(w) = \max_{A,B\in\binom{G}{k}} |w(A) - w(B)|$ where now $\binom{G}{k}$ is the set of all k-cliques of G. Generalizing the notion of $disc(r,k,n)$, for an r-graph G and $k > r$, we define

$$disc(G,k) = \min_{w} \frac{disc_k(w)}{disc(w)}$$

where $r < k$ and the minimum is taken over all non-constant weighings w of G.

Trivially, if G has no k-clique, then $disc(G,k) = 0$. But since the Turán number of hypergraphs is by itself a notoriously difficult problem (the Turán number of any K_r^k and where $k > r > 2$ is not known, even asymptotically [6]), we will restrict our attention to graphs. Let us consider all graphs with n vertices and minimum degree at least d for some $0 \le d < n$. Thus, let

$$disc([n,d],k) = \min_{G} disc(G,k)$$

where the minimum is taken over all graphs with n vertices and minimum degree at least d. The reason for looking at the family of graphs with a certain minimum degree requirement and not just the family of graphs with a certain number of edges is obvious. In the latter case we can construct graphs which are almost complete, say take a K_{n-1}, and an additional vertex v connected to just one other vertex of the K_{n-1}. We can assign a nonzero weight to the edge incident with v and weight 0 to all other edges. For $k \ge 3$, every K_k in this graph has weight 0, while the edge weighing is non-constant.

Once again, if there exist graphs with n vertices and minimum degree d that do not contain a K_k, then, trivially, $disc([n,d],k) = 0$. However, Turán's theorem [7] tells us what is the minimum d which guarantees that every graph with n vertices and that minimum degree has a K_k. That, however, is not enough since it may still be possible for a non-constant weighing to have all k-cliques with the same weight.

Theorem 2 *If $d \leq \lfloor \frac{k}{k+1}n \rfloor$, then $disc([n,d],k) = 0$. If $d > \lfloor \frac{k}{k+1}n \rfloor$, then $disc([n,d],k) = disc(2,k,n)$.*

Proof. Recall that the Turán graph $T(s,n)$ is the unique complete s-partite graph with n vertices and with each vertex part of size either $\lfloor n/s \rfloor$ or $\lceil n/s \rceil$. While $T(s,n)$ does not contain K_{s+1}, Turán's theorem states that if a graph with n vertices has more edges than $T(s,n)$ has, then it does contain K_{s+1}.

Recall also from the introduction that there are nonzero weighings of K_{k+1} such that every copy of K_k has the same weight, that is $disc(2,k,k+1) = 0$. So take such a weighing w of K_{k+1} on vertex set $[k+1]$ and notice that $T(k+1,n)$ is a *blowup* of K_{k+1}, that is we replace vertex v with an independent set X_v of size either $\lfloor n/(k+1) \rfloor$ or $\lceil n/(k+1) \rceil$ and add all possible edges between X_v and X_u when $u \neq v$. So we can also assign the weight $w(u,v)$ to all edges between X_u and X_v and obtain a nonzero weighing of $T(k+1,n)$ where each copy of K_k in $T(k+1,n)$ has the same weight. Now, since the minimum degree of $T(k+1,n)$ is $n - \lceil n/(k+1) \rceil \leq \lfloor \frac{k}{k+1}n \rfloor$ we have that whenever $d \leq \lfloor \frac{k}{k+1}n \rfloor$, then $disc([n,d],k) = 0$.

Now suppose that G is a graph with n vertices whose minimum degree d satisfies $d \geq \lfloor \frac{k}{k+1}n \rfloor + 1$. Every set of t vertices where $1 \leq t \leq k+1$ has a common neighbor. Indeed, suppose T is such a set, then the number of vertices not adjacent to at least one element of T is at most $t(n-1-d)$ so at least $n - t - t(n-1-d) > 0$ vertices are adjacent to all vertices of T. It follows that for each K_t-subgraph of G on the set of vertices T, there is a K_{k+2} subgraph of G that contains T.

Consider the graph H whose vertices are the edges of G and two vertices of H are adjacent in H if the corresponding edges appear in the same copy of a K_{k+2} in G. We claim that H is connected. Indeed, suppose not. Then it is possible to partition $V(H) = E(G)$ into two parts F_1 and F_2 with no edges of H connecting the two parts. For each $v \in V(G)$ let $F_i(v)$ be the set of all edges of G incident with v that belong to F_i for $i = 1,2$. Notice that $|F_1(v)| + |F_2(v) \geq d$. Since G is connected, we must have that for some $v \in V(G)$, $F_1(v) \neq \emptyset$ and also $F_2(v) \neq \emptyset$. Wlog, assume $|F_2(v)| \geq |F_1(v)|$ and let $(v,x) \in F_1(v)$. But since $k \geq 3$ and $|F_1(v)| + |F_2(v)| \geq d > \frac{k}{k+1}n$ we must have that $(x,y) \in E(G)$ for some y with $(v,y) \in F_2(v)$. But according to the previous paragraph, the triple (v,x,y) belongs to some K_{k+2}-subgraph of G. So it cannot be that (v,x) and (v,y) are in distinct components of H.

Now suppose that w is a non-constant weighing of G. By the connectivity of H, there is some copy of K_{k+2} of G which has two edges whose weight is not the same. thus $disc_k(w)/disc(w) \geq disc(2,k,k+2)$. It follows that $disc(G,k) \geq disc(2,k,k+2) = disc(2,k,n)$. Thus, $disc([n,d],k) \geq disc(2,k,n)$. Since trivially $disc([n,d],k) \leq disc(2,k,n)$, the theorem follows. ■

References

[1] D. De Caen. A note on the ranks of set-inclusion matrices. *The Electronic Journal of Combinatorics*, 8(1):N5–N5, 2001.

[2] A. Frumkin and A. Yakir. Rank of inclusion matrices and modular representation theory. *Israel Journal of Mathematics*, 71(3):309–320, 1990.

[3] D.H. Gottlieb. A certain class of incidence matrices. *Proceedings of the American Mathematical Society*, 17(6):1233–1237, 1966.

[4] J. Guo, F. Li, and K. Wang. Incidence matrices of finite attenuated spaces and class dimension of association schemes. *Discrete Mathematics*, 315:42–46, 2014.

[5] A. Shapira. Quasi-randomness and the distribution of copies of a fixed graph. *Combinatorica*, 28(6):735–745, 2008.

[6] A. Sidorenko. What we know and what we do not know about Turán numbers. *Graphs and Combinatorics*, 11(2):179–199, 1995.

[7] P. Turán. On an extremal problem in graph theory. *Mat. Fiz. Lapok*, 48(436-452):137, 1941.

[8] R.M. Wilson. A diagonal form for the incidence matrices of t-subsets vs. k-subsets. *European Journal of Combinatorics*, 11(6):609–615, 1990.