Bounds on the Number of Edges in Hypertrees

Gyula Y. Katona[a,b,*], Péter G. N. Szabó[a]

[a] *Budapest University of Technology and Economics*
Department of Computer Science and Information Theory
Magyar Tudósok krt. 2, Budapest, Hungary, H-1117
[b] *HAS Research Group "Numerical Analysis and Large Networks"*
Pázmány Péter sétány 1/C, Budapest, Hungary, H-1117

Abstract

Let $\mathcal{H}$ be a k-uniform hypergraph. A chain in $\mathcal{H}$ is a sequence of its vertices such that every k consecutive vertices form an edge. In 1999 Katona and Kierstead suggested to use chains in hypergraphs as the generalisation of paths. Although a number of results have been published on Hamiltonian chains in recent years, the generalisation of trees with chains has still remained an open area. We generalise the concept of trees for uniform hypergraphs. We say that a k-uniform hypergraph $\mathcal{H}$ is a hypertree if every two vertices of $\mathcal{H}$ are connected by a chain, and an appropriate kind of cycle-free property holds. An edge-minimal hypertree is a hypertree whose edge set is minimal with respect to inclusion. After considering these definitions, we show that a k-uniform hypertree on n vertices has at least $n-(k-1)$ edges if $n > n_0(k)$, and it has at most $\binom{n}{k-1}$ edges. The latter bound is asymptotically sharp in 3-uniform case.

Keywords: hypertree, cycle in hypergraph, path in hypergraph
2000 MSC: 05C65, 05C05

1. Introduction

Trees are among the most fundamental, the most useful and the best understood objects in all of graph theory. In the present paper, we generalize the notion of tree to uniform hypergraphs and prove bounds on the number of edges of such hypertrees.

The concepts of paths and cycles were generalised to hypergraphs in many different ways, although none of those definitions became dominant over the years. One of the earliest generalisations of the Hamiltonian cycle to hypergraphs was given by Bermond et. al. in 1976. They gave extensions of known results about Hamiltonian cycles to the hypergraph setting [3].

*Corresponding author
Email addresses: kiskat@cs.bme.hu (Gyula Y. Katona), szape@cs.bme.hu (Péter G. N. Szabó)

Other interesting generalisations such as Berge-paths and Berge-cycles were introduced by Claude Berge in 1985 [1]. These are basically graph paths and cycles, but the edges are blown up with extra vertices in addition. In comparison, the path concept we introduce in Definition 1 satisfies stricter structural properties.

Recently, deep results regarding Hamiltonicity in hypergraphs were proven by Szemerdi et al. Their concept of tight path closely aligns with the concept of chain used in the current paper which, in fact, can be viewed as a continuation of these generalization efforts. For more details see [2, 3, 5, 6, 7, 8].

In the following, we generalise the concept of tree for k-uniform hypergraphs, where chains play the role of paths. After introducing the basic definitions, we discuss lower and upper bounds on the number of edges of such hypertrees.

First, we want to clarify the notions of chain, cycle and semicycle. In the following sections we assume that there are no multiple edges in a hypergraph.

Simple chains plays the most important role in the definition of hypertrees because we intend to require a natural chain-connectivity property.

Definition 1 (Chain). *The k-uniform hypergraph $\mathcal{P} = (V, \mathcal{E})$ is a* chain *if there exists a sequence $v_1, v_2, \ldots, v_l$ of its vertices such that every vertex appears at least once (possibly more times), $v_1 \neq v_l$ and for all $1 \leq i \leq l-k+1$, $\{v_i, v_{i+1}, \ldots, v_{i+k-1}\}$ are distinct edges of $\mathcal{P}$. The length of the chain $\mathcal{P}$ is $l-k+1$, the number of its edges.*

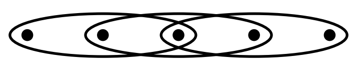

Figure 1: A 3-uniform chain of length 3

Definition 2 (Cycle). *The k-uniform hypergraph $\mathcal{C} = (V, \mathcal{E})$ is a* cycle *if there exists a sequence $v_1, v_2, \ldots, v_l$ of its vertices such that every vertex appears at least once (possibly more times) and for all $1 \leq i \leq l$, $\{v_i, v_{i+1}, \ldots, v_{i+k-1}\}$ are distinct edges of $\mathcal{C}$ (the indices should be understood modularly). The length of the cycle $\mathcal{C}$ is l, the number of its edges.*

Cycles are too special of structures for our purposes, so we use the weaker concept of semicycles. Hypergraphs without semicycles are in close resemblance with (ordinary) forests.

Definition 3 (Semicycle). *The k-uniform hypergraph $\mathcal{C} = (V, \mathcal{E})$ is a* semicycle *if there exists a sequence $v_1, v_2, \ldots, v_l$ of its vertices such that every vertex appears at least once (possibly more times), $v_1 = v_l$ and for all $1 \leq i \leq l-k+1$, $\{v_i, v_{i+1}, \ldots, v_{i+k-1}\}$ are distinct edges of $\mathcal{C}$. The length of the semicycle $\mathcal{C}$ is $l-k+1$, the number of its edges.*

A chain (cycle, semicycle) is called *self-intersecting* if there is a vertex appearing at least twice in its defining sequence. Similarly, a semicycle is *self-intersecting* if there is a vertex appearing at least twice in its defining sequence aside from the condition $v_1 = v_l$.

