Universitat Politècnica de Catalunya
Facultat de Matemàtiques i Estadística

Master in Advanced Mathematics and Mathematical Engineering

Master's thesis

Finding Partite Graphs Efficiently

Ferran Espuña Bertomeu

Supervised by Richard Lang

June 2025

Thanks to...

Abstract

Keywords

hypergraph, algorithm, graph, partite, extremal

1. Introduction

TODO: Write introduction

2. Preliminaries

In this section we introduce some basic definitions and results that will be used throughout the thesis.

Definition 2.1. For an integer $k \geq 2$ a finite *k-graph* is a tuple $G = (V, E)$ where V is a finite set and $E \subseteq \binom{V}{k}$. We call the elements of $V =: V(G)$ its *vertices* and those of $E =: E(G)$ its *edges*.

Remark 2.2. If we let $k = 2$ we recover the usual definition of a graph.

Definition 2.3. Let $G = (V, E)$ and $H = (W, F)$ be k-graphs. A *homomorphism* from G to H is a map $f : V \to W$ such that for every edge $e \in E$ the set $f(e) := \{f(v) \mid v \in e\}$ is an edge in H (that is, $f(e) \in F$). If such a homomorphism exists and is injective, we say that f is an embedding of G on H and that H contains G as a subgraph. If, furthermore, $f^{-1} : \mathrm{Im}(f) \to V$ is a homomorphism, we say that f is an *induced* embedding and that H contains G as an *induced* subgraph. We write $G \subseteq H$. If, in addition, f is a bijection, we say that f is an *isomorphism* and that G is *isomorphic* to H. We write $G \cong H$.

Remark 2.4. It is elementary to check that (induced) inclusion is an order relation and that isomorphism is an equivalence relation. Furthermore, isomorphism preserves (induced) inclusion. Therefore, we can talk about the (induced) subgraph condition up to isomorphism, both in the *host* k-graph (H) and in the *guest* k-graph (G).

Remark 2.5. Given a k-graph $G = (V, E)$ and a set W satisfying $|V| = |W|$, we can define an edge set E' on W such that $G \cong (W, E')$ by taking any bijection $f : V \to W$ and setting $E' = \{f(e) \mid e \in E\}$. This frees us, up to isomorphism, to change or reorder the vertices of a k-graph.

Proposition 2.6. *Let $G = (V, E)$ be a k-graph with nonempty edge set and $n \geq |V|$ be an integer. Then there exists an integer $M_0 = ex(n, G) \in \left[0, \binom{n}{k}\right)$ such that the condition*

"All k-graphs with n vertices and m edges contain G as a subgraph"

is true for all $\binom{n}{k} \geq m > M_0$ and false for all $0 \leq m \leq M_0$.

Proof. Note that, if M_0 exists, clearly it is unique. Also, the condition is clearly false for $m = 0$ and true for $m = \binom{n}{k}$ (the only graph H with vertex set W, $|W| = n$ and $\binom{|V|}{k}$ vertices is the one having all k-sets of vertices so any injective map $f : V \to W$ is an embedding of G in H). We only need to show that if the condition is true for m then it is true for all $m' \geq m$. Suppose it is true for m and let $m' \geq m$. Let $H = (W, F)$ be a k-graph with n vertices and m' edges. We can just take $F' \subseteq F$ with $|F'| = m$. By hypothesis, the graph $H' = (W, F')$ contains G as a subgraph, and the identity map in W is an embedding of H' in H:

$$G \subseteq H' \subseteq H \implies G \subseteq H \qquad \square$$

Remark 2.7. We call $ex(n, G)$ the *extremal number* of G. It is clearly invariant under isomorphism.

Definition 2.8. for an integer $p \geq k$, a k-graph $G = (V, E)$ is *p-partite* if there exists a partition $V = V_1 \cup \cdots \cup V_p$ such that every edge $e \in E$ intersects every part V_i in at most one vertex. We may write $G = (V_1, \ldots, V_p; E)$ and say that G is a partite k-graph on $V_1, \ldots, V_p$.

Remark 2.9. If $p = k$, every edge intersects every part in exactly one vertex, so we can identify the edges with a subset of $V_1 \times \cdots \times V_k$.

