



HAL
open science

Interval Hypergraphs and Catalan Numbers

Pascal Pr ea, Fran ois Brucker

► **To cite this version:**

Pascal Pr ea, Fran ois Brucker. Interval Hypergraphs and Catalan Numbers. [Research Report] Ecole Centrale Marseille (ECM); Aix-Marseille Universite. 2020. hal-02892780

HAL Id: hal-02892780

<https://hal.science/hal-02892780>

Submitted on 7 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.

Interval Hypergraphs and Catalan Numbers*

Pascal Préa François Brucker

Aix-Marseille Université, CNRS, Université de Toulon, LIS

École Centrale Marseille

Marseille, France

{pascal.prea, francois.brucker}@lis-lab.fr

Abstract

An *interval hypergraph* is a hypergraph $H = (V, E)$ such that there exists a linear order σ (which is said to be *compatible*) on V such that, when V is sorted along σ , every hyperedge is an interval of V . A *interval cover* of a set V is an interval hypergraph $H = (V, E)$ such that $\bigcup_{e \in E} e = V$ and $\forall e, e' \in E, e \not\subseteq e'$. In this note, we show that:

- Nearly all interval hypergraphs have only two compatible orders.
- The number of interval covers of $\{1, \dots, n\}$ is the Catalan number $C_n = \frac{1}{n+1} \binom{2n}{n}$.

keywords : Interval hypergraphs, Catalan numbers, PQ-trees, Robinson dissimilarities.

1 Definitions and notations

A *hypergraph* is a couple $H = (V, E)$ where V is a finite set whose elements are *vertices* and $E \subseteq 2^V$ is the *hyperedge* set. A hypergraph $H = (V, E)$ is an *interval hypergraph* if there exists a *compatible (linear) order* σ of V such that, when V is sorted along σ , all hyperedges are intervals. A interval hypergraph $H = (V, E)$ is a *interval cover* of V if $\bigcup_{e \in E} e = V$ and $\forall e, e' \in E, e \not\subseteq e'$. For $V' \subset V$, the *restriction* of $H = (V, E)$ to V' is the hypergraph $H|_{V'} = (V', E')$ where $E' = \{e \cap V' : e \in E\}$.

A *PQ-tree* T on a set V is a tree that represents a set of permutations on V , and if $H = (V, E)$ is an interval hypergraph, then the set of all compatible orders of H can be represented by a PQ-tree [1, 7]. The leaves of T are the elements of V , and the nodes of T are of two types: the *P-nodes* and the *Q-nodes*. We use the general convention of writing that is to represent P-nodes by circles, and Q-nodes by rectangles. On a P-node, one can apply any permutation of its children (equivalently, its children are not ordered). The children of a Q-node are ordered, and the only permutation we can apply on them is to reverse the order. For instance, the PQ-tree of figure 1 represents the set of permutations $\{(1,2,3,4,5), (1,3,2,4,5), (2,1,3,4,5), (2,3,1,4,5), (3,1,2,4,5), (3,2,1,4,5), (5,4,1,2,3), (5,4,1,3,2), (5,4,2,1,3), (5,4,2,3,1), (5,4,3,1,2), (5,4,3,2,1)\}$.

We say that a PQ-tree is *flat* if it has only one internal node, which is a Q-node. A flat PQ-tree represents only two permutations. Similarly, an interval hypergraph is *flat* if it has only two compatible orders, *i.e.* if its PQ-tree is flat. Conversely, a PQ-tree with one internal node, which is a P-node, is said to be *universal*. A universal PQ-tree represents all the permutations on S .

A *dissimilarity* on a finite set V is a symmetric function $d : V \times V \rightarrow \mathbb{R}^+$ such that $d(i, j) = 0$ if and only if $i = j$. Dissimilarities can be seen as a generalization of distances and can be represented by symmetric matrices on \mathbb{R}^+ with a null diagonal. For a dissimilarity d on a set V and $\delta \in \mathbb{R}$, the graph $G_d^\delta = (V, E_d^\delta)$ with $xy \in E_d^\delta \iff d(x, y) \leq \delta$ is a *threshold graph* of d . The set \mathcal{C}_d of all maximal cliques of all threshold

*This work was supported in part by ANR project DISTANCIA (ANR-17-CE40-0015).

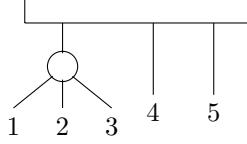


Figure 1: A PQ-tree

graphs of d is called the *class system* of d . Note that $H_d = (V, \mathcal{C}_d)$ is a hypergraph such that V and all singletons are hyperedges.

In the rest of this paper, we will only consider hypergraphs with vertex set $V = \{1, \dots, n\}$ and we will suppose that, if H is an interval hypergraph, then $(1, \dots, n)$ is one of its compatible orders. We will denote the number of interval hypergraphs on $\{1, \dots, n\}$ by I_n , the number of flat interval hypergraphs by F_n and the Catalan number $\frac{1}{n+1} \binom{2n}{n}$ by C_n .

2 Results on interval hypergraphs

Claim 1. $I_n = 2^{n(n+1)/2}$.