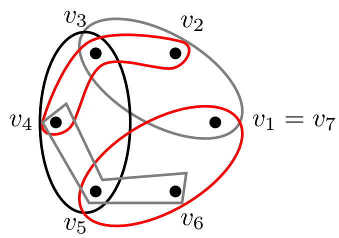

Figure 2: A 3-uniform semicycle of length 5

A non-self-intersecting chain (cycle) is also called a *tight-path* (*tight-cycle*) in related literature.

The length of a non-self-intersecting chain on n vertices is $n-k+1$, matching the fact that the length of a (2-uniform) path on n vertices is $n-1$.

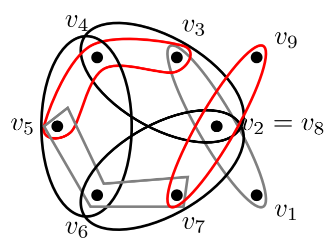

Figure 3: A 3-uniform self-intersecting chain of length 7

A k-uniform hypergraph is

- *chain-connected* if every pair of its vertices is connected by a chain, i.e. there exists a subhypergraph which is a chain and contains both vertices;
- *semicycle-free* if it contains no semicycle as a subhypergraph;

It is well-known from elementary graph theory that there are a handful of equivalent descriptions which define a tree. The analogous situation for hypergraphs is unclear, so we adopt the notion of chain-connectivity and semicycle-freeness in order to define a hypertree.

Definition 4 (Hypertree). *A k-uniform hypergraph is a* hypertree *if it is chain-connected and semicycle-free.*

We note that the 2-uniform hypertrees are just the common trees. We discuss interesting variations of Definition 4 in Section 4.

Our main results are the lower and upper bounds on the number of edges of hypertrees.

Theorem 1. *Let $\mathcal{H}$ be a chain-connected k-uniform hypergraph on $n \geq n_0(k)$ vertices. Then, $|\mathcal{E}(\mathcal{H})| \geq n-k+1$, where this bound is best possible.*

Theorem 2. *Let $\mathcal{H}$ be a semicycle-free k-uniform hypergraph on n vertices. Then, $|\mathcal{E}(\mathcal{H})| \leq \binom{n}{k-1}$. When $k = 3$, this bound is asymptotically best possible.*

The related proofs are presented separately in Section 2.

If we compare Theorem 1 and Theorem 2 to their graph counterparts, we can find similarities as well as surprising differences. Theorem 1 generalizes the well-known fact that a connected graph G on n vertices has $|E(G)| \geq n-1$ edges. When $k \geq 3$, the proof of Theorem 1 becomes quite non-trivial. For a glimpse into this difficulty, note that the property of chain-connectivity between a pair of vertices is not necessarily transitive and therefore, unlike with components in the graph case, chain-connectivity does not define an equivalence relation on the vertices. An other unexpected difficulty arises when we try to apply the lower bound of Theorem 1 for small hypertrees. The bound is true only if $n \geq n_0(k)$. We will prove bounds on $n_0(k)$ at the end of Section 2. We note that obviously $n_0(2) = 0$.

Theorem 2 is also true for graphs, as it implies that a cycle-free graph has at most n edges. Though this upper bound is not the best possible, it is still very close to the optimal bound, $n - 1$. In the k-uniform case, a trivial upper bound for the edge number would be $\binom{n}{k}$. For graphs, it is $\binom{n}{2}$ and the actual number of edges is a linear factor less. Theorem 2 shows the same trend for hypertrees.

The main difference for hypertrees is that the number of edges ranges from linear to degree $(k - 1)$, in other words, there is a $(k - 2)$-degree gap in between the lower and upper bounds, which is a really interesting fact and worth of further study. We give some thoughts on this topic in section 4.

Finally, we note that $\mathcal{H}$ does not need to be a hypertree neither in Theorem 1 nor in Theorem 2. For the lower bound only chain-connectivity, while for the upper bound only semicycle-freeness is assumed. We will also see that the latter condition can be further weakened.

We investigate hypertrees with bounded diameter as well. Such kind of restrictions get more and more attention in recent years and can be also found in related papers.

Definition 5 (l-hypertree). *A hypertree is called an l-*hypertree *if every chain in it is of length at most l.*

It turns out that the upper bound of Theorem 2 can be improved for l-hypertrees if l is less than k.

Theorem 3. *Let $1 \leq l \leq k$ and $\mathcal{H}$ be a k-uniform l-hypertree on n vertices. Then, $|\mathcal{E}| \leq \frac{1}{k-l+1}\binom{n}{k-1}$.*

The reader may find the proof and discussion of Theorem 3 in Section 3.

2. Proofs of the main results

In this section, we prove and discuss our main results Theorem 1 and Theorem 2. In Subsection 2.1 we present our results on the counterexamples related to Theorem 1.