Definition 2.10. A k-partite k-graph $G = (V_1, \dots, V_k; E)$ is *complete* if every k-set of vertices $(v_1, \dots, v_k)$ with $v_i \in V_i$ satisfies $\{v_1, \dots, v_k\} \in E$. We write $G = K(V_1, \dots, V_k)$.

Remark 2.11. $V_1, \dots, V_k, W_1, \dots, W_k$ are disjoint sets, and $|V_i| = |W_i| =: a_i$ for all i then it is elementary to check that

$$K(V_1, \dots, V_k) \cong K(W_1, \dots, W_k)$$

by a construction very similar to the one in Remark 2.5. This allows us to talk about *the* complete k-partite k-graph on $a_1, \dots, a_k$ vertices, which we denote by $K(a_1, \dots, a_k)$.

Remark 2.12. All k-partite k-graphs with part sizes $b_1 \leq a_1, \dots, b_k \leq a_k$ are contained in $K(a_1, \dots, a_k)$ as subgraphs. This lets us follow the exact same argument as in Proposition 2.6 to define the following:

Definition 2.13. let $0 < t_1 \leq v_1, \dots, 0 < t_k \leq v_k$ be integers. Then the *generalized Zarankiewicz number* $z(v_1, \dots, v_k; t_1, \dots, t_k)$ is the largest integer $0 \leq z < v_1 \dots v_k$ for which there exists k-partite k-graph H with part sizes $|V_1| = v_1, \dots, |V_k| = v_k$ and z edges such that no embedding f of $K(T_1, \dots, T_k)$ with $|T_i| = t_i$ in it exists satisfying $f(T_i) \subseteq V_i$ for all i.

Remark 2.14. Finding this number can help us upper bound the extremal number of $K(t_1, \dots, t_k)$ asymptotically: Assume that G is a $K(t_1, \dots, t_k)$-free n-vertex k-graph with m edges. pick $v_1, \dots, v_k$ such that $\sum_i v_i = n$ and $v_i \sim n/k$ (For example, $\lfloor n/k \rfloor \leq v_i \leq \lceil n/k \rceil$) Let $V_1, \dots, V_k$ be a random partition of $V(G)$ with $|V_i| = n_i$. for an edge $e \in E(G)$, the probability that e is an edge in $K(V_1, \dots, V_k)$ is greater than

$$k! \prod_i n_i \sim k!(1/k)^k$$

which is independent of n. Therefore, the expected number of edges satisfying this condition is a positive fraction of m. Applying the probabilistic method, we can conclude that

$$ex(n, K(t_1, \dots, t_k)) = O(z(\lceil n/k \rceil, \overset{k}{\cdots}, \lceil n/k \rceil; t_1, \dots, t_k))$$

The problem on finding the Zarankiewicz number was first posed by K. Zarankiewicz in 1951 for the case of bipartite 2-graphs (that is, finding $z(u, w; s, t)$), in terms of finding all-1 minors in a $0-1$ matrix. An upper bound for it in the case $m = n, s = t$ was found by Kővari, Sós and Turán in [4] in 1954. This was generalized to arbitrary complete partite 2-graphs by C. Hyltén-Cavallius in [3] in 1958. The result is stated and proved here for completeness:

Theorem 2.15. *Let $0 < m \leq s$ and $0 < n \leq t$ be integers. Then*

$$z(u, w; s, t) \leq (s-1)^{1/t}(w-t+1)u^{1-1/t} + (t-1)u$$

Proof. Suppose that we have a bipartite graph $G = (U, W; E)$ with $|U| = u$, $|W| = w$ and $|E| = z$ exceeding the bound stated above. Let us consider the set

$$P = \left\{ (x, Y) \in U \times \binom{W}{t} \middle| \forall y \in Y : \{x, y\} \in E \right\}$$

Counting on the first coordinate, and using Jensen's inequality, we get

$$|P| = \sum_{x \in U} \binom{d_G(x)}{t} = \sum_{x \in U} f(d_G(x)) \geq u \sum_{x \in U} f(z/u) = m\binom{z/u}{t}$$

Where we define

$$f(x) := \begin{cases} \binom{x}{t}, & \text{if } x \geq t-1 \\ 0, & \text{otherwise} \end{cases}$$

Which is convex, meaning we get the inequality as Jensen's inequality. The other equalities come from the fact that $f(d)$ agrees with $\binom{d}{t}$ for all integers $d \geq 0$; and that by our bound on z, $z \geq (t-1)u \implies z/u \geq t-1$.

