UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2013:20

Extremal Hypergraphs

Teo Asplund

Examensarbete i matematik, 15 hp
Handledare och examinator: Vera Koponen
Juni 2013

Department of Mathematics
Uppsala University

Extremal Hypergraphs

Teo Asplund

Abstract

In this paper we take a look at some problems and results in *extremal graph theory*. We prove some theorems regarding the maximal size of graphs with forbidden subgraphs and find some results relating to the structure of hypergraphs with certain forbidden subgraphs.

1 Introduction

History. The foundations of graph theory lie in a paper written by Euler on *the Königsberg bridge problem* [1]. Published in 1736, the paper answered negatively the question of whether a walk through the city of Königsberg crossing all its seven bridges exactly once was possible. In his proof Euler reformulated the problem in more abstract terms by noting that the only important feature of a route is the sequence in which bridges are crossed. Thus the separate parts of the city can be replaced by *vertices* and the bridges may be shown as connections, or *edges*, between these vertices. This abstraction results in a *graph* (see figure 1) and the result may be regarded as the first theorem of graph theory.

Much later, in 1907, the proof of *Mantel's theorem* was the beginning of *extremal graph theory*. The name comes from the fact that this branch of graph theory deals with *maximal* or *minimal* graphs, with regard to some property, that satisfy some requirement. The maximal (or minimal) graph property may, for example, be its number of edges or its number of vertices. In the case of Mantel, the theorem answers the question of how many edges a graph may have *at most* without containing any triangles. In 1941 Pál Turán proved a generalized version of this theorem. In section 2 we will prove both these theorems.

Notation and terminology. First of all, note that in this essay *log* always denotes the natural logarithm. We define $X^{(r)} = \{Y : Y \subset X, |Y| = r\}$. A set Y where $|Y| = r$ is called an r-set. By graph we mean an ordered pair of sets (V, E) where V is a non-empty set of *vertices* and $E \subset V^{(2)}$ is the set of *edges*. An edge $\{x, y\}$ *joins* x to y. We say that a vertex x is *incident* with the edge $\{x, y\}$. If G is a graph (V, E) then $V(G)$ is the set of vertices V and $E(G)$ is E. Two vertices are *adjacent* or *connected* if they are part of the same edge. The set of vertices adjacent to a vertex x is denoted by $\Gamma(x)$ and $d(x) = |\Gamma(x)|$ is said to be the *degree* of x. The *minimum degree* of a graph is the smallest degree of all the vertices in a graph, and is written $\delta(G)$. The number of vertices in a graph G is called the *order* of G and is denoted by $|G|$. The number of edges of G is known as its *size* and is denoted by $e(G)$. G^n denotes a graph of order n. A graph of order n and size m is written as $G(n, m)$. A graph $G' = (V', E')$ is a subgraph of $G = (V, E)$ if $V' \subset V$ and $E' \subset E$. This is written as $G' \subset G$. A subgraph $G' \subset G$ is said to be induced by G if for every pair of vertices $x, y \in V(G')$, $\{x, y\} \in E(G') \iff \{x, y\} \in E(G)$.

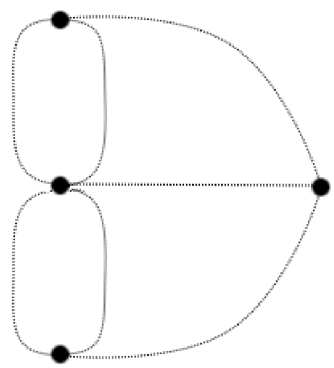

Figure 1: The city of Königsberg had seven bridges connecting four separate land masses. This can be illustrated with a graph where the land masses are represented by vertices and the bridges are represented by edges between them.

A graph G is called *multipartite* or *m-partite* if $V(G)$ is the disjoint union of m sets, called *vertex classes*, $V_1, V_2, ..., V_m$, such that every edge in G joins vertices from different vertex classes. This is denoted by $G_m(n_1, n_2, ..., n_m)$, where $n_i = |V_i|$. If all the vertex classes have the same number of vertices we may write $G_m(n)$. *Bipartite* means the same as 2-partite. A complete graph K is one where all the vertices are joined to each other. A complete graph of order n is denoted K^n. A complete m-partite graph has m vertex classes and for every vertex class all vertices are connected to every other vertex in the graph not in the same vertex class. We write this as $K_m(n_1, n_2, ..., n_m)$, where vertex class number i has n_i vertices. If all the n_i are the same we may write $K_m(n)$.

To *color* the vertices of a graph means to associate each vertex with some color (number). Suppose G is a graph with colored vertices using k colors. A vertex coloring of some graph G, is a way of coloring the vertices of G, such that no adjacent vertices share the same color.

Let's now define *hypergraphs* and some related concepts:

Definition 1.1 (i) Suppose V is a non-empty set of vertices and $E \subset V^{(n)}$, $n \in \mathbb{N}$, is a set of *hyperedges*, then the ordered pair (V, E) is called an *n-uniform hypergraph* or *n-graph*.

(ii) If $H = (V, E)$ is a hypergraph then $V(H) = V$ and $E(H) = E$.

(iii) If $E = \{e_1, e_2, ..., e_n\}$ is a hyperedge then the vertices $e_1, e_2, ..., e_n$ are *adjacent*.

(iv) Let $m \in \mathbb{N}, m \geq 1$. A hypergraph $H = (V, E)$ is m-partite if there is some function $f : V \to \{1, 2, ..., m\}$ such that for every $e \in E$ there are two vertices $e_1, e_2 \in e$ for which $f(e_1) \neq f(e_2)$.

A vertex coloring of a hypergraph is a way of coloring its vertices such that in every set of adjacent vertices (i.e. every edge) there is a pair of vertices with different colors.

Turán number and Turán graphs. We begin by defining the Turán number $ex(n, F)$ for a fixed r-uniform hypergraph (r-graph) F. This is the maximum number of edges

in an r-graph on n vertices such that it does not contain F. For $r = 2$ and $F = K^k$ the problem of finding this number is solved by Turán [2]. One example of K^k-free graphs are complete, *balanced*, $(k-1)$-partite graphs. One can construct these by partitioning the set of vertices into $k-1$ parts equal or nearly-equal in size (any two differing by at most 1) and pairwise connecting the appropriate vertices. Turán's theorem (see Theorem 2.1) states that this construction yields the largest number of edges for a K^k-free graph on n vertices. It is also unique. We call this graph the Turán graph and denote it by $T(n, k-1)$.
Let us find the size of $T(n, k-1)$. One can easily see that it has $n \bmod (k-1)$ vertex classes of size $\lceil \frac{n}{(k-1)} \rceil$ and $(k-1) - (n \bmod (k-1))$ vertex classes of size $\lfloor \frac{n}{(k-1)} \rfloor$, i.e.

$$T(n, k-1) = K_{k-1}(\lceil \frac{n}{(k-1)} \rceil, \lceil \frac{n}{(k-1)} \rceil, ..., \lfloor \frac{n}{(k-1)} \rfloor, \lfloor \frac{n}{(k-1)} \rfloor).$$

Lemma 1.2 *If $T(n, k-1)$ is a Turán graph, then*

$$e(T(n, k-1)) \leq \frac{(k-2)n^2}{2(k-1)}.$$

Proof. Let $p = n \bmod (k-1)$ and $q = (k-1) - p$, then the number of edges in $T(n, k-1)$ is

$$e(T(n, k-1)) = \binom{n}{2} - p\binom{\lceil \frac{n}{k-1} \rceil}{2} - q\binom{\lfloor \frac{n}{k-1} \rfloor}{2} =$$

$$= \frac{1}{2}\left(n^2 - n - p\left(\left\lceil \frac{n}{k-1} \right\rceil^2 - \left\lceil \frac{n}{k-1} \right\rceil\right) - q\left(\left\lfloor \frac{n}{k-1} \right\rfloor^2 - \left\lfloor \frac{n}{k-1} \right\rfloor\right)\right) =$$

$$= \frac{1}{2}\left(n^2 - n + p\left\lceil \frac{n}{k-1} \right\rceil + q\left\lfloor \frac{n}{k-1} \right\rfloor - p\left\lceil \frac{n}{k-1} \right\rceil^2 - q\left\lfloor \frac{n}{k-1} \right\rfloor^2\right) =$$

$$= \frac{1}{2}\left(n^2 - p\left\lceil \frac{n}{k-1} \right\rceil^2 - q\left\lfloor \frac{n}{k-1} \right\rfloor^2\right) \leq \frac{n^2}{2}\left(1 - \frac{1}{(k-1)}\right) = \frac{(k-2)n^2}{2(k-1)}$$

with equality when $k-1$ divides n. □

For arbitrary graphs F, we have no way of computing the Turán number exactly, but there is a result by Erdős and Stone [3] that provides an approximate answer.