Proof. [Theorem 1]

First of all, if the bound of Theorem 1 is true, then it is best possible because non-self-intersecting chains have exactly $n-k+1$ edges. This proves the second statement of Theorem 1. We note that there exist other examples of hypertrees with minimal number of edges, such as *tight stars*. A k-uniform hypergraph is a tight star if its edges intersect each other in a set of size $k-1$. Tight stars are also the most basic examples for 2-hypertrees.

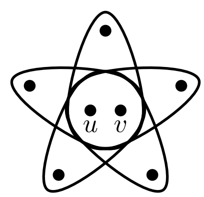

Figure 4: A 3-uniform tight star on 7 vertices

In order to prove the lower bound, we define the tight line graph of a uniform hypergraph, which partially preserves the structure of chains.

Let $\mathcal{H}' = (V', \mathcal{E}')$ be a k-uniform hypergraph. The *tight line graph* of $\mathcal{H}'$ is the graph $L_{\mathcal{H}'} = (\mathcal{E}', E')$, where $E' = \{\{e, f\} \colon e, f \in \mathcal{E}', |e \cap f| = k-1\}$. This definition is not equivalent to the definition of the usual line graph, where all pairs of intersecting edges are adjacent. Here, we are only interested in the intersections of size $(k-1)$ since the edges of chains are joined in this way.

The hypergraph $\mathcal{H}'$ is *line-graph-connected* if its tight line graph is connected. It will be easier to prove Theorem 1 for both line-graph-connected hypergraphs, while the general case can be traced back to the tight line graph components of $\mathcal{H}$.

Let $L = (\mathcal{E}, E)$ be the tight line graph of $\mathcal{H}$, and let $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_l$ denote the vertex sets of the connected components of L. Furthermore, let $V_i = \bigcup\{e \colon e \in \mathcal{E}_i\}$, $i = 1, \ldots, l$ be the projections of the components to V, called the classes of $\mathcal{H}$. Then, it can be easily shown that the following hold:

- $\bigcup_{i=1}^{l} V_i = V$ since every vertex of $\mathcal{H}$ is contained in some edges;
- for every $u, v \in V$, there exists an index i such that $u, v \in V_i$.

The second point follows from the fact that $\mathcal{H}$ is chain-connected. The edges of the chain connecting u and v forms a path in L, hence they are in the same component. The projection of this component contains both u and v.

In the next step, we prove a lemma which states that 1 is true for line-graph-connected hypergraphs, hence it is true for every class of $\mathcal{H}$.

Lemma 1. *Let $\mathcal{H}' = (V', \mathcal{E}')$ be a line-graph-connected k-uniform hypergraph that contains no isolated vertex. Then, $|\mathcal{E}'| \geq |V'| - (k-1)$.*

Proof. Let $m = |\mathcal{E}'|$, $e \in \mathcal{E}'$ be an arbitrary edge and $\mathcal{X}_1 = (e, \{e\})$ be a subhypergraph of $\mathcal{H}'$.

We define a subhypergraph-sequence $\mathcal{X}_1 < \mathcal{X}_2 < \ldots < \mathcal{X}_m$ for which $\mathcal{X}_m = \mathcal{H}'$ and for every i, $|V(\mathcal{X}_i)| - (k-1) \leq |\mathcal{E}(\mathcal{X}_i)| = i$. For $i = m$, the lemma follows.

We create this sequence by adding a yet unseen edge from $\mathcal{E}'$ in every step, which intersects the current vertex-set in at least $(k-1)$ vertices. We continue this process until all of the edges of $\mathcal{H}'$ were added. In the last step $\mathcal{X}_m = \mathcal{H}'$ will hold.

Assume that we have already defined $\mathcal{X}_{i-1}$ for some $i-1 < m$, and let L' denote the tight line graph of $\mathcal{H}'$. Since we are not in the mth step and L' is connected, there must be an edge in L' between $\mathcal{E}(\mathcal{X}_{i-1})$ and $\mathcal{E}' \backslash \mathcal{E}(\mathcal{X}_{i-1})$. Let $f_{i-1} \in \mathcal{E}(\mathcal{X}_{i-1})$ and $e_i \in \mathcal{E}' \backslash \mathcal{E}(\mathcal{X}_{i-1})$ denote the two endpoints of this edge. Then, $|f_{i-1} \cap e_i| = k-1$ by the definition of L', thus $|e_i \cap V(\mathcal{X}_{i-1})| \geq k-1$. Let $\mathcal{X}_i = (V(\mathcal{X}_{i-1}) \cup e_i, \mathcal{E}(\mathcal{X}_{i-1}) \cup \{e_i\})$.