Proof. Since an interval on $\{1, \dots, n\}$ is either a pair of elements (its extremities) or a singleton, there are $n(n+1)/2$ interval in $\{1, \dots, n\}$. \square

Theorem 1. $\lim_{n \rightarrow \infty} F_n/I_n = 1$.

Proof. Let $H = (V, E)$ be an interval hypergraph. Suppose that its PQ-tree T_H has an internal node α , different from the root of T_H , and that the leaves under α are $i, \dots, i+k$. We have $1 < k < n-i$ and we set $V_\alpha := \{i, \dots, i+k\}$.

For every permutation σ represented by T_H , V_α is an interval of V and, if for some $e \in E$, there exists $x \in e \cap V_\alpha$ and $y \in e \setminus V_\alpha$, then $V_\alpha \subset e$. So H is entirely determined by $H|_{V_\alpha}$ and $H|_{(V \setminus V_\alpha) \cup \{i\}}$ and thus the number of interval hypergraphs with a node like α (i.e. such that the leaves under it are $i, \dots, i+k$) as internal node of their PQ-tree is $I_n^{[k]} := I_k \cdot I_{n-k+1}$. This value is maximum for $k = 2$ and we have $I_n^{[2]} = 2^{n(n-1)/2+3}$. So the proportion of interval hypergraphs having a PQ-tree with a node like α as leftmost internal node is $\leq \frac{1}{2^{n-3}}$. There are less than $(n-1)^2$ possible choices for the type of α (at most $n-1$ choices for i and $n-1$ choices for k), so the proportion of interval hypergraphs on $\{1, \dots, n\}$ which are compatible with a PQ-tree having internal nodes tends to 0 when n tends to ∞ . If the PQ-tree of an interval hypergraph H is universal, the hyperedges of H are either singletons or V . So there are only 2^{n+1} interval hypergraphs whose PQ-tree is universal, and the result follows. \square

Note that, even if we do not impose $(1, \dots, n)$ to be a compatible order, nearly all interval hypergraphs have only two compatible orders.

Theorem 2. *The number of interval covers of $\{1, \dots, n\}$ is C_n .*

Proof. There are C_n lattice paths in $\mathbb{Z} \times \mathbb{Z}$ from $(1, 1)$ to (n, n) with steps $(0, 1)$, $(1, 0)$ and $(1, 1)$ never going below the line $y = x$ and such that steps $(1, 1)$ only appear on the line $y = x$ (Exercice 52 of [9]). In such a path, a *summit* is a point with neither a step $(1, 0)$ just before it nor a step $(0, 1)$ just after.

If we associate with each summit (i, j) the interval $[i, j]$ (see Figure 2), we build a one-to-one correspondence between these paths and interval covers of $\{1, \dots, n\}$. \square

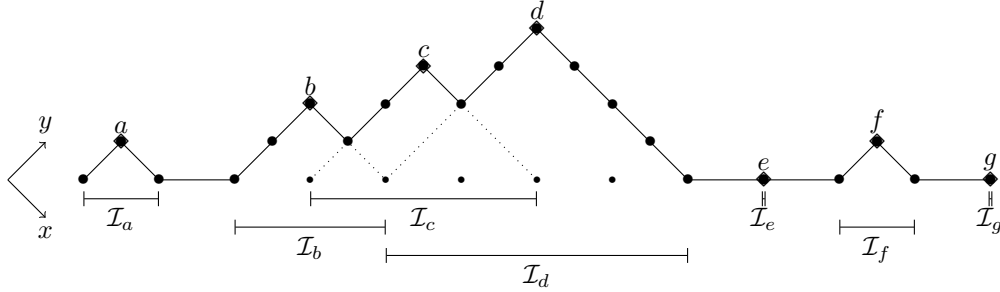


Figure 2: A lattice path on $\mathbb{Z} \times \mathbb{Z}$ from $(1, 1)$ to $(13, 13)$ with summits a, b, \dots, g and the corresponding intervals $\mathcal{I}_a, \mathcal{I}_b, \dots, \mathcal{I}_g$, which form an interval cover of the points $(1, 1), (2, 2), \dots, (13, 13)$.

3 Robinson dissimilarities

A dissimilarity d on a finite set V is *Robinson* [8] if there exists a linear order σ on V (which is said to be *compatible*) such that, when V is ordered along σ , $i < j < k \implies d(i, j) \leq d(i, k)$ and $d(j, k) \leq d(i, k)$. A dissimilarity d on V is Robinson if and only if (V, \mathcal{C}_d) is an interval hypergraph [3, 4]; in addition, d and (V, \mathcal{C}_d) have the same set of compatible orders [7]. We will identify two dissimilarities d and d' if $\mathcal{C}_d = \mathcal{C}_{d'}$; and, as for interval hypergraphs, we will suppose that $V = \{1, \dots, n\}$ and $(1, \dots, n)$ is a compatible order.