2 Forbidden Subgraphs and the Number of Edges

In this section we look at a proof of Mantel's theorem and of Turán's theorem. Mantel's theorem tells us how many edges a graph can contain at the most and still be triangle-free and is actually a special case of Turán's theorem. Turán's theorem was proved in 1941 by Pál Turán and is one of the earliest and most well-known results in extremal graph theory.

Let us first define the term *k-clique*: A k-clique in a graph G is a complete graph K^k on k vertices.

We now look at a well known result in graph theory, the *theorem of Turán* [2][4]:

Theorem 2.1 (Turán) *Let $G(V,E)$ be a graph on n vertices without a k-clique, then*

$$|E| \leq \frac{(k-2)n^2}{2(k-1)}$$

We will first prove this theorem for the special case when $t = 3$, in other words:

Theorem 2.1′ (Mantel's theorem)[5] *If $G(V,E)$ is a graph containing no triangles, then*

$$|E| \leq \frac{n^2}{4}$$

For the proof we will need the *degree sum formula* (also called the *handshaking lemma*)

Lemma 2.2 (Degree sum formula) *Let $G(V,E)$ be a graph. Then*

$$\sum_{v\in V} d(v) = 2|E|.$$

Proof. To count the number of pairs $(v,e), v \in V, e \in E$ of incident vertices and edges in G one could count the number of incident edges to every vertex, i.e. $\sum_{v\in V} d(v)$. Since every edge connects exactly two vertices each edge will contribute two pairs of incident vertices and edges, i.e. counting this way the number of pairs is $2|E|$. Since the two expressions count the same thing they are equal

$$\sum_{v\in V} d(v) = 2|E|.$$

□

We are now ready to prove Mantel's theorem

Proof of Theorem 2.1′. Let $\{x,y\} \in E$. Since G contains no triangles, each vertex is joined to at most one of x and y (see figure 2), so we have $(d(x)-1)+(d(y)-1) \leq n-2$, thus $d(x)+d(y) \leq n$. Summing over all the edges we get

$$\sum_{\{x,y\}\in E} (d(x)+d(y)) \leq n|E|.$$

Note that every $d(x)$ appears $d(x)$ times in the sum, so

$$\sum_{x\in V} d(x)^2 = \sum_{\{x,y\}\in E} (d(x)+d(y)) \leq n|E|.$$

By the Cauchy-Schwarz inequality $(\sum x_i y_i)^2 \leq (\sum x_i^2)(\sum y_i^2)$ we get:

$$n^2|E| \geq n\sum_{x\in V} d(x)^2 \geq \left(\sum_{x\in V} d(x)\right)^2 = 4|E|^2$$

since $\sum_{x\in V} d(x) = 2|E|$ according to Lemma 2.2 and therefore $|E| \leq \frac{n^2}{4}$. □

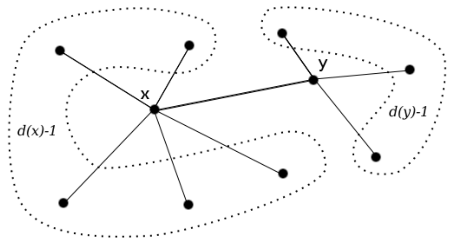

Figure 2: Figure showing possible edges containing x and y.

We will now prove the general case (Theorem 2.1):

Proof of Theorem 2.1. Let $G = (V, E)$ be a graph on n vertices without a k-clique with a maximal number of edges.

Claim 1. G does not contain any three vertices u, v, w such that $\{u, v\} \in E$, $\{u, w\} \notin E$ and $\{v, w\} \notin E$.

Proof by contradiction:

Case 1: $d(w) < d(u)$ or $d(w) < d(v)$. Without loss of generality we assume that $d(w) < d(u)$. Construct the graph G' from G by removing w and copying u (call it u') along with its edges from G. G' does not contain any k-clique since no clique contains both u and u' (they do not constitute an edge) and therefore the largest clique cannot possibly be larger for G' than for G. However $e(G') = e(G) - d(w) + d(u) > e(G)$ which contradicts the assumption about the size of G.

Case 2: $d(w) \geq d(u)$ and $d(w) \geq d(v)$ Similarly to case 1, create a new graph G'' by removing u and v and create two new copies of vertex w along with its edges. Like previously, G'' does not contain any k-clique, but $e(G'') = e(G) - (d(u) + d(v) - 1) + 2d(w) \geq e(G) + 1$

Claim 2. The number of edges in a complete m-partite graph is maximal when the number of elements in the vertex classes differ by at most 1.

Assume we have two vertex classes A and B such that $|A| > |B| + 1$, if we move one vertex v from A to B in such a way that the graph remains complete and m-partite (i.e. removing all edges containing $\{v, b\}, b \in B$ and adding new edges $\{v, a\}, a \in A$) then the number of edges increases; the graph loses $|B|$ edges, but gains $|A| - 1$ edges. Thus in total gaining $|A| - |B| - 1 \geq 1$ edges.

We have now proved both claims. Notice that claim 1 is equivalent to the statement that

$$u \sim v \iff \{u, v\} \notin E$$

defines an equivalence relation. Therefore we can partition the vertices of G into equivalence classes where two vertices are in the same equivalence class if they are not adjacent, implying that G is a complete, multipartite graph. There are obviously $(k-1)$-cliques in the graph, since otherwise we would have to add more edges satisfy the assumption of a maximal number of edges. The existence of a $(k-1)$-clique implies that there has to be at least $(k-1)$ vertex classes, since G is multipartite. On the other hand, there cannot be any larger cliques, so there are in fact exactly $(k-1)$ vertex classes. So, to summarize, any graph G without a k-clique and with a maximal number of edges will be complete, $(k-1)$-partite and the vertex classes will differ in size by at most 1. This describes a Turán graph and thus proves that the Turán graph $T(n, k-1)$ is the unique graph with the maximal number of edges without a k-clique, and since according to Lemma 1.2 we know that $e(T(n, k-1)) \leq \frac{(k-2)n^2}{2(k-1)}$ the proof of Theorem 2.1 is complete. □

3 Extremal Graph Theory

In this section we look at some results in extremal graph theory that are similar to Turán's theorem in that they deal with questions regarding maximal sizes of graphs that do not contain some forbidden subgraph. We also prove a theorem by Bollobás, Erdős and Simonovits that gives a lower bound on the size of a graph that is guaranteed to contain some subgraph.

Given m, n, s and t, $z(m, n; s, t)$ denotes the maximal size of a graph $G_2(m, n)$ not containing a $K_2(s, t)$. The problem of determining z is known as the *Zarankiewicz problem.*

We prove the following theorems from [6]:

Theorem 3.1 $z(m, n; s, t) < (s-1)^{1/t}(n-t+1)m^{1-1/t} + (t-1)m$

We will use the following lemma to prove this theorem by *Kővári-Sós-Turán* [7]:

Lemma 3.2 *Let m, n, s, t be integers, $0 \leq s \leq m$, $0 \leq t \leq n$ and let $G = G_2(m, n)$ be a graph without a $K(s, t)$. Then*

$$m\binom{e(G)/m}{t} \leq (s-1)\binom{n}{t}$$

Proof. Denote by V_1, V_2 the vertex classes of G. A t-set T of V_2 *belongs* to $x \in V_1$ if x is joined to every vertex in T. Then $\binom{d(x)}{t}$ t-sets belong to a vertex $x \in V_1$ and by assumption a t-set in V_2 belongs to at most $s-1$ vertices of V_1. Thus

$$\sum_{x \in V_1} \binom{d(x)}{t} \leq (s-1)\binom{n}{t}. \quad (1)$$

Using *Jensen's inequality* [8]:

$$f\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right) \leq \frac{1}{n}\sum_{k=1}^{n} f(x_k),$$

where f is a convex function we get:

$$m\frac{1}{m}\sum_{x \in V_1}\binom{d(x)}{t} \geq m\binom{\frac{1}{m}\sum_{x \in V_1} d(x)}{t} = m\binom{e(G)/m}{t}$$