If we had s different elements of P with the same second coordinate T, they would all necessarily have different first coordinates (say $S = \{x_1, \dots, x_s\}$). But now, by definition of P, for all $a \in S, b \in T$, we have $\{a, b\} \in E$. This would mean that the inclusion map from $S \cup T$ to $U \cup W$ is an embedding of $K(s, t)$ in G, as described in Definition 2.13. Supposing that this is not the case, by the pigeonhole principle, we have:

$$|P| \leq (s-1)\binom{w}{t}$$

Putting the two inequalities together, we get:

$$u\binom{z/u}{t} \leq (s-1)\binom{w}{t}$$

Now, because we can see E as a subset of $U \times W$, we get $z \leq uw \implies z/u \leq w$. In particular, we have:

$$\frac{(z/u-(t-1))^t}{\binom{z/u}{t}} \leq \frac{(w-(t-1))^t}{\binom{w}{t}}$$

which is true for each factor when expanding the denominators. Multiplying the two inequalities, we get:

$$m\,(z/u-(t-1))^t \leq (s-1)(w-(t-1))^t$$

which, by algebraic manipulation, gives

$$z \leq (s-1)^{1/t}(w-t+1)u^{1-1/t} + (t-1)u$$

In contradiction with our assumption. □

Remark 2.16. Following Remark 2.14, we can use this bound to get an upper bound on the extremal number of $K(s, t)$:

$$ex(n, K(s,t)) = O\left((s-1)^{1/t}(n-t+1)n^{1-1/t} + (t-1)n\right) = O\left(n^{2-1/t}\right)$$

Note that if $s < t$, we get the better bound $O\left(n^{2-1/s}\right)$ by interchanging the roles of s and t.

In 1964, Erdős [2] generalized this result to arbitrary complete partite k-graphs in the following theorem:

Theorem 2.17. $ex(n, K(t, \overset{k}{\cdots}, t)) = O\left(n^{k-\frac{1}{t^{k-1}}}\right)$

A more modern proof of this result can be found in [1], which also generalizes it to arbitrary complete k-partite k-graphs (not necessarily with equal part sizes). They in fact prove a bound for the generalized Zarankiewicz number in a similar way we proved the bound for the Zarankiewicz number in Theorem 2.15, which then following Remark 2.14 gives the result in Theorem 2.17.

3. Our Algorithm

In this section we present a polynomial-time algorithm to find a balanced partite k-graph in a given k-graph G with n vertices and m edges with part size in the same order of magnitude as stated in Theorem 2.17.

Remark 3.1. If we let q be the size of each part in the partite k-graph we are looking for, we need

$$m \geq ex(n, K(t, \overset{k}{\cdots}, t)) = O\left(n^{k-\frac{1}{t^{k-1}}}\right)$$

Defining $d = \frac{m}{n^k}$, and taking logarithms, this is true iff

$$\log d \geq -\frac{\log n}{t^{k-1}} + O(1)$$

which implies

$$t = O\left(\left(\frac{\log n}{\log(1/d)}\right)^{\frac{1}{k-1}}\right)$$

This algorithmic problem has already been solved for $k = 2$ by Mubayi and Turán [5]. The algorithm in that case follows very closely the structure of the proof of Theorem 2.15. We outline the algorithm for $k = 2$ here for context and clarity. The variable names have been altered to match the notation used in this thesis.