Claim 2. *The number of Robinson dissimilarities on V is $2^{n(n-1)/2-1}$*

Proof. $V \cup \{V\}$ is always a subset of \mathcal{C}_d , so the number of intervals that can be, or not, in \mathcal{C}_d is $n(n-1)/2 - 1$. \square

Theorem 3. *The number of Robinson dissimilarities on $\{1, \dots, n\}$ with values in $\{0, 1, 2\}$ is $C_n - 1$.*

Proof. When d takes its values in $\{0, 1, 2\}$, the relation R on V defined by $iRj \iff d(i, j) \leq 1$ is reflexive, symmetric and such that if $1 \leq i < j < k \leq n$, and iRk , then iRj and jRk . There are C_n such relations on V (Exercice 187 of [9]). Since the dissimilarities defined by $i \neq j \implies d(i, j) = 1$ and by $i \neq j \implies d(i, j) = 2$ are equivalent, the result follows. \square

If we represent a Robinson dissimilarity d by a matrix with a horizontal diagonal (see Figure 3), we get “stacks” of lattice paths (like the one in Figure 2) with a “peak” in foreground. If d takes its values in $\{0, \dots, k\}$, the stack is made of k lattice paths (including the peak in foreground); equivalently, \mathcal{C}_d does not contain e_1, e_2, \dots, e_k with $\forall 1 \leq i \leq k, 1 < |e_i| < n$ and $e_1 \subset e_2 \subset \dots \subset e_k$. Table 1 shows the number of

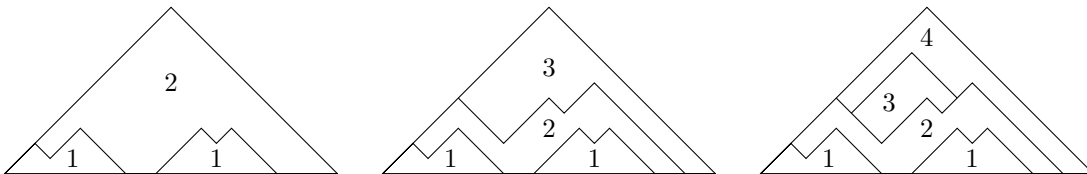


Figure 3: Robinson dissimilarities represented by matrices with a horizontal diagonal.

Robinson dissimilarities on $\{1, \dots, n\}$ taking their values in $\{0, \dots, k\}$ for small values of k and n . Sequences for $k = 3, 4, 5$ match no sequence in OEIS [6] and we know no formula for them.

Table 1: Number of Robinson dissimilarities on $\{1, \dots, n\}$ that take their values in $\{1, \dots, k\}$ for small values of n and $k \in \{2, 3, 4, 5, \infty\}$. The boxes with // needed too much computation time to be filled.

$n \setminus k$	2	3	4	5	∞
2	1	1	1	1	1
3	4	4	4	4	4
4	13	32	32	32	32
5	41	266	512	512	512
6	131	2254	9938	16384	16384
7	428	19934	208717	708314	1048576
8	1429	185383	//	//	134217728
9	4861	//	//	//	34359738368
n	$C_n - 1$?	?	?	$2^{n(n-1)/2-1}$

4 Conclusion

Our original goal was to prove that nearly all Robinson dissimilarities have a flat PQ-tree (a direct consequence of Claim 2 and of the proof of Theorem 1), which was intended to be a first step toward the determination of the average complexity of algorithms like those of [2] or [5]. This led us to lattice paths, and thus to a combinatorial interpretation of Catalan numbers (Theorem 2). This interpretation, although very simple, natural and easy to prove, seems curiously to have been unknown until now.

References

- [1] K.S. BOOTH & G.S. LUEKER (1976), Testing for the Consecutive Ones Property, Interval Graphs and Graph Planarity Using PQ-Tree Algorithm, *Journal of Computer and System Sciences* 13, 335–379.
- [2] V. CHEPOI & B. FICHET (1997), Recognition of Robinsonian Dissimilarities, *Journal of Classification* 14, 311–325.
- [3] E. DIDAY (1986), Orders and Overlapping Clusters by Pyramids, in *Multidimensionnal Data Analysis*, J. de Leeuw, W. Heiser, J. Meulman and F. Critchley Eds., pp 201–234. DSWO.
- [4] C. DURAND & B. FICHET (1988), One-to-One Correspondences in Pyramidal Representation: a Unified Approach, in *Classification and Related Methods of Data Analysis*, H.H. Bock Ed., pp 85–90. North-Holland.
- [5] M. LAURENT & M. SEMINAROTI (2017), Similarity-First Search: a New Algorithm with Application to Robinsonian Matrix Recognition, *SIAM J. Discrete Math.* 31-3, 1765-1800.
- [6] OEIS: The On-Line Encyclopedia of Integer Sequences, <https://oeis.org>.
- [7] P. PRÉA & D. FORTIN (2014). An Optimal Algorithm to Recognize Robinsonian Dissimilarities. *Journal of Classification* 31-3, 351–385.
- [8] W.S. ROBINSON (1951), A Method for Chronologically Ordering Archeological Deposits, *American Antiquity* 16, 293–301.
- [9] R.P. STANLEY (2015), *Catalan Numbers*. Cambridge University Press.