Compact Representation of Multi-Granular Topological Hierarchies

José Fuentes Sepúlveda, **Diego Gatica**, Gonzalo Navarro, Andrea Rodríguez, Diego Seco

March 24th, 2021

Introduction

Problem

- Nowadays, it is common to deal with the management of information associated with a specific space.
- Interesting questions to answer are of topological nature.

Region	Date	Total
Valparaiso	September	5080
Santiago	September	8054
Biobio	September	3204

Table: Number of birth per region

Province	Week	Sex	Total
Quillota	18	Male	102
Arauco	18	Female	52
Concepcion	18	Male	67

Table: Number of birth per province

Introduction

Problem with current solutions

- Solutions focused on models using spatial object geometry
- High space usage and slow query times
- Lack of research focused on topological models

Introduction

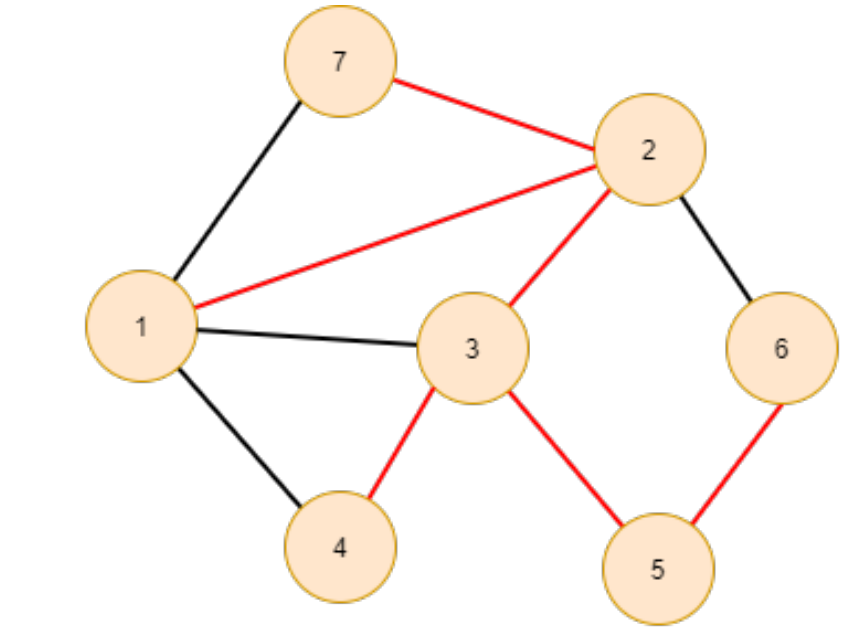
Proposed solution

- To develop a compact data structure with the objective of answering containment and adjacency queries.
- Proposed structure focused on answering topological queries over a partition of space.

Related work

Compact planar graph representation

- Turan's representation: Representation built in two steps.
- A sequence of parentheses and brackets representing the planar graph is obtained.