Algorithm 1 Finding a balanced bipartite graph in a 2-graph

Require: A graph $G = (V, E)$ with $|V| = n$, $E = m$

```
d ← m/n^2
assert d ≥ 3n^{1/2}
t ← ⌊log(n/2) / log(2e/d)⌋ , w ← ⌊t/d⌋
W ← The set of t vertices with highest degree in G
U ← V \ W
for all T ∈ (W choose t) do
    S ← {x ∈ U : x, y ∈ E for all y ∈ T}
    if |S| ≥ t then
        return (S, T)
    end if
end for
```

If the set S is too large we can simply take a subset of it of size t. The algorithm is correct if at some point it returns a pair of sets (S, T). The argument of why this is the case boils down to showing that there is a large number of edges between U and W and then applying Theorem 2.15 with $u = |U| = n - w$ and $s = t$. Finally, the algorithm runs in polynomial time because the number of iterations of the loop is

$$\binom{w}{t} \leq \left(\frac{ew}{t}\right)^t \leq \left(\frac{1}{d}\right)^t e^t < e^{t\log(1/d)+\log n} < e^{2\log n} = n^2$$

Remark 3.2. The requirement for a minimum density is because if $d = o\left(n^{1/2}\right)$ then there may not even be a $K(2, 2)$ in G.

We will now describe an extended algorithm which will find a balanced partite k-graph in a k-graph G with n vertices and $m = dn^k$ with part size

$$t(n,d,k) = \left\lfloor \left(\frac{\log(n/2^k)}{\log(2^{k+1}/d)} \right)^{\frac{1}{k-1}} \right\rfloor$$

Algorithm 2 Finding a balanced partite k-graph in a k-graph

1: **function** $\mathrm{FIND\_PARTITE}(G, k)$
2:     **assert** G is a k-graph
3:     **if** $k = 1$ **then**
4:         **return** $\{x : \{x\} \in E(G)\}$
5:     **end if**
6:     **assert** $d \geq 4^k n^{-\frac{1}{2^{k-1}}}$
7:     $n \leftarrow |V(G)|$, $m \leftarrow |E(G)|$, $d \leftarrow m/n^k$
8:     $t \leftarrow t(n,d,k)$, $w \leftarrow \lceil 2t/d \rceil$, $s \leftarrow \lfloor d^t n^{k-1} \rfloor$
9:     $W \leftarrow$ the set of t vertices with highest degree in G
10:     $U \leftarrow \binom{V \setminus W}{k-1}$
11:     **for all** $T \in \binom{W}{t}$ **do**
12:         $S \leftarrow \{\, y \in U : \{x\} \cup y \in E$ for all $x \in T \,\}$
13:         **if** $|S| \geq s$ **then**
14:             **return** $(S) \circ \mathrm{FIND\_PARTITE}\,(G' := (V \setminus W,\ S),\ k-1)$
15:         **end if**
16:     **end for**
17: **end function**

Where the operator $\circ$ denotes the concatenation of tuples, and we understand a 1-graph to be a subset of a set. We will now argue that this algorithm is correct (as long as the the density condition is met on the first call) and runs in polynomial time. The following lemmas are stated assuming that $k \geq 2$ and the minimum density requirement of line 6 of Algorithm 2 is met.

Remark 3.3. The $t = 1$ case is trivial, as we can simply select one edge of G. The minimum density requirement implies $t \geq 2$, although the converse is not true for any n. However, it *is* true that, for all k, $t \geq 3$ implies the minimum density requirement for $n \geq n_0(k)$. Therefore, we can simply solve the problem for $t = 2$ or $n < n_0(k)$ by checking all possible embeddings (which clearly can be done in polynomial time for every k) and apply 2 otherwise.

Lemma 3.4. *The selection of t, w, s in line 8 is sound in the sense that $t \leq w \leq n$, $k - 1 \leq n - w$ and $s \leq \binom{n-w}{k-1}$.*

Proof. $t \leq w$ is clear. We will in fact show that $w < \frac{n}{2}$. If not,

$$\frac{n}{2} \leq w \leq 1 + \frac{2t}{d} < 1 + \frac{2\log(n/2^k)}{4^k n^{-\frac{1}{2^{k-1}}}} < + \frac{\log(n)}{2^{2k-1} n^{-\frac{1}{2}}} = 1 + \frac{\log n \cdot \sqrt{n}}{8} < 1 + \frac{n}{8} \implies n < 3$$

But because we have positive density this implies $n = k = 2$, with at most one edge, which clearly does not satisfy the density condition.