By induction, it can be shown that for every index i,

$$|\mathcal{E}(\mathcal{X}_i)| \geq |V(\mathcal{X}_i)| - (k-1). \tag{1}$$

For $i = 1$ (1) trivially holds. Assume that (1) is true for $i-1$, namely, $|\mathcal{E}(\mathcal{X}_{i-1})| \geq |V(\mathcal{X}_{i-1})| - (k-1)$. In the ith step we add one new edge and at most one new vertex to $\mathcal{X}_{i-1}$ because e_i intersects $V(\mathcal{X}_{i-1})$ in at least $k-1$ vertices. Thus, $|\mathcal{E}(\mathcal{X}_i)| \geq |V(\mathcal{X}_i)| - (k-1)$ also holds. $\square$

All of the hypergraphs $(V_i, \mathcal{E}_i)$ meet the conditions of Lemma 1, so $|\mathcal{E}_i| \geq |V_i| - (k-1)$ for $i = 1, 2, \ldots, l$. Therefore,

$$|\mathcal{E}| = \sum_{i=1}^{l} |\mathcal{E}_i| \geq \sum_{i=1}^{l} (|V_i| - (k-1)) = \left(\sum_{i=1}^{l} |V_i| \right) - l(k-1).$$

Let σ denote $\sum_{i=1}^{l} |V_i|$. In order to prove $|\mathcal{E}| \geq n - (k-1)$ it is enough to show that $\sigma - l(k-1) \geq n - (k-1)$, or equivalently

$$\sigma \geq n + (l-1)(k-1). \tag{2}$$

We note that $\sigma \geq n$ certainly holds since every vertex of $\mathcal{H}$ is covered by one of the classes.

We may assume that $|V_i| < n$ for every i, otherwise $|\mathcal{E}| \geq |\mathcal{E}_i| \geq |V_i| - (k-1) = n - (k-1)$ and we are done. Let $r = \min_{x \in V} |\{i \colon x \in V_i\}|$. This parameter will help us in splitting the problem into two parts. We may assume that $r \geq 2$, otherwise there would be a vertex x which is covered by only one class, say V_i, but for every $y \in V$ the pair $\{x, y\}$ of vertices has to be covered by a class too, which means that $V_i = V$, contradicting our previous assumption.

To complete the proof, we only need two inequalities on σ.

Lemma 2.

1. *$\sigma \geq rn$,*
2. *$\sigma \geq n + r - 1 + (l-r)k$.*

Proof.

We can compute σ in two ways. We can sum up the sizes of the classes or for all of the vertices we can sum up the number of classes containing them.

In the first case, we obtain

$$\sigma = \sum_{x \in V} |\{i\colon x \in V_i\}| \geq rn,$$

which verifies the first inequality.

In the second case, we choose a vertex x such that it is contained in exactly r classes. The union of the classes containing x must be V since they have to cover every pair of vertices. Now,

$$\sigma = \sum_{i\colon x \in V_i} |V_i| + \sum_{i\colon x \notin V_i} |V_i|.$$

On one hand, $\sum_{i\colon x \in V_i} |V_i| \geq n + r - 1$ since we count every vertex at least once and x exactly r times. On the other hand, $\sum_{i\colon x \notin V_i} |V_i| \geq (l-r)k$ because the size of every class is at least k. Combining these inequalities we obtain $\sigma \geq n + r - 1 + (l-r)k$, which proves the second part of the lemma. □

Using Lemma 2, it is enough to show that either $rn \geq n + (l-1)(k-1)$ or $n + r - 1 + (l-r)k \geq n + (l-1)(k-1)$ holds since both inequalities imply (2), thus $m \geq n - (k-1)$.

It is easy to see that

$$rn \geq n + (l-1)(k-1) \Leftrightarrow l - 1 \leq \frac{(r-1)n}{k-1},$$

and

$$n + r - 1 + (l-r)k \geq n + (l-1)(k-1) \Leftrightarrow l - 1 \geq (r-1)(k-1).$$

Therefore, $|\mathcal{E}| \geq n-(k-1)$, does not hold only if $\frac{(r-1)n}{k-1} < l-1 < (r-1)(k-1)$, which implies $n < (k-1)^2$ (we have assumed $r \neq 1$ above). So, if $n \geq (k-1)^2$, then $|\mathcal{E}| \geq n - (k-1)$. □

From the proof of Theorem 1, $n_0(k) \leq (k-1)^2$ follows. We will further discuss bounds on $n_0(k)$ in Subsection 2.1. Now, let us continue with the proof of Theorem 2. Before the actual proof, we state two basic properties of semicycles and chains. These claims will come in handy during the proof.

Claim 3. *A semicycle has at least 3 edges.*

Claim 4. *If a hypergraph is semicycle-free, then it does not contain a self-intersecting chain.*

The graph-counterpart of Claim 4 would be the following: if a graph is cycle-free, it does not contain a closed walk. Claim 4 is important because it guarantees that as long as we are concerned with semicycle-free hypergraphs, we do not have to deal with self-intersecting chains (remember that a chain can traverse a point more than one times by Definition 1, causing unwanted complexity). So, choosing the semicycle as the generalisation of the graph-theoretical cycle turns out to be a fruitful idea here.

Proof. [Theorem 2]
The proof of Theorem 2 consists of two parts: the first part is dedicated to proving the upper bound, while the second part shows the asymptotic sharpness in 3-uniform case.

Part 1. We give an injective function $\varphi\colon \mathcal{E} \to \binom{V}{k-1}$ below.

Let us construct a maximal chain $\mathcal{P}$ in $\mathcal{H}$ (i.e. it cannot be continued with more edges). Such a chain exists because $\mathcal{H}$ is semicycle-free. Starting with any edge, we can construct $\mathcal{P}$ edge-by-edge by adding a new edge to it in every step. Claim 4 implies that we must add an unseen vertex to the chain in every step, hence, sooner or later we run out of vertices, and we cannot extend our chain anymore.

