Bounding the Number of Minimal Transversals in Tripartite 3-Uniform Hypergraphs

Alexandre Bazin — **Le2i - Université de Bourgogne**
Laurent Beaudou — **LIMOS - Université Clermont Auvergne**
Giacomo Kahn[1] — **LIMOS - Université Clermont Auvergne**
Kaveh Khoshkhah — **LIMOS - Université Clermont Auvergne**

Abstract

We bound the number of hypergraph transversals that arise in tripartite 3-uniform hypergraphs, a class commonly found in data mining and other applications. It is more than $c3.35^{(n/3)}$ and less than $3.45^{n/3}$.

1 Introduction

A *triadic binary dataset* is a tuple $(S_1, S_2, S_3, \mathcal{R})$ where S_1, S_2, S_3 are sets and $\mathcal{R}$ is a ternary relation between them. A *closed 3-itemset* is a maximal box of incidence in the dataset, up to permutation of rows, columns and layers. More formally, a closed 3-itemset is a triple (X_1, X_2, X_3) where $X_i \subseteq S_i$, $X_1 \times X_2 \times X_3 \subseteq \mathcal{R}$, and all components are maximal for this property.

Closed 3-itemsets are central in some fields of data mining, and their enumeration is an interesting problem [3]. Few results exists on this topic, and in particular no non-trivial upper bound is known.

Hypergraphs are generalizations of graphs, where edges may have arity different than 2. Formally, a hypergraph $\mathcal{H} = (V, \mathcal{E})$ is a pair where V is a set of vertices and $\mathcal{E}$ is a set of subsets of V called (hyper)edges. If all the edges of a hypergraph are of size k, we call it a k-uniform hypergraph.

Binary datasets are essentially hypergraphs. For example, in a 2-dimensional setting, a dataset is a bipartite graph, and its closed 2-itemsets are maximal complete bipartite subgraphs. In the 3-dimensional case, a dataset is a tripartite, 3-uniform hypergraph and its closed 3-itemsets are then complete tripartite sub-hypergraphs.

A subset S of the vertices is called a *transversal* of $\mathcal{H}$ if it intersects every edge of $\mathcal{H}$. A transversal of $\mathcal{H}$ is minimal if no subset of S is a transversal.

When considering the complement hypergraph of a dataset, each edge represents a "hole" in the dataset. In this representation, the closed itemsets of the dataset are minimal transversals of the hypergraph.

In the following, we use n to denote the number of vertices of the hypergraph we are considering. We are interested in the maximal number of minimal transversals in 3-uniform, tripartite hypergraphs with n vertices.

In the 2-dimensional case it is known that the maximal number of closed itemsets (and thus bicliques in the associated bipartite graphs) is bounded by $2^{n/2}$. It is attainable when n is even and the graph can be partitioned into two independent sets of size $n/2$. The graph that has $2^{n/2}$

[1]This research was partially supported by the European Union's "*Fonds Européen de Développement Régional (FEDER)*" program.

minimal transversals (or minimal dominating sets) is a matching. This can be generalized to 3-dimensional datasets and produces $3^{n/3}$ itemsets. In this case, the hypergraph is a set of disjoint hyperedges of arity 3.

The datasets that generate each case are shown in Fig. 1.

	a	b	c	d
1		×	×	×
2	×		×	×
3	×	×		×
4	×	×	×	

	a	b	c	a	b	c	a	b	c
α		×	×	×	×	×	×	×	×
β	×	×	×	×		×	×	×	×
γ	×	×	×	×	×	×	×	×	
		1			2			3	

Figure 1: A dataset $(Numbers, Latin, \mathcal{R})$ with 2^4 closed 2-itemsets (left) and a dataset $(Number, Greek, Latin, \mathcal{R})$ with 3^3 closed 3-itemsets (right).

In a tripartite, 3-uniform hypergraph of size n, $2^{2n/3} = 4^{n/3}$ is a trivial upper bound on the number of minimal transversals. Indeed, partitioning V into three independent sets and choosing elements from two of them to construct a minimal transversal leaves only one possible choice in the third one. By supposing that all possible subsets of the two smallest independent sets can appear, we obtain $2^{n/3} \times 2^{n/3} \times 1$ minimal transversals.

Let us denote by $\mathscr{N}$ the maximal number of minimal transversals in all triadic hypergraphs of size n. We know that $3^{n/3} \leq \mathscr{N} \leq 4^{n/3}$.

In the following, we sketch the proofs for Theorem 1 and Theorem 2.