We can also show that $k < \frac{n}{2}$. If not,

$$1 \geq d \geq 4^{\frac{n}{2}} n^{-\frac{1}{2^{n/2-1}}} \geq e^{\frac{n}{2}\log 4 - \frac{\log n}{2^{n/2-1}}} \implies \frac{n}{2}\log 4 \leq \frac{\log n}{2^{n/2-1}}$$

which is false for all $n > 0$.
Therefore,

$$k + w < n \implies k - 1 < n - w$$

Finally, suppose $s > \binom{n-w}{k-1}$. Then, using the fact that $w < \frac{n}{2}$,

$$\left(\frac{n}{2k}\right)^{k-1} \leq \left(\frac{n-w}{k-1}\right)^{k-1} \leq \binom{n-w}{k-1} < s \leq d^t n^{k-1} \implies \left(\frac{1}{2k}\right)^{k-1} < d^t \leq \left(\frac{1}{k!}\right)^2$$

Where in the last inequality we use that $t \geq 2$ and there are at most $\binom{n}{k} \leq \frac{n^k}{k!}$ edges in G.
We can show that $k!^2 \geq (2k)^{k-1}$ for all k , which means we have reached a contradiction.

□

Lemma 3.5. *With $W \subset V$ as defined in line 9, There are at least $\frac{3}{2}dwn^{k-1}$ edges of G with exactly one vertex in W.*

Proof. The degree sum over V is kdn^k. Thus, by the pigeonhole principle, the degree sum over W is at least $\frac{w}{n}kdn^k = wkdn^{k-1}$. For $2 \leq j \leq n$, consider the contribution to this sum by edges with exactly j vertices in W. Each such edge contributes j to the sum, and there are at most $\binom{w}{j}\binom{n-w}{k-j} \leq \frac{w^j n^{k-j}}{j!} \leq \frac{w^j n^{k-j}}{j}$ of them. Thus, the total contribution of these edges is at most $w^j n^{k-j} \leq w^2 n^{k-2}$. The number of edges with only one vertex in W is then at least

$$wkdn^{k-1} - (k-1)w^2 n^{k-2} = dwn^{k-1}\left(k - \frac{(k-1)w}{nd}\right)$$

Suppose, by way of contradiction, that $k - \frac{(k-1)w}{nd} < \frac{3}{2}$. Using that $\frac{k-1}{k-3/2} \leq 2$ for $k \geq 2$, we arrive at

$$2 \geq \frac{nd}{w} \geq \frac{nd}{2t/d+1} \cdots$$

TODO Applying our minimum density, this means
TODO
which is false for all n. □

Lemma 3.6. *Line 14 of Algorithm 2 is reached at some point in the for loop in line 11.*

Proof. TODO □

Now, for the base case, we need:

Lemma 3.7. *For $k = 2$, Algorithm 2 finds $s \geq t$.*

Proof. TODO □

For the recursive step, we need:

Lemma 3.8. *For $k \geq 3$, in the recursive call in line 14 of Algorithm 2, the density condition of line 6 is met for $k-1$. That is,*

$$d' := \frac{s}{(n-w)^{k-1}} \geq 4^{k-1}(n-w)^{-\frac{1}{2^{k-2}}}$$

Furthermore, the resulting part size t' satisfies

$$t' := t(n-w, d', k-1) \geq t$$

Proof. TODO □

All in all, we can now state the following theorem:

Theorem 3.9. *Algorithm 2 finds a balanced partite k-graph in a k-graph G with n vertices and $m = dn^k$ with part size $t(n, d, k)$ in polynomial time.*

Proof. TODO □

4. Bibliography

References

[1] Matías Azócar Carvajal, Giovanne Santos, and Mathias Schacht. Canonical ramsey numbers for partite hypergraphs. *arXiv preprint arXiv:2411.16218*, 2024.

[2] P. Erdös. On extremal problems of graphs and generalized graphs. *Israel Journal of Mathematics*, 2(3):183–190, September 1964.

[3] C. Hyltén-Cavallius. On a combinatorical problem. *Colloquium Mathematicae*, 6(1):61–65, 1958.

[4] T. Kővari, V. Sós, and Turán P. On a problem of k. zarankiewicz. *Colloquium Mathematicae*, 3(1):50–57, 1954.

[5] Dhruv Mubayi and György Turán. Finding bipartite subgraphs efficiently. *Information Processing Letters*, 110(5):174–177, 2010.