Together with (1) this proves the lemma. □

We will now prove Theorem 3.1

Proof of Theorem 3.1. Applying Lemma 3.2 to an extremal graph of size $z(m, n; s, t)$ yields

$$m\binom{e(G)/m}{t} \leq (s-1)\binom{n}{t}.$$

We have the bound $\frac{(n-(k-1))^k}{k!} \leq \binom{n}{k}$, which decreases when n decreases. Since $e(G)/m \leq n$ we get

$$m\frac{(e(G)/m-(t-1))^t}{t!} \leq (s-1)\frac{(n-(t-1))^t}{t!} \quad ,$$

hence

$$(e(G)/m-(t-1))^t \leq \frac{(s-1)(n-(t-1))^t}{m}$$

so

$$e(G) \leq (s-1)^{1/t}(n-t+1)m^{1-1/t}+(t-1)m$$

□

Theorem 3.3 *Suppose G is a graph of order n not containing a $K_2(s,t)$, $2 \leq s, 2 \leq t$. Then*

$$2e(G) \leq z(n, n; s, t)$$

so

$$e(G) \leq \frac{1}{2}(s-1)^{1/t}(n-t+1)n^{1-1/t}+\frac{1}{2}(t-1)n.$$

Proof. Construct a graph $G' = G_2(n, n)$ from G as follows. Take two copies, V_1 and V_2, of $V(G)$. If $\{x, y\} \in E(G)$ and $x' \in V_1$ and $y' \in V_2$ are the vertices corresponding to x and y respectively, then let $\{x', y'\}$ be an edge in G'. Then $G' = G_2(n, n)$ has $2e(G)$ edges and does not contain a $K(s,t)$ so $2e(G) \leq z(n, n; s, t)$. Using Theorem 3.1, $e(G) \leq \frac{1}{2}(s-1)^{1/t}(n-t+1)n^{1-1/t}+\frac{1}{2}(t-1)n$. □

Theorem 3.4 [9] [10]

(i) If $2 \leq t < n$ then

$$z(n, n; t, t) < (t-1)^{1/t}n^{2-1/t}+\frac{t-1}{2}n.$$

(ii) If a graph of order n does not contain a $K_2(t,t)$ then its size is at most

$$\frac{1}{2}\left((t-1)^{1/t}n^{2-1/t}+\frac{t-1}{2}n\right).$$

Proof. Part (ii) follows from part (i) by the same reasoning as in Theorem 3.3. To prove (i) it is, by Lemma 3.2, enough to show that, for an extremal graph $G_2(n, n)$ of size $z(n, n; t, t)$ if

$$U = (t-1)^{1/t}n^{1-1/t}+\frac{t-1}{2} = e(G)/n \tag{2}$$

then

$$n\binom{U}{t} > (t-1)\binom{n}{t} \tag{3}$$

for any integer $2 \le t < n$.

We look at two cases:

1. If t is even then (3) can be written as the product of the t factors

$$(n^{1/t}U)(n^{1/t}(U-1))(n^{1/t}(U-2))\dots(n^{1/t}(U-\frac{1}{2}(t-2)))(n^{1/t}(U-\frac{1}{2}t))\dots(n^{1/t}(U-t+1)) >$$

$$((t-1)^{1/t}n)\dots((t-1)^{1/t}(n-\frac{1}{2}(t-2)))((t-1)^{1/t}(n-\frac{1}{2}t))\dots((t-1)^{1/t}(n-t+1)).$$

Multiplying the first factor with the last factor, the second with the next to last and so on we get:

$$(n^{2/t}U(U-t+1))(n^{2/t}(U-1)(U-t+2))\dots(n^{2/t}(U-\frac{1}{2}(t-2))(U-\frac{1}{2}t)) >$$

$$((t-1)^{2/t}n(n-t+1))((t-1)^{2/t}(n-1)(n-t+2))\dots((t-1)^{2/t}(n-\frac{1}{2}(t-2))(n-\frac{1}{2}t)).$$

By the condition $t < n$ together with (2) we have that:

$$U > (t-1)^{1/t}(t-1)^{1-1/t} + \frac{t-1}{2} = t-1+\frac{t-1}{2},$$

so all factors on both sides of the inequality are positive. Therefore we only need to show that the r:th factor on the left-hand side is larger than the r:th factor on the right-hand side, where $1 \le r \le t/2$ (there are $t/2$ factors on each side), to prove that the inequality holds. I.e.

$$n^{2/t}(U-r+1)(U-t+r) > (t-1)^{2/t}(n-r+1)(n-t+r) \tag{4}$$

should hold for $1 \le r \le t/2$. From (2) we get

$$n^{2/t}((t-1)^{1/t}n^{1-1/t} + \frac{t-1}{2} - r + 1)((t-1)^{1/t}n^{1-1/t} - \frac{t-1}{2} - r - 1)$$
$$> (n-r+1)(n-t+r)(t-1)^{2/t}.$$

Notice that the left-hand side is $n^{2/t}(a+b)(a-b)$, where $a = (t-1)^{1/t}n^{1-1/t}$ and $b = \frac{t-1}{2} - r + 1$, so the inequality becomes

$$(t-1)^{2/t}n^2 > n^{2/t}\left(\frac{t-1}{2} - r + 1\right)^2 + (n-r+1)(n-t+r)(t-1)^{2/t}.$$

Since $(t-1)^{2/t}(n-r+1)(n-t+r) = (t-1)^{2/t}(r-1)(t-r)+(t-1)^{2/t}n^2-(t-1)^{1+2/t}n$, we get

$$(t-1)^{1+2/t}n > n^{2/t}\left(\frac{t-1}{2} - r + 1\right)^2 + (r-1)(t-r)(t-1)^{2/t}. \tag{5}$$

The following inequaltiy holds (because $r \le t/2$):

$$(r-1)(t-r)(t-1)^{2/t} < \frac{(t-1)^2(t-1)^{2/t}}{2} < \frac{(t-1)^2 n^{2/t}}{2} \tag{6}$$

Since $2 \leq j < n$, we have $n^{1-2/t} \geq t^{1-2/t}$. Multiplying by $n^{2/t}(t-1)^{1+2/t}$ we get

$$(t-1)^{1+2/t}n \geq (t-1)^{1+2/t}t^{1-2/t}n^{2/t} \geq (t-1)^2 n^{2/t} > \frac{(t-1)^2}{4}n^{2/t} + \frac{(t-1)^2}{2}n^{2/t}$$

$$\geq n^{2/t}\left(\frac{t-1}{2} - r + 1\right)^2 + \frac{(t-1)^2 n^{2/t}}{2}.$$

Together with (6) this proves (5).

2. If t is odd the proof is similar. Write (3) as

$$(n^{1/t}U)(n^{1/t}(U-1))\dots(n^{1/t}(U-\frac{1}{2}(t-1)))(n^{1/t}(U-\frac{1}{2}(t+1)))\dots(n^{1/t}(U-t+1)) >$$

$$((t-1)^{1/t}n)\dots((t-1)^{1/t}(n-\frac{1}{2}(t-1)))((t-1)^{1/t}(n-\frac{1}{2}(t+1)))\dots((t-1)^{1/t}(n-t+1)).$$

Multiply as before:

$$(n^{2/t}U(U-t+1))\dots(n^{2/t}(U-\frac{1}{2}(t-3))(U-\frac{1}{2}(t+1)))(n^{1/t}(U-\frac{1}{2}(t-1))) >$$

$$((t-1)^{2/t}n(n-t+1))\dots((t-1)^{2/t}(n-\frac{1}{2}(t-3))(n-\frac{1}{2}(t+1)))((t-1)^{1/t}(n-\frac{1}{2}(t-1)).$$

Notice the last factor on both sides of the inequality; they are the factors that could not be paired up because t is odd. Using the same reasoning as before we see that it is enough to show that the r:th factor on the left-hand side is larger than the r:th factor on the right-hand side where $1 \leq r \leq (t-1)/2$ *and* that the last factor ($r = (t-1)/2+1$) of the expression on the left is greater than the last factor of the expression on the right to prove that the inequality holds. I.e. if

$$n^{2/t}(U-r+1)(U-t+r) > (t-1)^{2/t}(n-r+1)(n-t+r) \tag{7}$$

holds for $1 \leq r \leq (t-1)/2$ and

$$\left(n^{1/t}(U-\frac{1}{2}(t-1))\right) > \left((t-1)^{1/t}(n-\frac{1}{2}(t-1))\right) \tag{8}$$

is true, then the inequality (3) is true for odd numbers.

Let's begin by proving (8). Using (2) the right-hand side becomes:

$$n^{1/t}\left((t-1)^{1/t}n^{1-1/t} + \frac{t-1}{2} - \frac{t-1}{2}\right) = (t-1)^{1/t}n > (t-1)^{1/t}(n-\frac{1}{2}(t-1)).$$

See (4) for the proof of (7).

We have now proved that (3) holds for any integer $2 \leq t < n$. The proof of Theorem 3.4 is therefore complete.