Theorem 1. *Let $r = (n/3) \mod 5$. There exists a tripartite, 3-uniform hypergraph of size n with $3^r \times 3.35^{(n/3)-r}$ hypergraph transversals.*

Theorem 2. *A tripartite, 3-uniform hypergraph with n vertices has at most $1.85532\ldots^{2n/3} = 3.44221\ldots^{n/3} \leq 3.45^{n/3}$ minimal transversals.*

2 Lower Bound

Lemma 3. *If there exists a dataset of size $k \times k \times k$ with N closed 3-itemsets and a dataset of size $k' \times k' \times k'$ with N' closed 3-itemset, then there exists a dataset of size $k'' \times k'' \times k''$ with $k'' = k + k'$ with $N \times N'$ closed 3-itemsets.*

This lemma is illustrated in 2 dimensions in Fig. 2. The dataset $\mathcal{C} = (S_1, S_2, S_3, \mathcal{R})$ is constructed from the datasets $\mathcal{C}^1 = (S_1^1, S_2^1, S_3^1, \mathcal{R}^1)$ and $\mathcal{C}^2 = (S_1^2, S_1^2, S_1^2, \mathcal{R}^2)$ in this way :

$$\mathcal{S}_i = \mathcal{S}_i^1 \cup \mathcal{S}_i^2, \forall i \in \{1, 2, 3\}$$

and

$$\begin{cases} (x_1, x_2, x_3) \in \mathcal{R} \text{ if} (x_1, x_2, x_3) \in \mathcal{R}^1 \\ (x_1, x_2, x_3) \in \mathcal{R}^2 \\ \exists i, j \in \{1, 2, 3\}, i \neq j \text{ such that } x_i \in \mathcal{S}_i^1, x_j \in \mathcal{S}_j^2 \end{cases}$$

This construction allows to consider the hypergraph transversals in each separate tripartite 3-uniform hypergraph, and combine the results. In small instances, an exhaustive search allows

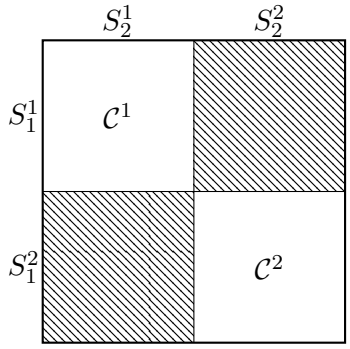


Figure 2: The striped areas contain only crosses. In this dataset, a maximal box of incidence in $\mathcal{C}^1$ can be combined with any maximal box of incidence in $\mathcal{C}^2$.

to find some tripartite 3-uniform hypergraphs with more than $3^{n/3}$ hypergraph transversals. In a $5 \times 5 \times 5$ dataset, there is an instance with $428 \approx 3.35^5$ hypergraph transversals.

By using the multiplicative construction on $3 \times 5k$ vertices, and adding r disjoint edges of arity 3, with $r = (n/3) \mod 5$, we can construct a hypergraph that shows Theorem 1. The resulting structure has $3(5k + r) = n$ vertices, and $3^r \times 3.35^{(n/3)-r}$ minimal transversals.

3 Upper Bound

Our proof for the upper bound on the number of minimal transversals in tripartite 3-uniform hypergraphs relies on principles introduced in [1] and is similar to the proof given in [2].

A condition is a pair (A^+, A^-) of disjoint subsets of vertices. A minimal transversal T satisfies a condition (A^+, A^-) if it contains all the vertices in A^+ and none of the vertices in A^-. A condition is non-trivial if $A^+ \cup A^- \neq \emptyset$, which we will suppose in the rest of this paper.

Let $\mathcal{H}$ be a hypergraph and A a condition. The hypergraph $\mathcal{H}_A$ is constructed from $\mathcal{H}$ and A using the following procedure :

1. remove every edge E that contains a vertex that is in A^+;
2. remove from every remaining edge the vertices that are in A^-;
3. remove redundant edges.

Clearly, T is a minimal transversal of $\mathcal{H}_A$ if $T \cup A$ is a minimal transversal of $\mathcal{H}$. A family of conditions is complete if it is non-empty, non-trivial, and every transversal of $\mathcal{H}$ satisfies at least one condition of the family.

Let $\mathscr{C}$ be the class of tripartite hypergraphs in which edges are at most of size 3 and there is a set I of independent vertices that intersects all the edges. If no vertex of I appears in conditions, then $\mathscr{C}$ is closed for the aforementioned procedure.