Let us take the last edge of $\mathcal{P}$, and assign the set consisting of the last $k-1$ vertices of the chain to this edge. Obviously, this $(k-1)$-set is contained only in the edge to which we assigned it, otherwise the chain could be further continued or a semicycle would occur.

Now, let us delete this edge from $\mathcal{H}$, and repeat the same process to the remaining hypergraph (i.e. find a maximal chain, define φ on the last edge before deleting it). We can do this because the subhypergraphs inherit the semicycle-free property. We continue deleting edges until there are no more left, i.e. φ is defined everywhere.

It remains to prove that φ is injective. Assume indirectly that we assign the same vertex set T to two different edges e_1 and e_2. Without loss of generality, we may assume that during the assigning process we have deleted e_1 earlier than e_2. However, when we delete e_1, $\varphi(e_1)$ is contained in two edges, e_1 and e_2, which contradicts the definition of φ.

Because of the injectivity of φ, we have $|\mathcal{E}| = |\varphi(\mathcal{E})| \leq \binom{n}{k-1}$, which is what we wanted to prove. $\square$

Part 2. The following construction proves the asymptotic sharpness of the upper bound in 3-uniform case. Actually, we prove a little bit more: the bound is asymptotically the best possible for 3-uniform hypertrees.

Let $V = \{v_i\}_{i=1}^n$ be a set of vertices and $\mathcal{H} = (V, \mathcal{E})$ be an arbitrary 3-uniform hypertree. Now let us define the 3-uniform hypergraph $B(\mathcal{H}) = (V \cup V', \mathcal{E} \cup \mathcal{E}')$, where $V' = \{0,1\}^n$ and

$$\mathcal{E}' = \left\{ \{v_i, u, w\} \,\middle|\, \begin{array}{l} v_i \in V,\ u, w \in V' \text{ and the } i\text{th bit is the} \\ \text{first bit where } u \text{ and } w \text{ differ} \end{array} \right\}.$$

Lemma 5. *If $\mathcal{H}$ is a hypertree, then $B(\mathcal{H})$ is a hypertree as well.*

Proof.
Chain-connectivity. Any two vertices from V are connected by a chain because $\mathcal{H}$ is a hypertree and all of its edges are edges $B(\mathcal{H})$ as well.

For $u, w \in V'$ let i denote the position of the first bit where they differ. Then, by definition, $\{v_i, u, w\} \in \mathcal{E}'$, thus u and w are connected by a chain of length one in $B(\mathcal{H})$.

In case of $u \in V'$, $v_i \in V$ consider a vertex $w \in V'$ whose first $i-1$ bits are the same as the first $i-1$ bits of u, but it differs in the ith bit. Then, by definition, $\{v_i, u, w\} \in \mathcal{E}'$, so u and v_i are connected with a chain of length one.
Semicycle-freeness. Assume indirectly that $B(\mathcal{H})$ contains a semicycle $\mathcal{C}$.

Notice that all edges of $\mathcal{C}$ belong to either $\mathcal{E}$ or $\mathcal{E}'$ since

$$\forall\, e \in \mathcal{E}, e' \in \mathcal{E}' : |e \cap e'| \leq 1. \tag{3}$$

In the former case, $\mathcal{C}$ would be a semicycle in $\mathcal{H}$, which is in contradiction with the hypertree property.

Hence, all edges of $\mathcal{C}$ are contained in $\mathcal{E}'$. Notice that only one edge can contain a pair $\{u, w\} \subseteq V'$ since the first place where the bits of u and v differ, uniquely determines the third vertex of the edge containing $\{u, w\}$, which is in V. So, if two adjacent vertices of the semicycle are in V', then these vertices are the last two vertices of the semicycle since only one edge of $\mathcal{C}$ can contain them.

Let $\{v_{i_1}, u_1, u_2\}$ be the first edge of $\mathcal{C}$. If we write down the vertices of the semicycle in a sequence, denoting the vertices from V by v_i and those from V' by u_j, there are 3 possible sequences (the point at the end indicates that the given sequence cannot be continued):

- $v_1 u_1 u_2 \cdot$: this obviously cannot be a semicycle.
- $u_1 v_1 u_2 u_3 \cdot$: still cannot be a semicycle, it must have at least three edges because of Claim 3.
- $u_1 u_2 v_1 u_3 u_4 \cdot$: there are three edges now, but to form a semicycle, the first and the last vertices must be identical. This is impossible since u_1 and u_2, u_2 and u_3, u_3 and u_4 differs in the ith bit, thus u_1 and u_4 differs in the ith bit too.

Here, we have taken into consideration that every edge of $\mathcal{C}$ comes from $\mathcal{E}'$, and such an edge contains two vertices from V'.