□

Theorem 3.5 *[11] There is an absolute constant $\alpha > 0$ such that if $0 < \varepsilon < \frac{1}{r}$ and*

$$m > \left(1 - \frac{1}{r} + \varepsilon\right)\frac{n^2}{2} \tag{9}$$

then every $G(n, m)$ contains a $K_{r+1}(t)$, where

$$t = \left\lfloor \frac{\alpha \log n}{r \log 1/\varepsilon} \right\rfloor \tag{10}$$

Proof. Let us first look at the case where $r = 1$. Choose $\alpha = 1$ and suppose for contradiction that there are arbitrarily large values of n for which some graph $G(n, m)$ satisfies (9) *without* containing a $K_2(t)$, where t is as in (10). Then, by Theorem 3.4

$$\varepsilon n^2 \leq (t-1)^{1/t} n^{2-1/t} + tn$$

and so, for sufficiently large n

$$2 \leq \varepsilon n^{1/t} \leq (t-1)^{1/t} + tn^{-1+1/t} < 2$$

which is obviously a contradiction, so $\alpha = 1$ will do for large n. This implies the existence of the appropriate constant for $r = 1$. In the rest of the proof we will assume that $r \geq 2$.

We will use the following lemma to replace (10) with a more convenient condition:

Lemma 3.6 *Let $\frac{1}{2} > c > \varepsilon > 0$. Suppose $n > 2c/\varepsilon$ and $e(G^n) > cn^2$. Then G^n contains a subgraph G' of order p such that*

$$p > \varepsilon^{1/2} n$$

$$\delta(G') > 2(c - \varepsilon)p$$

and

$$e(G') > e(G^n) - (c - \varepsilon)(n - p)(n + p + 1).$$

For sufficiently large n we may require that

$$p > (2\varepsilon)^{1/2} n.$$

Proof. Let us define a sequence of graphs $S_n = G^n, G_{n-1}, G_{n-2}, \ldots$ in the following manner: If $\delta(G_k) > 2(c - \varepsilon)k$ then stop the sequence with G_k. Otherwise choose some vertex $x_k \in G_k$ with $d_{G_k} \leq 2(c-\varepsilon)k$ and put $G_{k-1} = G_k \setminus x_k$. This will generate a finite sequence (S_n contains graphs G_i where $p \leq i \leq n$, n is finite and $0 < p \leq n$) where the member G_i contains i vertices. The last member is a graph G_p with smallest degree $\delta(G_p) > 2(c - \varepsilon)k$. Put $G' = G_p$. We have

$$\binom{p}{2} \geq e(G_p) \geq e(G^n) - 2(c - \varepsilon) \sum_{k=p+1}^{n} k.$$

The first inequality holds since there is obviously at most one edge between every pair of vertices. The second inequality is true because in the construction of the sequence we remove $n - p$ vertices x_k, where $p + 1 \leq k \leq n$, from the original graph G^n to get to G_p.

These vertices have degree at most $2(c-\varepsilon)k$ each. So the number of edges we remove is at most:

$$\sum_{k=p+1}^{n} 2(c-\varepsilon)k = 2(c-\varepsilon)\sum_{k=p+1}^{n} k.$$

We see that the sum is an arithmetic series, so we can rewrite the lower bound for $e(G_p)$ as:

$$\begin{aligned} e(G^n) - 2(c-\varepsilon)\sum_{k=p+1}^{n} k &= e(G^n) - (c-\varepsilon)(n-p)(n+p+1) \\ &= e(G^n) - (c-\varepsilon)(n-p)(n+p) - (c-\varepsilon)(n-p) \\ &= e(G^n) - (c-\varepsilon)(n^2-p^2) - (c-\varepsilon)(n-p) \\ &= e(G^n) - c(n^2-p^2) + \varepsilon(n^2-p^2) - (c-\varepsilon)(n-p) \\ &> cn^2 - cn^2 + cp^2 + \varepsilon(n^2-p^2) - (c-\varepsilon)(n-p) \\ &= cp^2 + \varepsilon(n^2-p^2) - (c-\varepsilon)(n-p). \end{aligned}$$

Now we have that

$$\binom{p}{2} = \frac{p(p-1)}{2} \geq cp^2 + \varepsilon(n^2-p^2) - (c-\varepsilon)(n-p)$$

Rearranging we get

$$p^2 - p - 2cp^2 + 2\varepsilon p^2 \geq 2\varepsilon n^2 - 2(c-\varepsilon)(n-p)$$

and finally

$$p^2(1-2c+2\varepsilon) \geq 2\varepsilon n^2 - 2(c-\varepsilon)(n-p) + p > 2\varepsilon n^2 - 2cn,$$

which implies both lower bounds on p when n is sufficiently large.

□

To prove Theorem 3.5 we will first show that by Lemma 3.6 it is enough to prove the following theorem (3.7). We then prove this reformulated theorem instead.

Theorem 3.7 *If $0 < \beta < 1$ is a sufficiently small absolute constant, $2 \leq r$, $0 < \varepsilon < 1/r$, and a graph G of order n satisfies*

$$\delta(G) \geq \left(1 - \frac{1}{r} + \varepsilon\right) n \tag{11}$$

then G contains a $K_{r+1}(M)$, where

$$M = \left\lfloor \frac{\beta \log n}{r \log 1/\varepsilon} \right\rfloor. \tag{12}$$

To go from Theorem 3.5 to Theorem 3.7 let us suppose that $0 < \varepsilon < \frac{1}{r}, r \geq 2$ and $m > (1 - \frac{1}{r} + \varepsilon)\frac{n^2}{2}$. Let $G = G(n,m)$ and

$$c = \frac{1 - 1/r + \varepsilon}{2}.$$

Then $e(G) > cn^2$ and $\frac{1}{2} > c > \frac{\varepsilon}{4} > 0$. Applying Lemma 3.6 we see that G contains a subgraph G' of order p, where

$$p > \frac{\varepsilon^{1/2}}{2}n$$

$$\delta(G') > 2(c - \frac{\varepsilon}{4})p$$

and

$$e(G') > e(G) - (c - \frac{\varepsilon}{4})(n-p)(n+p+1)$$

Hence

$$\delta(G') > 2\left(\frac{1 - 1/r + \varepsilon}{2} - \frac{\varepsilon}{4}\right)p = \left(1 - \frac{1}{r} + \varepsilon - \frac{\varepsilon}{2}\right)p = \left(1 - \frac{1}{r} + \frac{\varepsilon}{2}\right)p$$

If we assume that Theorem 3.7 is true, then there is some absolute constant $0 < \beta < 1$ such that G' contains a $K_{r+1}(M)$, where

$$M = \left\lfloor \frac{\beta \log p}{r \log 2/\varepsilon} \right\rfloor = \left\lfloor \frac{\beta \log p}{r(\log 2 + \log(1/\varepsilon))} \right\rfloor > \left\lfloor \frac{\beta \log p}{r(1 + \log(1/\varepsilon))} \right\rfloor = \left\lfloor \frac{\beta \log p}{r + r\log(1/\varepsilon)} \right\rfloor >$$

$$\left\lfloor \frac{\beta \log \frac{\varepsilon^{1/2}}{2} n}{r + r\log(1/\varepsilon)} \right\rfloor \geq \left\lfloor \frac{\alpha \log n}{r \log(1/\varepsilon)} \right\rfloor$$

for large enough n if $\alpha > 0$ is chosen small enough (depending only on β and r).

I.e. if we make the required assumptions in Theorem 3.5 and also assume that Theorem 3.7 is true, then Theorem 3.5 holds. Therefore it is enough to show that Theorem 3.7 is true to prove Theorem 3.5. The proof follows below:

Proof (continued).

If $M < 1$ the theorem is trivial, so we assume that $M \geq 1$. From (12) we have that

$$n \geq (1/\varepsilon)^{r/\beta} \tag{13}$$

We will now use induction on r. If $r = 1$ then $\delta(G) \geq \varepsilon n$ so according to Lemma 2.2

$$|E| = \frac{1}{2}\sum_{v \in V(G)} d(v) \geq \frac{1}{2}\sum_{v \in V(G)} \varepsilon n = \frac{\varepsilon n^2}{2}.$$

Using Theorem 3.5 (which we have shown to be true for $r = 1$) we see that the result holds. Let us now suppose that $r \geq 2$ and the result holds for smaller values of r.

Put

$$\varepsilon' = \frac{1}{r-1} - \frac{1}{r} + \varepsilon$$

and

$$p_0 = \left\lfloor \frac{\beta \log n}{(r-1)\log(1/\varepsilon')} \right\rfloor \tag{14}$$

Since by assumption

$$\delta(G) \geq \left(1 - \frac{1}{r} + \varepsilon\right)n = \left(1 - \frac{1}{r-1} + \varepsilon'\right)n$$

by the induction hypothesis G contains a $K_r(p_0)$. We will make use of this fact later but first we need to do some work.