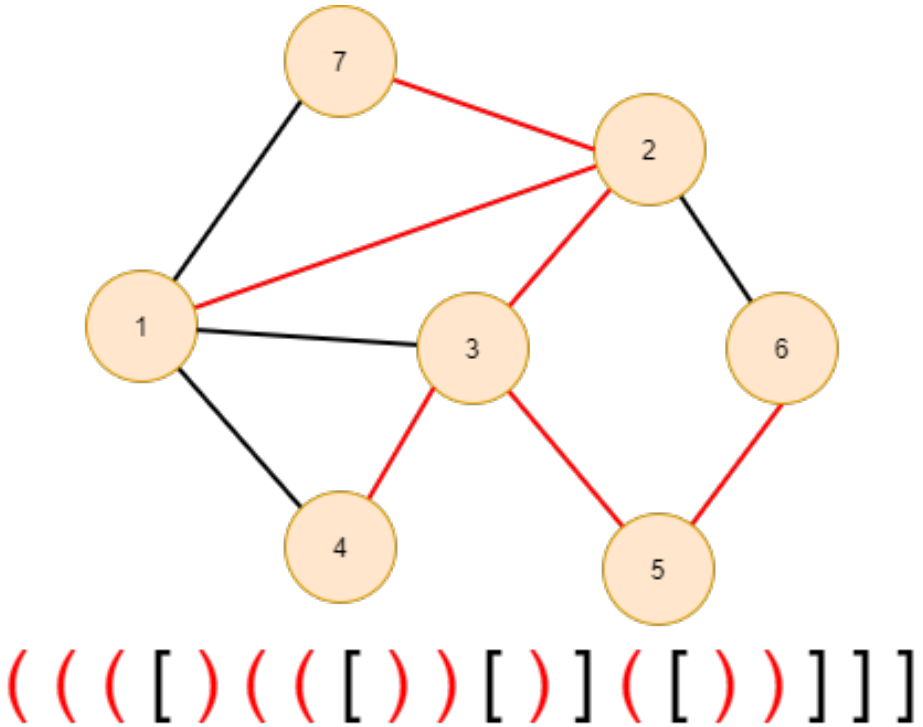


Example of compact planar graph

Related work

Originally, this representation has no primitives to navigate the graph.

- Ferres et. al. provides primitives for navigating the graph and answering topological queries
- Fuentes-Sepúlveda et al. improved the bounds of the operations.

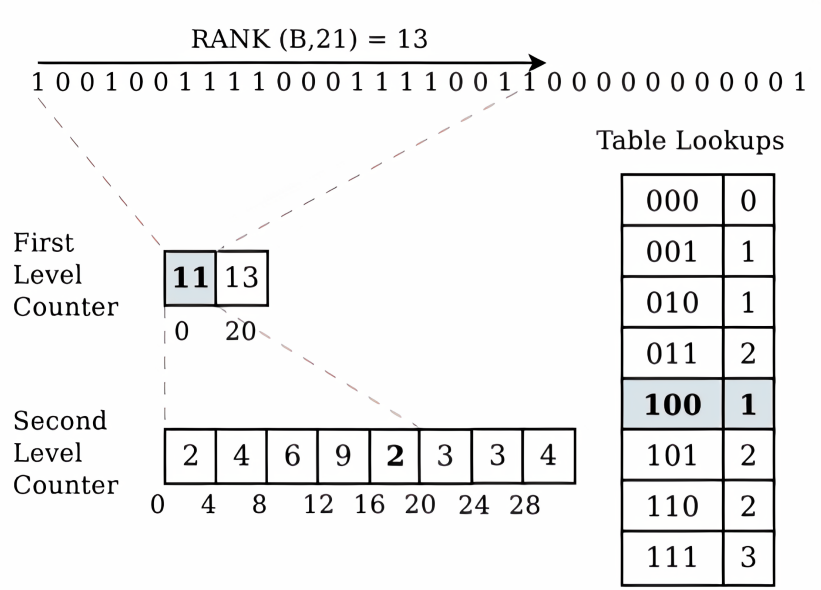


Example of compact planar graph

Related work

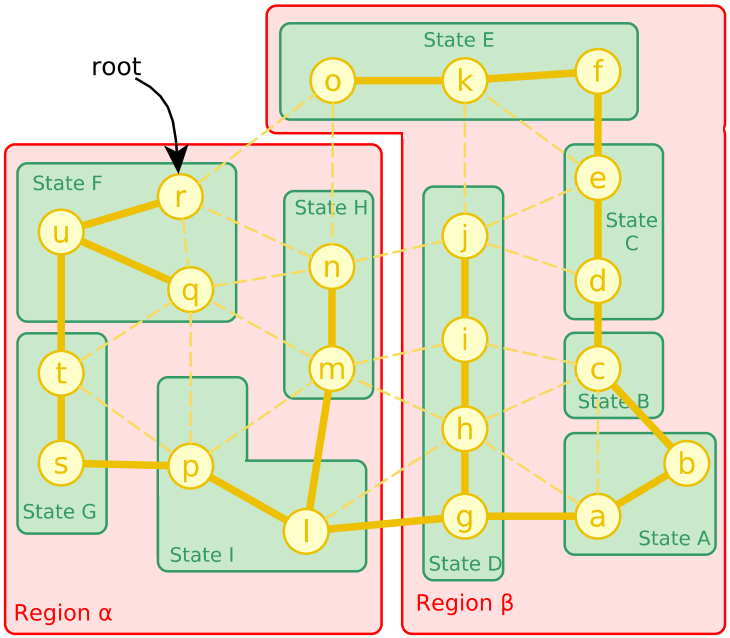
Succinct bitmap: Data structure that stores a secuenque of 0's and 1's

- $rank_b(B, i)$ returns the number of bits set to b in B between the positions 0 and i (both included).
- $select_b(B, i)$, which is the complementary operation of rank, returns the position in which the i-th bit of kind b is located in B.