We get a contradiction again, thus $B(\mathcal{H})$ must be a hypertree. $\square$

Let us count the number of vertices and edges of the hypertree obtained in this way. $|V \cup V'| = n + 2^n$, $|\mathcal{E} \cup \mathcal{E}'| \geq |\mathcal{E}'| = \binom{2^n}{2}$ since exactly one edge of $\mathcal{E}'$ belongs to every pair of vertices of V'. So,

$$\frac{|\mathcal{E}(B(\mathcal{H}))|}{\binom{|V(B(\mathcal{H}))|}{2}} \geq \frac{\binom{2^n}{2}}{\binom{n+2^n}{2}} \to 1 \text{ as } n \to \infty.$$

The fraction cannot exceed 1 either because of the upper bound we have just proven in Part 1. Thus, if $\{\mathcal{H}_i\}_{i=1}^{\infty}$ is a hypertree sequence such that $\lim_{i\to\infty} |V(\mathcal{H}_i)| = \infty$, then $|\mathcal{E}(B(\mathcal{H}_i))| \sim \binom{|V(B(\mathcal{H}_i))|}{2}$. □

We can apply the operator $B()$ more times on a hypertree in order to rapidly get close to the upper bound.

Theorem 2, in particular, is also true for hypertrees as Part 2 of the proof shows. It is an interesting question whether this construction can be generalised to k-uniform case.

Theorem 2 remains true even if we require $\mathcal{H}$ to be only cycle-free because we can also find a maximal chain in that case (otherwise there would be a $(k-1)$-tuple of vertices, which appears at least twice as $k-1$ consecutive vertices of the chain). This implies that from the viewpoint of the edge number, requiring cycle-free or semicycle-free property does not mean a big difference: in 3-uniform case, cycle-freeness would already guarantee that the number of edges is at most $\binom{n}{2}$, and not even the stronger semicycle-free property can lower this bound.

We close this section with an important consequence of Theorem 2. At the end of Section 3 we will show how this result might be a part of a bigger picture.

Claim 6. *If $\mathcal{H}$ is a 3-hypertree, then $\mathcal{B}(\mathcal{H})$ is a 3-hypertree as well.*

Proof. Let $\mathcal{H} = (V, \mathcal{E})$ and $B(\mathcal{H}) = (V \cup V', \mathcal{E} \cup \mathcal{E}')$ be as in Part 2 of the proof of Theorem 2. We have already shown that $B(\mathcal{H})$ is a hypertree, so we only need to prove that every chain in it is of length at most 3.

Observation (3) implies that the edges of a chain $\mathcal{P}$ is contained either in $\mathcal{E}$ or in $\mathcal{E}'$. In the former case, $\mathcal{P}$ is a chain in $\mathcal{H}$, thus it is of length at most 3. In the latter case, $\mathcal{P}$ can have at most 5 vertices (thus 3 edges) because of the case analysis we have encountered at the end of the proof of Theorem 2. □

2.1. Small counterexamples to Theorem 1

Inspired by Theorem 1, we call a chain-connected k-uniform n-vertex hypergraph $\mathcal{H}$ a *counterexample* if $|\mathcal{E}(\mathcal{H})| < n - k + 1$. A natural question is that what are the sharp lower and upper bounds on the order of a k-uniform counterexample and in particular, whether a k-uniform counterexample exists (i.e. $n_0(k) = 1$). We have partial answers for these questions. Also, notice that any upper bound on the order of the counterexamples is a good choice for $n_0(k)$ in Theorem 1.

Claim 7.

- *For $k = 3, 4, 5$, there are no k-uniform counterexamples.*
- *For each $k \geq 6$, there exists a k-uniform counterexample.*

Proof. The proof of the first part is a fairly straightforward case analysis using combinatorial arguments and relations between the parameters n, m, k, r and l

(see the definitions of parameters r and l in the proof of Theorem 1), hence it is left to the reader.

The following simple construction verifies the second part. Let $V' = \{v_1, v_2, \ldots, v_{k-6}\}$, $V_x = \{x_1, x_2, x_3\}$, $V_y = \{y_1, y_2, y_3\}$, $V_z = \{z_1, z_2, z_3\}$, $V = V_x \cup V_y \cup V_z \cup V'$ and $\mathcal{E} = \{V_x \cup V' \cup V_y, V_x \cup V' \cup V_z, V_y \cup V' \cup V_z\}$. Then, the hypergraph $\mathcal{F} = (V, \mathcal{E})$ is chain-connected and a counterexample for $k \geq 6$ because it has $k+3$ vertices and 3 edges, hence $|\mathcal{E}| = 3 < 4 = n - (k-1)$. □

Let us see, how many vertices a counterexample can have for a fixed k.

Theorem 4.

- *If $k \geq 6$, then there exists a k-uniform counterexample of order $k+3$; this is the minimal order of counterexamples;*
- *if $k \geq 6$ is even, then there exists a k-uniform counterexample of order $\frac{k(k-2)}{2}$;*
- *if $k \geq 6$ is odd, then there exists a k-uniform counterexample of order at least $\frac{(k-1)(k-4)}{2} + 1$;*

Proof. The hypergraph $\mathcal{F}$ from the previous construction proves the first part. It is easy to see that $k+3$ is a sharp lower bound since every chain-connected hypergraph with 2 edges must have $k+1$ vertices, thus a counterexample with $k+2$ vertices cannot exist.