We will need the following lemma:

Lemma 3.8 *Let $X \subset V(G)$ and denonte by Y the set of those vertices of $V(G) \setminus X$ that are joined to at least $(1 - 1/r + \varepsilon/2)|X|$ vertices of X. Then*

$$\frac{1}{2} r\varepsilon n - |X| < |Y|.$$

Proof. Denote by S the number of edges joining vertices in X to the vertices in $V(G) \setminus X$. Then

$$|X| \left(\left(1 - \frac{1}{r} + \varepsilon \right) n - |X| \right) \leq S$$

since $\delta(G) \geq (1 - 1/r + \varepsilon)n$ and for each vertex $x \in X$ the number of edges $\{x, y\} \in G$, $y \in X$ is less than $|X|$. We also have

$$S \leq |Y||X| + (n - |X| - |Y|) \left(1 - \frac{1}{r} + \frac{\varepsilon}{2} \right) |X|$$

The first term on the right-hand side of the inequality, $|Y||X|$, is the maximal number of edges between the vertices in Y and the vertices in X and is greater than or equal to the actual number of edges between the two sets. The second term, $(n - |X| - |Y|)\,(1 - 1/r + \varepsilon/2)\,|X|$ is the number of edges between $Z = G \setminus X \setminus Y$ and X if each vertex in Z was joined to $(1 - 1/r + \varepsilon/2)|X|$ vertices in $|X|$. But the actual number of edges between these two sets is less, since each of the vertices in Z is joined to fewer than $(1 - 1/r + \varepsilon/2)|X|$ vertices of X. So the sum of these terms is greater than or equal to $|S|$. Finally

$$|X| \left(\left(1 - \frac{1}{r} + \varepsilon \right) n - |X| \right) \leq |Y||X| + (n - |X| - |Y|) \left(1 - \frac{1}{r} + \frac{\varepsilon}{2} \right) |X|,$$

simplifies to

$$\frac{\varepsilon n}{2} - \frac{|X|}{r} + \frac{\varepsilon |X|}{2} \leq |Y| \left(\frac{1}{r} - \frac{\varepsilon}{2} \right)$$

so

$$\frac{\varepsilon n}{2} - \frac{|X|}{r} + \frac{\varepsilon |X|}{2} < |Y| \frac{1}{r},$$

therefore

$$\frac{1}{2} r\varepsilon n - |X| < |Y|.$$

□

Let us return to the proof of Theorem 3.7. Put $P = \lceil (2/\varepsilon) M \rceil < (3/\varepsilon) M$. Consider subgraphs $K_r(p_1, p_2, ..., p_r)$ of G such that

$$p_i \leq p + M, 1 \leq i \leq r, \tag{15}$$

where

$$p = \frac{1}{r} \sum_{i=1}^{r} p_i.$$

The rest of the proof is in two parts. First we will show that if $p \geq P$ then G does contain a $K_{r+1}(M)$. Then we show that it cannot be true for every $K_r(p_1, p_2, ..., p_r)$ satisfying (15) that $p < P$ and therefore G does contain a $K_{r+1}(M)$.

Part 1. For the first part, suppose that G contains a $K = K_r(p_1, p_2, ..., p_r)$ such that (15) holds and

$$p \geq P.$$

We may in fact assume that

$$P \leq p \leq P + 1, \tag{16}$$

by removing some vertices of K if necessary. From (13) we get

$$\varepsilon n \geq \varepsilon(1/\varepsilon)^{r/\beta} > 4(P + 1) \tag{17}$$

for small enough β.

If W denotes the set of vertices of $G \setminus K$ joined to at least

$$\left(1 - \frac{1}{r} + \frac{\varepsilon}{2}\right) rp = p(r-1) + \frac{\varepsilon pr}{2}$$

vertices of K then, by Lemma 3.8,

$$|W| > \frac{r}{2}\varepsilon n - |K| \geq \frac{r}{2}\varepsilon n - pr = \frac{r}{2}(\varepsilon n - 2p) > \frac{\varepsilon rn}{4}, \tag{18}$$

since $2(P + 1) < \frac{\varepsilon n}{2}$ according to (17) and (16) gives us that $p \leq (P + 1)$, we can see that $2p < \frac{\varepsilon n}{2}$, so

$$\frac{r}{2}(\varepsilon n - 2p) > \frac{\varepsilon rn}{4}.$$

By p_i we sometimes refer to the actual number p_i but other times we mean the vertex class with which this number is associated (being its size). The meaning should be clear from context.

Now, consider the vertex $v \in W$ and the vertex class p_m of K such that the number of edges between v and vertices in p_m is minimal. There are $r - 1$ vertex classes that are not p_m and the number of nodes in each class is at most $p + M$. Therefore v can be adjacent to a maximum of $(r - 1)(p + M)$ vertices not in p_m. Since we know that v has to form $p(r - 1) + \frac{\varepsilon pr}{2}$ edges with vertices in K we can deduce the following:

The number of vertices in any given vertex class p_i of K that v is joined to is at least

$$p(r - 1) + \frac{\varepsilon pr}{2} - (r - 1)(p + M).$$

By (16), together with the definition of P we have

$$p(r - 1) + \frac{\varepsilon pr}{2} - (r - 1)(p + M) \geq \frac{\varepsilon Pr}{2} - (r - 1)M \geq rM - (r - 1)M = M.$$

I.e. every vertex v of W is adjacent to at least M vertices of every vertex class p_i of K. So, for every vertex v of W there is a complete subgraph $K_r(M) \subset K$ such that v is joined to each vertex of this subgraph. The number of such subgraphs of K is at most

$$\binom{P+1+M}{M}^r, \tag{19}$$

since there are at most $P+1+M$ vertices in each vertex class and we choose M of these to be a vertex class of the complete subgraph and there are r vertex classes.

The Maclaurin series expansion of the exponential function is

$$e^k = \sum_{i=0}^{\infty} \frac{k^i}{i!}$$

so

$$e^k = \sum_{i=0}^{\infty} \frac{k^i}{i!} < \frac{k^k}{k!},$$

since every term in the sum is positive. Therefore

$$k! < \frac{k^k}{e^k}.$$

Thus

$$\binom{n}{k} \leq \frac{n^k}{k!} < \left(\frac{en}{k}\right)^k. \tag{20}$$

Now we can apply these inequalities to (19):

$$\binom{P+1+M}{M}^r < \binom{2P}{M}^r < \left(\frac{2eP}{M}\right)^{rM},$$

since $P = \lceil (2/\varepsilon)M \rceil \leq (3/\varepsilon)M$, we have

$$\binom{P+1+M}{M}^r < \binom{2P}{M}^r < \left(\frac{2eP}{M}\right)^{rM} < \left(\frac{2e(3/\varepsilon)M}{M}\right)^{rM} = \left(\frac{6e}{\varepsilon}\right)^{rM} \leq \left(\frac{6e}{\varepsilon}\right)^{\frac{\beta \log n}{\log 1/\varepsilon}}$$

Let $c = 6e/\varepsilon$. Then $c > 1$, since $\varepsilon < 1$. So

$$\left(\frac{6e}{\varepsilon}\right)^{\log n} = c^{\frac{\log_c n}{\log_c e}} = n^{(\log_c e)^{-1}} = n^{\log 6e/\varepsilon},$$

since

$$\log_c e = \frac{\log_e e}{\log_e c} = \frac{1}{\log \frac{6e}{\varepsilon}}.$$

Therefore

$$\left(\frac{6e}{\varepsilon}\right)^{\frac{\beta \log n}{\log 1/\varepsilon}} = \left(\left(\frac{6e}{\varepsilon}\right)^{\log n}\right)^{\frac{\beta}{\log 1/\varepsilon}} = n^{\frac{\beta \log 6e/\varepsilon}{log 1/\varepsilon}} < n^{\zeta\beta}$$

for some arbitrary constant

$$\zeta > \frac{\log 6e/\varepsilon}{\log 1/\varepsilon} \quad \text{(where } 0 < \varepsilon < 1/r \text{ is constant)}$$

Obviously $n^{\zeta\beta}$ decreases with β, and therefore, for small enough β we have

$$n^{\zeta\beta} < \frac{\varepsilon n \log 2}{\beta \log n}$$

Notice that $r^2/4 \geq 1$, thus

$$n^{\zeta\beta} < \frac{\varepsilon n \log 2}{\beta \log n} < \frac{r^2 \varepsilon n \log(1/\varepsilon)}{4\beta \log n} \leq \frac{r\varepsilon n}{4M} \text{ and by (18)}$$

$$\frac{r\varepsilon n}{4M} \leq \frac{|W|}{M}.$$

To recap; for every vertex v of W there is a complete subgraph $K_r(M) \subset K$ such that v is joined to each vertex of this subgraph. The number of such subgraphs is at most

$$\binom{P+1+M}{M}^r < n^{\zeta\beta} < \frac{|W|}{M}, \tag{21}$$

so if there are n such subgraphs, then there has to be more than nM vertices in W. Since every vertex $v \in W$ has to be joined to every vertex in one of the complete subgraphs, there obviously will be at least M vertices from W that are joined in this way to at least one complete subgraph. Thus W contains a subset W' of M vertices and K contains a $K_r(M)$ subgraph K' such that every vertex of W' is joined to every vertex of K'. Since none of the vertices in W are in K (and consequently K'), W' can form a new vertex class in a complete graph $K_{r+1}(M)$ consisting of the vertices in K' and W'.