Let ρ be a function that assigns to each hypergraph $\mathcal{H}$ in $\mathscr{C}$ a complete family of conditions. For each such hypergraph, we define a labeled tree $\mathcal{T}_\mathcal{H}$ such that if $\emptyset \in \mathcal{E}$ or $\mathcal{E} = \emptyset$ then $\mathcal{T}_\mathcal{H}$ is a single node labeled $\mathcal{H}$, and otherwise, we create a node labeled $\mathcal{H}$ that is the parent of all trees $\mathcal{T}_{\mathcal{H}_A}$ for all $A \in \rho(\mathcal{H})$.

Proposition 4. *Let ρ be a function that assigns to each hypergraph $\mathcal{H} \in \mathscr{C}$ a complete family of condition, where $\mathscr{C}$ is a class closed by removing edges and removing vertices from edges. Then for all hypergraphs $\mathcal{H} \in \mathscr{C}$, the number of transversals is bounded by the number of leaves of $\mathcal{T}_{\mathcal{H}}$.*

In [1], the author gives a way to estimate (bound) the number of leaves in a rooted tree, when some weight assigned to the edge from an internal node of the tree to each of its children sums to 1. We use this result to estimate the number of leaves in our tree $\mathcal{T}_{\mathcal{H}}$.

Let $\mu(\mathcal{H})$ be a real number, associated with a hypergraph $\mathcal{H}$. We call $\mu(\mathcal{H})$ the *measure* of hypergraph $\mathcal{H}$. We denote by $\Delta(\mathcal{H}, \mathcal{H}_A) = \mu(\mathcal{H}) - \mu(\mathcal{H}_A)$ the weight of the edge between $\mathcal{H}$ and $\mathcal{H}_A$ in $\mathcal{T}_{\mathcal{H}}$.

Proposition 5. *Let $\mathcal{H} \in \mathscr{C}$, then if $0 \leq \mu(\mathcal{H}) \leq |V(\mathcal{H})|$ and for all $A \in \rho(\mathcal{H}), \mu(\mathcal{H}_A) \leq \mu(\mathcal{H})$, there is a unique positive real number $\tau \geq 1$ such that*

$$\sum_{A \in \rho(\mathcal{H})} \tau^{-\Delta(\mathcal{H}, \mathcal{H}_A)} = 1$$

We say that ρ is μ-compatible in this case.

From here on, we say that the function ρ is μ-bounded by τ_0 if for all hypergraph $\mathcal{H}$ in $\mathscr{C}$, $\tau_{\mathcal{H}} \leq \tau_0$.

Let μ be a measure, and let ρ be a function that assigns a complete family of conditions to each hypergraph $\mathcal{H}$ in a class $\mathscr{C}'$ closed under the aforementioned procedure. If ρ is μ-compatible and μ-bounded by τ_0, then for every hypergraph $\mathcal{H}$ in $\mathscr{C}'$, we have that the number of leaves of $\mathcal{T}_{\mathcal{H}}$ is bounded by $\tau_0^{H(\mathcal{T}_{\mathcal{H}})}$ where $H(\mathcal{T}_{\mathcal{H}})$ is the height of $\mathcal{T}_{\mathcal{H}}$.

Lemma 6. *There is a measure μ defined for every hypergraph $\mathcal{H} \in \mathscr{C}$ and a function ρ for $\mathscr{C}$ that is μ compatible, μ-bounded by* $1.85532\ldots$.

The proof of this lemma requires a case analysis that is too long to be part of this short paper. Lemma 6 concludes the proof for Theorem 2.

4 Conclusion

By combining Theorem 1 and Theorem 2, we have that $3^4 \times 3.35^{(n/3)-4} \leq \mathscr{N} \leq 3.44221^{n/3}$. Theorem 2 yields the following corollary.

Corollary 7. *Every triadic binary dataset of size n has at most $3.44221\ldots^{n/3}$ closed 3-itemsets.*

References

[1] O. Kullmann. New Methods for 3-SAT decision and Worst-Case Analysis. *Theoretical Computer Science* 223 (1-2), 1-72, 1999.

[2] Z. Lonc, M. Truszczyński On the Number of Minimal Transversals in 3-Uniform Hypergraphs. *Discrete Mathematics* 308 (16), 3668-3687, 2008

[3] L. Wei, T. Qian, Q. Wan, J. Qi. A Research Summary about Triadic Concept Analysis. *International Journal of Machine Learning and Cybernetics*, 2016