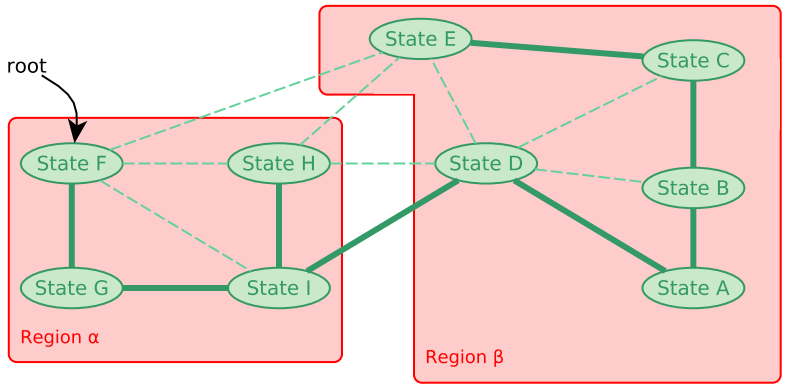


Example of a Bitmap

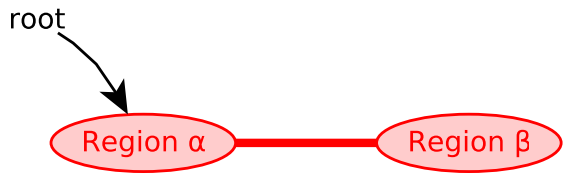
Proposed structure



County level planar graph



State level planar graph

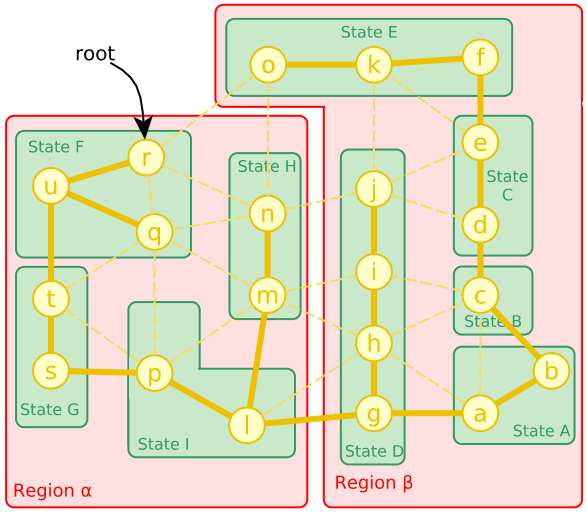


Region level planar graph

Proposed data structure

Compaction realized using the following components

- Compaction of each level using the same strategy of Fuentes-Sepúlveda *et al.*
- Bitmap B marks the entry and exit point to each region in the aggregation levels different from the highest level of aggregation.

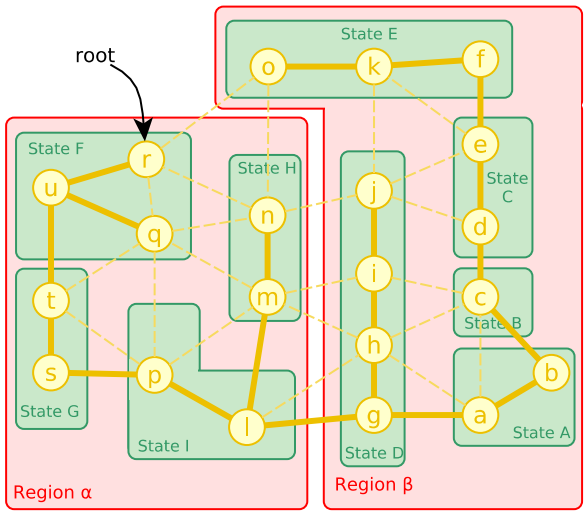


$$B_2 = 10101011011010000101101 0...$$

Compact representation

Change in the way of constructing the compact representation.