This proves that if $p \geq P$ then G does contain a $K_{r+1}(M)$ and thereby the first part of the proof is complete.

Now all we need to prove is that it cannot be true for every $K_r(p_1, p_2, ..., p_r)$ satisfying (15) that $p < P$ and therefore G does contain a $K_{r+1}(M)$.

Part 2. We will assume that $p < P$ for every $K_r(p_1, p_2, ..., p_r)$ satisfying (15) and show that this leads to a contradiction.

Let $K = K_r(p_1, \ldots, p_r)$ be a subgraph of G for which p is maximal under the conditions in (15). We know from (14) that G contains a subgraph $K_r(p_0) = \left\lfloor \frac{\beta \log n}{(r-1)\log(1/\varepsilon')} \right\rfloor$ and since $\varepsilon' > \varepsilon$, $M < p_0$. It is also true that $p_0 \leq p$, since p is maximal and $p < P$ according to our assumption. Thus

$$M < p_0 \leq p < P. \tag{22}$$

Denote by $U \subset G$ the set of vertices that are each joined to at least M vertices of each vertex class of K. Suppose first that $|U|$ is large, say $|U| \geq n^{1/2}$. Then by the same reasoning as in part 1, the number of $K_r(M)$ subgraphs of K is at most

$$\binom{p+M}{M}^r < n^{\zeta\beta} < \frac{n^{1/2}}{M} < \frac{|U|}{M}$$

according to (21), if β is sufficiently small. Then, just as in part 1, there are M vertices in U that are each joined to every vertex of a $K_r(M) \subset K$, so G does contain a $K_{r+1}(M)$. We will now show that if we assume that $|U| < n^{1/2}$ there will be a contradiction. So either $|U| \geq n^{1/2}$ or $p \geq P$. Regardless, we will have shown that there exists a complete r-partite subgraph $K_{r+1}(M)$ of G and the proof will be complete. Let us continue.

Assume $|U| < n^{1/2}$. As in part 1 denote by W the set of vertices of $G \setminus K$ such that every vertex of W is joined to at least

$$\left(1 - \frac{1}{r} + \frac{\varepsilon}{2}\right) rp \tag{23}$$

vertices of K. Then, as seen in (18) $|W| > \varepsilon rn/4$. Let $V = W \setminus U$.

$$|V| \geq |W| - |U| > \frac{\varepsilon rn}{4} - n^{1/2},$$

so if β is sufficiently small, since $n \geq (1/\varepsilon)^{r/\beta}$ then

$$|V| > \frac{\varepsilon rn}{5}. \tag{24}$$

Recall that the set of vertices adjacent to a vertex x is denoted by $\Gamma(x)$. We define an equivalence relation on V by

$$x \sim y \iff \Gamma(x) \cap K = \Gamma(y) \cap K,$$

i.e. for two vertices x and y where $x, y \in V$, $x \sim y$ if x is adjacent to exactly the same vertices of K as y is. We now want to show that there exists an equivalence class containing at least $\lfloor p + M \rfloor$ elements. The proof of this will rely on showing that there are relatively few equivalence classes.

Let C_i denote the ith vertex class of K. For every $x \in V$ there exists some index i_0, $1 \leq i_0 \leq r$, such that x is joined to fewer than M vertices of C_{i_0} (this follows from the definition of V, where we remove U). As x is joined to more than

$$\left(1 - \frac{1}{r}\right) rp = (r-1)p \tag{25}$$

vertices of K according to (23), the number of vertices in the remaining classes $K \setminus C_{i_0}$ that are *not* joined to x are fewer than

$$(r-1)(p+M) - ((r-1)p - M) = rM,$$

since there are at most $(r-1)(p+M)$ nodes in $K \setminus C_{i_0}$ according to (15) and x needs to be joined to more than $(r-1)p - M$ more vertices to satisfy (23). For any given $x \in V$ there are at most

$$\sum_{\lambda < M} \binom{p_{i_0}}{\lambda}$$

ways that x could be joined to vertices in C_{i_0}, since there are p_{i_0} vertices in C_{i_0} and x is joined to at most M of them. There are at most

$$\sum_{\mu < rM} \binom{rp - p_{i_0}}{\mu}$$

ways to choose vertices from $K \setminus C_{i_0}$ that are not joined to x. After choosing which vertices in C_{i_0} that x is joined to and which vertices in $K \setminus C_{i_0}$ that x is not joined to we have implicitly decided that x is joined to the remaining vertices in K. Therefore if the same choices of i_0, adjacent vertices in C_{i_0} and non-adjacent vertices in $K \setminus C_{i_0}$ are made for two different vertices in V they will belong to the same equivalence class. Since there are r possible values of i_0, the number of equivalence classes in V is less than

$$\sum_{i=1}^{r} \left\{ \sum_{\lambda<M} \binom{p_i}{\lambda} \sum_{\mu<rM} \binom{rp-p_i}{\mu} \right\} < \sum_{i=1}^{r} \left\{ \sum_{\lambda<M} \binom{2p}{M} \sum_{\mu<rM} \binom{rp}{rM} \right\},$$

since $p_i \leq p + M$ and according to (22), $M < p$. Furthermore

$$\sum_{i=1}^{r} \left\{ \sum_{\lambda<M} \binom{2p}{M} \sum_{\mu<rM} \binom{rp}{rM} \right\} \leq r \left\{ M \binom{2p}{M} rM \binom{rp}{rM} \right\} = r^2 M^2 \binom{2p}{M} \binom{rp}{rM}$$

$$< r^2 M^2 \binom{2p}{M} \binom{rp}{rM}$$

using the inequalities derived earlier (20) we get

$$r^2 M^2 \binom{2p}{M} \binom{rp}{rM} < r^2 M^2 \left(\frac{2ep}{M}\right)^M \left(\frac{erp}{rM}\right)^{rM} = r^2 M^2 e^{rM+M} \left(\frac{2p}{M}\right)^M \left(\frac{rp}{rM}\right)^{rM}$$

$$< r^2 M^2 e^{M(r+1)} \left(\frac{2P}{M}\right)^M \left(\frac{P}{M}\right)^{rM}$$

and since $P < (3/\varepsilon)M$

$$r^2 M^2 e^{M(r+1)} \left(\frac{2P}{M}\right)^M \left(\frac{P}{M}\right)^{rM} < r^2 M^2 e^{M(r+1)} \left(\frac{6}{\varepsilon}\right)^M \left(\frac{3}{\varepsilon}\right)^{rM} <$$

$$r^2 M^2 e^{M(r+1)} \left(\frac{e^2}{\varepsilon}\right)^M \left(\frac{e^2}{\varepsilon}\right)^{rM} = r^2 M^2 e^{M(r+1)} \left(\frac{e^2}{\varepsilon}\right)^{M(r+1)} =$$

$$r^2 M^2 e^{3M(r+1)} \left(\frac{1}{\varepsilon}\right)^{M(r+1)} < r^2 M^2 e^{5rM} \left(\frac{1}{\varepsilon}\right)^{2rM},$$

since $r \geq 2$. From the definition of M it follows that

$$rM \leq \frac{\beta \log n}{\log(1/\varepsilon)}.$$

We also have the following bounds

$$0 < \varepsilon < \frac{1}{2},$$

which means that $\frac{1}{\varepsilon} > 2$ and therefore

$$\frac{1}{\log(1/\varepsilon)} < \frac{1}{\log(2)} < \frac{1}{0.6} < 2,$$

so

$$rM < 2\beta \log n.$$

Thus the following two inequalities hold:

$$r^2M^2 < 4\beta^2(\log n)^2$$

and

$$e^{5rM} < e^{10\beta \log n} = n^{10\beta}.$$

Let K be some constant such that $e^K \geq 1/\varepsilon$. Then

$$r^2M^2e^{5rM}\left(\frac{1}{\varepsilon}\right)^{2rM} < 4\beta^2(\log n)^2n^{10\beta}e^{2rMK} < 4\beta^2(\log n)^2n^{10\beta}e^{4K\beta\log n} =$$

$$4\beta^2(\log n)^2n^{10\beta}n^{4K\beta} = 4\beta^2(\log n)^2n^{(10+4K)\beta} < n^\gamma,$$

where γ depends only on β (and ε) and tends to 0 when β goes toward 0. Therefore, if β is small enough, the number of equivalence classes is less than

$$\begin{aligned} n^\gamma &< \frac{\varepsilon^3 n}{8\log n} \leq \frac{\varepsilon n r^2(1/4)}{8(1/\varepsilon)^2\log n} < \frac{\varepsilon n r^2}{10(3/\varepsilon)\log n(1/\varepsilon)+1} \\ &< \frac{\varepsilon n r}{10(3/\varepsilon)\beta(\log n)(1/\varepsilon)(1/r)+1} < \frac{\varepsilon n r}{10(3/\varepsilon)\beta(\log n)/(r\log(1/\varepsilon))+1} \\ &< \frac{\varepsilon n r}{10(P/M)M} = \frac{\varepsilon n r}{10P} < \frac{\varepsilon n r}{10p} < \frac{\varepsilon n r}{5(p+M)} \end{aligned}$$

and since according to (24) there are more than $\varepsilon rn/5$ vertices in V there has to be an equivalence class V_1 in V consisting of $\lfloor p+M \rfloor$ vertices.