As for the second part, we define $c = c(k)$ clusters of vertices, each of size $\frac{k}{2}$ such that each pair of clusters form an edge. Now, we have $n = \frac{ck}{2}$ vertices and $|\mathcal{E}| = \binom{c}{2}$ edges. The condition being a counterexample can be formulated as

$$\binom{c}{2} < \frac{ck}{2} - k + 1.$$

To maximize n for a fixed k, we must maximize c subject to the constraint above. The parameter c attains its maximum at $k-2$, and the hypergraph (which is obviously chain-connected) obtained in this way has $\frac{k(k-2)}{2}$ edges.

The third part of the claim can be proven with a minor modification of the proof of the second part. Take the construction above with c clusters, each of size $\frac{k-1}{2}$. Now, add an extra vertex to all edges, and maximize c subject to the appropriate constraint. □

We note that these counterexamples are all hypertrees, but semicycle-freeness does not play any particular role here. The asymptotically sharp upper bound on the order of k-uniform counterexamples remains the most important open question of this topic.

3. Proof of Theorem 3

In this section, we prove Theorem 3, an upper bound on the number of edges of l-hypertrees, which is an interesting addition to Theorem 2.

Proof. [Theorem 3] The proof is the direct generalization of Part 1 of the proof of Theorem 2. We give $k-l+1$ injective functions, $\varphi_1, \varphi_2, \dots, \varphi_{k-l+1} \colon \mathcal{E} \to \binom{V}{k-1}$ for which $\mathrm{Im}\varphi_i \cap \mathrm{Im}\varphi_j = \emptyset$ hold for all $1 \le i < j \le k-l+1$.

Let $\mathcal{P}$ be a maximal chain in $\mathcal{H}$. Such a chain exists as we have seen in the proof of Theorem 2. Let e be the last edge of $\mathcal{P}$. The length of this chain is at most $l \le k$, therefore there are at least $k-l+1$ vertices contained in the intersection of all edges of $\mathcal{P}$. Let us denote this intersection by U. If $u \in U$ and some edge $f \in \mathcal{E}$, $f \ne e$ covers $e \backslash \{u\}$, then $\mathcal{P}$ could be continued by f (because we have a freedom in ordering the elements of U, and $\mathcal{H}$ is semicycle-free), which is a contradiction.

Let $u_1, \dots, u_{k-l+1} \in U$ be distinct vertices and define $\varphi_i(e)$ to be $e \backslash \{u_i\}$, for $i = 1, 2, \dots, k-l+1$. Then, delete e and repeat the whole process for the remaining hypergraph. In the end, φ_i will be defined everywhere, for $i = 1, 2, \dots, k-l+1$. The previous observation shows that the conditions required for the φ_is are satisfied.

So, $(k-l+1)|\mathcal{E}| = |\mathrm{Im}\varphi_1| + \ldots + |\mathrm{Im}\varphi_{k-1}| = |\mathrm{Im}\varphi_1 \cup \ldots \cup \mathrm{Im}\varphi_{k-1}| \le \binom{n}{k-1}$, hence $|\mathcal{E}| \le \frac{1}{k-l+1}\binom{n}{k-1}$. $\square$

Semicycle-freeness is a crucial premise of this proof, and there is no trivial extension to the cycle-free case because a maximal chain of length l in a cycle-free hypergraph might be extended with an edge without forming a longer path (forming a semicycle of length $l+1$ instead).

The extremal values of l are quite interesting. Notice that by our definitions, $S(j, k, n)$ Steiner systems are 1-hypertrees. In the case of $l = 1$, Theorem 3 gives the upper bound $\frac{1}{k}\binom{n}{k-1}$ which is best possible because $S(k-1, k, n)$ Steiner systems have exactly that many edges. The existence of such Steiner systems for arbitrarily large n, which was in fact a long-standing open problem, was recently solved by Peter Keevash in [4].

In the case of $l = k$, Theorem 3 falls back to the upper bound seen in Theorem 2. Moreover, the upper bound can be actually approximated by 3-hypertrees in 3-uniform case, as Claim 6 shows. Based on these observations, we formulate the following conjecture.

Conjecture 8. *The upper bound of Theorem 2 can be reached by a sequence of k-hypertrees.*

Interestingly, even in the first nontrivial case, $l = 2$, the asymptotic sharpness of Theorem 3 is still unknown. Though, we have investigated many constructions for 2-hypertrees based on Steiner systems, the best asymptotic edge-number achieved is still closer to $\frac{1}{k}\binom{n}{k-1}$ than to $\frac{1}{k-1}\binom{n}{k-1}$.

4. Open problems

First of all, we would like to enlist the open question related to the main results.

- Determine the maximal size of counterexamples to Theorem 1 for all n. What we know now is that $n_0(k) \sim ck^2$, where $\frac{1}{2} \leq c \leq 1$.
- Refine the upper bound of Theorem 2 and show asymptotic sharpness for all k. For $k = 3$, the asymptotic sharpness was proven by Theorem 2. For all k, our best constructions are based on $S(k-1, k, n)$ Steiner systems with approximately $\frac{1}{k-1}\binom{n}{k-1}$ edges.
- Prove that the bound of Theorem 3 is asymptotically sharp for all k and l. The asymptotic sharpness is only partially known: in the cases $l = 1 \wedge k = 2$ and $k = l = 3$. In the case of $l = 2$, $S(k-1, k, n)$ Steiner systems are close to the bound.