Finally we will show that this means that there exists a $K' = K_r(q_1, \ldots, q_r)$ subgraph of G with a p' as p in (15) that contradicts the maximality of p for $K = K_r(p_1, \ldots, p_r)$. Denote by $\bar{C}_i$ the set of vertices of the ith class of K, C_i, joined to the vertices of V_1. Without loss of generality, assume that the $\bar{C}_i$ are ordered by increasing size, $|\bar{C}_1| \leq |\bar{C}_2| \leq \cdots \leq |\bar{C}_r|$, in particular, $|\bar{C}_1| < M$ (follows from definition of V). Let us consider two cases when constructing the contradicting K': $|\bar{C}_2| \leq p$ or $|\bar{C}_2| > p$.

If $|\bar{C}_2| \leq p$, let the classes of $K' = K_r(q_1, \ldots, q_r)$ be

$$C_1^* = V_1, \quad C_2^* = \bar{C}_1 \cup \bar{C}_2 \quad \text{and} \quad C_j^* = \bar{C}_j, j = 3, \ldots, r.$$

Since every vertex of V is joined to more than $(r-1)p$ vertices of K according to (25)

$$|\bigcup_1^r \bar{C}_j| > (r-1)p$$

we know that $|C_1^*| = |V_1| \geq M > |\bar{C}_1|$, $|C_2^*| = |\bar{C}_1 \cup \bar{C}_2| = |\bar{C}_1| + |\bar{C}_2|$ and the C_j^*, $j = 1, \ldots, r$ are pairwise disjoint. Thus,

$$rp' = |\bigcup_1^r C_j^*| > |\bigcup_1^r \bar{C}_j| - |\bar{C}_1| + \lfloor p+M \rfloor - |\bar{C}_2| + |\bar{C}_1| + |\bar{C}_2| =$$

$$(r-1)p + \lfloor p+M \rfloor > rp.$$

Furthermore, $|C_1^*| \leq p+M < p'+M$, $|C_2^*| = |\bar{C}_1| + |\bar{C}_2| < M+p < p'+M$ and obviously $|C_i^*| \leq p'+M$ for $i = 3, \ldots, r$. Thus K' satisfies (15) and contradicts the maximality of K.

If $|\bar{C}_2| > p$, select $q = \lfloor p+1 \rfloor \leq p + M$ vertices from W_1 and also from each $|\bar{C}_j|, j = 2, \ldots, r$ (which is possible since the classes are ordered in ascending order by size and $|C_2| > p$ according to the assumption). $K_r(q)$ is a complete subgraph in G with

$$\frac{1}{r}\sum_{i=1}^{r} q = q = \lfloor p+1 \rfloor > p,$$

which contradicts the maximality of K. This completes the proof of Theorem 3.7 and Theorem 3.5 as well.

□

In this section we mentioned the Kővári-Sós-Turán Theorem (3.1) mostly for completeness. In the following section we will use theorems 3.3, 3.4 and 3.5.

4 Some Generalizations to Hypergraphs

We try to find a generalization of theorems 3.3, 3.4 and 3.5 to k-hypergraphs with certain forbidden subgraphs

Definition 4.1 Suppose $H = (V, E)$ is a 3-hypergraph and $v \in V(H)$ a vertex. Let V_v be the set of all vertices adjacent to v. We then construct the set E_v (see figure 3)

$$E_v \ = \ \{\{v_1, v_2\} \ : \ v_1, v_2 \ \in \ V_v \ \textit{and} \ \{v, v_1, v_2\} \in E\}.$$

Denote the ordered pair (V_v, E_v) by H_v. Obviously H_v is 2-uniform i.e. an ordinary graph.

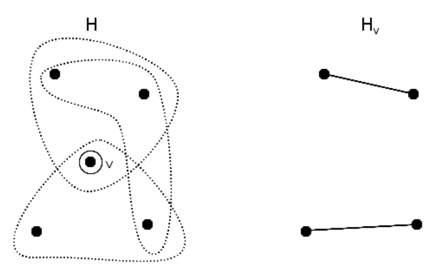

Figure 3: Example of H and H_v.

Definition 4.2 We generalize Definition 4.1. Suppose $H = (V, E)$ is a k-uniform hypergraph and $v_1, v_2, ..., v_{k-2} \in V(H)$ are vertices. Let $V_{v_1,v_2,...,v_{k-2}}$ be the set of all vertices adjacent to $v_1, v_2, ..., v_{k-2}$. We then construct the set

$$E_{v_1,v_2,...,v_{k-2}} = \{\{v_{k-1}, v_k\} : v_{k-1}, v_k \in V_{v_1,v_2,...,v_{k-2}} \text{ and } \{v_1, v_2, ..., v_k\} \in E\}.$$

Denote the ordered pair $(V_{v_1,v_2,...,v_{k-2}}, E_{v_1,v_2,...,v_{k-2}})$ by $H_{v_1,v_2,...,v_{k-2}}$. By this construction each selection of $k-2$ vertices from $V(H)$ gives rise to a 2-uniform graph, cf. the previous Definition 4.1.

We use the Definition 4.1 in the proof of the following theorem:

Theorem 4.3 *Suppose H is a 3-hypergraph of order n where, for all $v \in H$, H_v does not contain a $K(s,t)$, $2 \leq s, 2 \leq t$. Then:*

$$e(H) \leq \frac{1}{2}(s-1)^{1/t}(n-t)(n-1)^{1-1/t}n + \frac{1}{2}(t-1)(n-1)n$$

Proof. For every graph $H_v, v \in H$ it is obvious from the construction that $v \notin H_v$, thus the order of H_v is no greater than $|H|-1$. Since we suppose that no H_v contains a $K(s,t)$ we can use Theorem 3.3 to get:

$$e(H_v) \leq \frac{1}{2}(s-1)^{1/t}(n-t)(n-1)^{1-1/t} + \frac{1}{2}(t-1)(n-1)$$

We then, by summation over all vertices in H, get:

$$e(H) \leq \frac{1}{2}(s-1)^{1/t}(n-t)(n-1)^{1-1/t}n + \frac{1}{2}(t-1)(n-1)n$$

□

Generalising this theorem to k-uniform hypergraphs we get:

Theorem 4.4 *Suppose H is a k-hypergraph of order n where, for all $v_1, v_2, ..., v_{k-2} \in H$, $H_{v_1,v_2,...,v_{k-2}}$ does not contain a $K(s,t)$, $2 \leq s, 2 \leq t$. Then:*

$$e(H) \leq \frac{1}{2}\binom{n}{k-2}\left((s-1)^{1/t}(n-k-t+3)(n-k+2)^{1-1/t} + (t-1)(n-k+2)\right)$$

Proof. For every graph $H_{v_1,v_2,...,v_{k-2}}$ we note that $v_1, v_2, ..., v_{k-2} \notin H_v$, thus the order of H_v is no greater than $|H|-(k-2)$. According to the supposition, no $H_{v_1,v_2,...,v_{k-2}}$ contains a $K(s,t)$, therefore we can use Theorem 3.3 to get:

$$e(H_{v_1,v_2,...,v_{k-2}}) \leq \frac{1}{2}(s-1)^{1/t}(n-(k-2)-t+1)(n-(k-2))^{1-1/t} + \frac{1}{2}(t-1)(n-(k-2)) =$$

$$e(H_{v_1,v_2,...,v_{k-2}}) \leq \frac{1}{2}(s-1)^{1/t}(n-k-t+3)(n-k+2)^{1-1/t} + \frac{1}{2}(t-1)(n-k+2)$$

Since there are $\binom{n}{k-2}$ ways to choose $H_{v_1,v_2,...,v_{k-2}}$, we get:

$$e(H) \leq \frac{1}{2}\binom{n}{k-2}\left((s-1)^{1/t}(n-k-t+3)(n-k+2)^{1-1/t} + (t-1)(n-k+2)\right)$$

□

Generalizing Theorem 3.4 in a similar manner yields:

Theorem 4.5 *Suppose H is a k-hypergraph of order n where, for all $v_1, v_2, ..., v_{k-2} \in H$, $H_{v_1,v_2,...,v_{k-2}}$ does not contain a $K(t,t)$, $2 \leq t$. Then:*

$$e(H) \leq \frac{1}{2}\binom{n}{k-2}\left((t-1)^{1/t}(n-1)^{2-1/t} + \frac{t-1}{2}(n-1)\right)$$

Proof. The proof is the same as for Theorem 4.3 except we use Theorem 3.4 to get:

$$e(H_{v_1,v_2,\ldots,v_{k-2}}) \le \frac{1}{2}\left((t-1)^{1/t}(n-1)^{2-1/t} + \frac{t-1}{2}(n-1)\right)$$

and thus

$$e(H) \le \frac{1}{2}\binom{n}{k-2}\left((t-1)^{1/t}(n-1)^{2-1/t} + \frac{t-1}{2}(n-1)\right)$$

□

Using Theorem 3.5 we can prove the following:

Theorem 4.6 *There is an absolute constant $\alpha > 0$ such that if $0 < \varepsilon < \frac{1}{r}$ and*

$$m > \left(1 - \frac{1}{r} + \varepsilon\right)\frac{n^2\binom{n}{k-2}}{2} \tag{26}$$

then for every k-hypergraph $H(n,m)$, there are some $v_1, v_2, \ldots, v_{k-2} \in H$, such that $H_{v_1,v_2,\ldots,v_{k-2}}$ contains a $K_{r+1}(t)$, where

$$t = \left\lfloor \frac{\alpha \log n}{r \log 1/\varepsilon} \right\rfloor \tag{27}$$

Proof. Let $H = H(n,m)$ be a k-hypergraph. Suppose $\alpha > 0$, $0 < \varepsilon < \frac{1}{r}$ and

$$m > \left(1 - \frac{1}{r} + \varepsilon\right)\frac{n^2\binom{n}{k-2}}{2} \tag{*}$$

Suppose also that for every selection of $v_1, \ldots, v_{k-2} \in H$, $e(H_{v_1,\ldots,v_{k-2}}) \le \left(1 - \frac{1}{r} + \varepsilon\right)\frac{n^2}{2}$. This means that no selection of $k-2$ vertices from H belongs to more than $\left(1 - \frac{1}{r} + \varepsilon\right)\frac{n^2}{2}$ hyperedges and therefore (since there are $\binom{n}{k-2}$ such selections), $m \le \left(1 - \frac{1}{r} + \varepsilon\right)\frac{n^2\binom{n}{k-2}}{2}$, which contradicts (*). Hence, assuming (*) there is some selection of vertices $v_1, \ldots, v_{k-2}$ such that $e(H_{v_1,v_2,\ldots,v_{k-2}}) > \left(1 - \frac{1}{r} + \varepsilon\right)\frac{n^2}{2}$. According to Theorem 3.5 we then, for this selection, have that $H_{v_1,v_2,\ldots,v_{k-2}}$ contains a $K_{r+1}(t)$, where $t = \left\lfloor \frac{\alpha \log n}{r \log 1/\varepsilon} \right\rfloor$. □

Definition 4.7 *If $H = (V,E)$ is a hypergraph and $H' \subset H$ is an induced subgraph of H, then the vertex set $V(H')$ is an independent set of H if H' does not contain any edges.*

Finally, using Theorem 4.6 we can prove the following result about hypergraphs and independent sets.

Theorem 4.8 *There is an absolute constant $\alpha > 0$ such that if $0 < \varepsilon < \frac{1}{r}$ and*

$$m > \left(1 - \frac{1}{r} + \varepsilon\right)\frac{n^3}{2} \tag{28}$$

then for every 3-hypergraph $H(n,m)$ the following is true: if H has no independent set with $t+1$ vertices, where

$$t = \left\lfloor \frac{\alpha \log n}{r \log 1/\varepsilon} \right\rfloor, \tag{29}$$

then H is not 2-colorable.

Proof. Let $H = H(n, m)$ be a 3-hypergraph. Suppose $0 < \varepsilon < \frac{1}{r}$ and

$$m > \left(1 - \frac{1}{r} + \varepsilon\right) \frac{n^3}{2} \tag{*}$$

then Theorem 4.6 tells us that there is an absolute constant $\alpha > 0$ for which some selection of a vertex v results in a graph H_v containing a $K_{r+1}(t)$ (See figure 4 for an example where $r = 1$). I.e. we have $r + 1$ t-sets of vertices, in each of which, none of the vertices are adjacent (in H_v) Let us call these t-sets $I_1, I_2, \ldots, I_{r+1}$. Now there are two cases:

- Case 1: For every I_i, $1 \le i \le r + 1$ there is a choice of three vertices $v_1, v_2, v_3 \in I_i$ that make up an edge in H (i.e. $\{v_1, v_2, v_3\} \in E(H)$), in which case H is not 2-colorable. To see this, try to color the graph using colors 1 and 2. Let us assume (without loss of generality) that v has color number 1. Since the edge a (See fig 5) has to contain at least one vertex of color 1 and the same is true for edge c, and since every pair of vertices in different groups form an edge together with v, we can conclude that there is an edge in the graph where all the vertices must be the same color. I.e. the graph is not 2-colorable.

- Case 2: For some $1 \le i \le r + 1$, there is no such choice for I_i. Therefore at least one of these t-sets I_i together with v form an independent set containing $t + 1$ vertices.

Therefore, if there are no independent sets of order $t + 1$, then H is not 2-colorable. On the other hand, the reverse is not necessarily true, since there may be subgraphs that are not 2-colorable and that do not contain vertices from H_v.

□

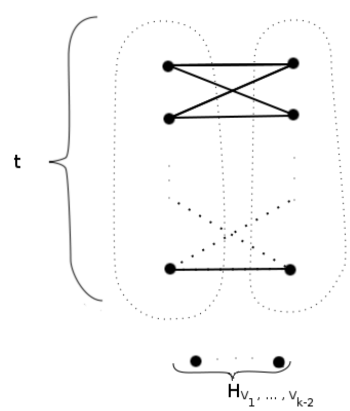

Figure 4: Example of $H_{v_1,v_2,\ldots,v_{k-2}}$ and the contained $K_2(t)$ when $r = 1$ and Theorem 4.6 is applicable.

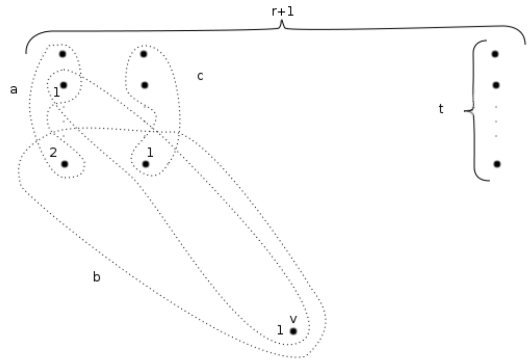

Figure 5: The $K_{r+1}(t)$ in H_v and some of the edges of H.

References

[1] R. Diestel, Graph Theory (2nd edition), Springer-Verlag 2000, (1997), 18-20.

[2] P. Turán, *Egy gráfelméleti szélsöértékfeladatról* (On an Extremal Problem in Graph Theory.) (in Hungarian), Mat. Fiz. Lapok 48, (1941), 436-452.

[3] P. Erdős and A.H. Stone, On the Structure of Linear Graphs, Bull. Am. Math. Soc. 52, (1946), 1087-1091.

[4] M. Aigner and G. M. Ziegler, Proofs from THE BOOK (3rd edition), Springer-Verlag, (2004), 207-211.

[5] M. Aigner, Turán's Graph Theorem, The American Mathematical Monthly, Vol. 102, No. 9 (Nov., 1995), 808-816.

[6] B. Bollobás, Extremal Graph Theory, Academic Press, (1978), 309-311 and 327-336.

[7] P. Kövári, V. T. Sós and P. Turán, On a problem of K. Zarankiewicz, Colloq. Math. 3 (1954) 50-57.

[8] T. Needham, A Visual Explanation of Jensen's Inequality, The American Mathematical Monthly, Vol. 100, No. 8 (Oct., 1993), 768-771.

[9] Hyltén, Cavallius, On a Combinatorical Problem, Colloq. Math. 6 (1958) 59-65.

[10] Š. Znam, On a Combinatorial Problem of K. Zarankiewicz, Colloq. Math. 11 (1963) 81-84.

[11] B. Bollobás, P. Erdős and M. Simonovits, On the Structure of Edge Graphs II, J. London Math. Soc. 12 (2), (1976) 219-224.