Consider the following definitions, motivated by the gap between Theorem 1 and Theorem 2.

Definition 6 (Edge-minimal hypertree). *The hypergraph $\mathcal{H} = (V, \mathcal{E})$ is an* edge-minimal hypertree *if it is a hypertree, and deleting any edge $e \in \mathcal{E}$ from it, $\mathcal{H} \setminus \{e\}$ is not a hypertree any more (i.e. chain-connectivity does not hold).*

Definition 7 (Edge-maximal hypertree). *The hypergraph $\mathcal{H} = (V, \mathcal{E})$ is an* edge-maximal hypertree *if it is a hypertree, and adding any new edge $e \notin \mathcal{E}$ to it, $\mathcal{H} \cup \{e\}$ is not a hypertree any more (i.e. semicycle-freeness does not hold).*

Edge-minimal and edge-maximal hypertrees are the extremal cases of hypertrees, where the edge-set is minimal/maximal with respect to inclusion. Examples of edge-minimal hypertrees are non-self-intersecting chains, tight stars and $S(2, k, n)$ Steiner systems.

There are two important related extremal problems.

- What is the maximal number of edges in a k-uniform edge-minimal hypertree of order n?
- What is the minimal number of edges in a k-uniform edge-maximal hypertree of order n?

Based on our research, we propose the following two conjectures in 3-uniform case.

Conjecture 9. *Every 3-uniform edge-minimal hypertree has at most $\frac{1}{2}\binom{n}{2}$ edges.*

Conjecture 10. *Every 3-uniform edge-maximal hypertree on n vertices has at least $\frac{1}{2}\binom{n}{2} - \varepsilon(n)$ edges, where $\epsilon(n) = O(n)$.*

A special kind of 3-uniform hypertree is the simultaneously edge-minimal and edge-maximal 3-uniform hypertree which have nearly $\frac{1}{2}\binom{n}{2}$ edges by Conjecture 9 and Conjecture 10. It seems possible that there are infinitely many such hypertrees, and this is another interesting open question.

Certain variations of these definitions might also be of some interest.

Instead of edge-minimal hypertrees, one can consider edge-minimal chain-connected hypergraphs or hypergraphs which can contain long semicycles, but no short ones. Similarly, edge-maximal semicycle-free hypergraphs can be compared to edge-maximal hypertrees.

We could have allowed a chain to use an edge more times. In this case, one can show forbidden substructures in edge-minimal chain-connected hypergraphs such as the complete hypergraph $K_{k+2}^{(k)}$.

In conclusion, hypertrees are new, interesting, non-trivial generalizations of trees, with a both nice analogies and remarkable differences. Theorem 1 and Theorem 2 gives reasonable lower and upper bounds on the number of edges in a hypertree and also draw attention to the gap between these bounds, inspiring further study in the direction of edge-minimal and edge-maximal hypertrees. Another class of hard questions involve l-hypertrees which seem to be closely related to Steiner systems. We believe that the sheer amount of intriguing open problems in the topic makes it well worth of future study.

Acknowledgement

We would like to thank the anonym referees of our paper for their valuable advices.

The work reported in the paper has been developed in the framework of the project "Talent care and cultivation in the scientific workshops of BME" project. This project is supported by the grant TÁMOP - 4.2.2.B-10/1–2010-0009.

The first author is partially supported by the Hungarian National Research Fund (grant number NK 78439).

The authors are partially supported by the Hungarian National Research Fund (grant number K108947).

[1] C. Berge: Hypergraphs, North-Holland, Amsterdam, (1989) 389–396.

[2] H. Hàn, M. Schacht: Dirac-type Results for Loose Hamilton Cycles in Uniform Hypergraphs, *Journal of Combinatorial Theory, Series B* **100** (2010) 332–346.

[3] G. Y. Katona, H. Kierstead: Hamiltonian Chains in Hypergraphs, *Journal of Graph Theory* **30** (1999) 205–212.

[4] P. Keevash, *The existence of designs*, 2014, arXiv:`1401.3665 [math.CO]`

[5] D. Kühn, D. Osthus: Loose Hamilton Cycles in 3-uniform Hypergraphs of High Minimum Degree, *Journal of Combinatorial Theory, Series B* **96** (2006) 767–821.

[6] D. Kühn, R. Mycroft, D. Osthus: Hamilton l-cycles in k-graphs, *J. Combin. Theory Ser. A* **117** (2010) 910–927.

[7] V. Rödl, A. Ruciński, E. Szemerédi: An Approximate Dirac-Type Theorem for k-Uniform Hypergraphs, *Combinatorica* **28** (2) (2008) 229–260.

[8] V. Rödl, A. Ruciński, E. Szemerédi: Dirac-type conditions for Hamiltonian paths and cycles in 3-uniform hypergraphs, *Advances in Mathematics* **227** (2011) 1225–1299.