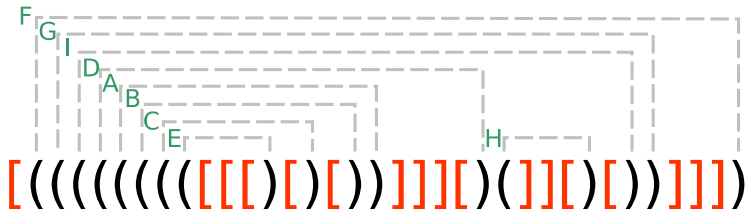
- DFS traversal performed at the highest granularity level L_h.
- Edge (u, v) will be traversed if both vertex v and the regions containing v at different levels have not been visited, or if v has not been visited and both u and v belong to the same region at a lower level of granularity.



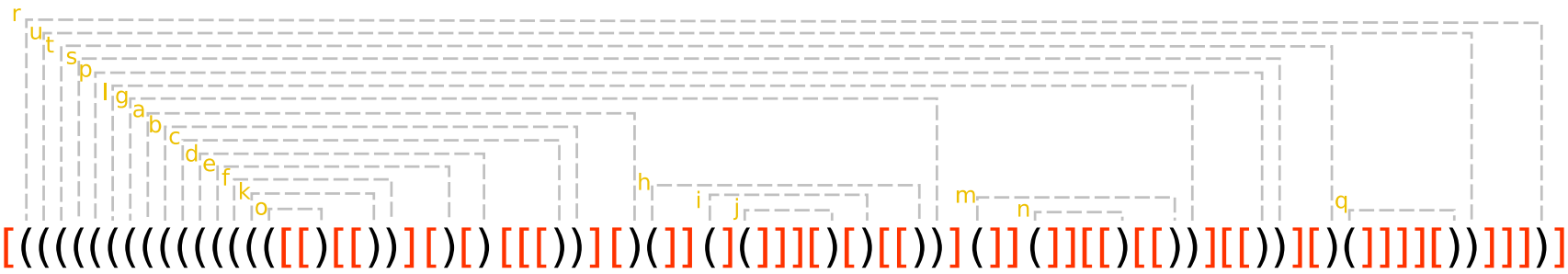
$$B_2 = 10101011011010001011010...$$

Operators

- go_up_$L_h(c, i)$: Find the first region at level L_h that is contained by c. First we obtain $z = \text{select}_(S_i, c)$, then we compute $y = \text{select}_1(B_i, \text{rank}_()(S_i, z))$, finally we report $\text{select}_()(S_h, y)$.



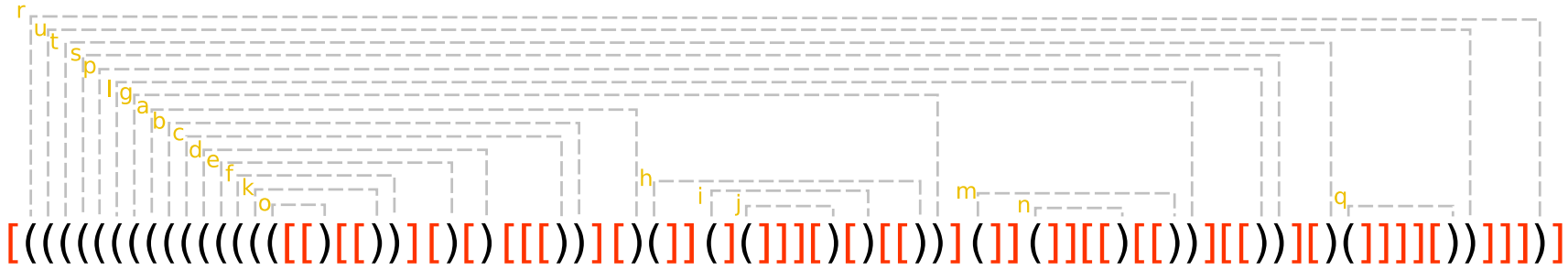
Compact representation state level



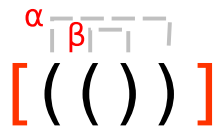
Compact representation county level

Operators

- go_down_$L_h(c, d)$: Find the region at level d that contains c. First we compute $p = \text{rank}_{()}(S_h, \text{select}_{(}(S_h, c))$, then we obtain $q = \text{select}_{()}(S_{h-d}, rank_1(B_{h-d}, p))$.



Compact representation county level



Compact representation region level

Operators

3 operations supported

- **Contains operator**: Does region r_1 contains region r_2?. Operation supported in $O(1)$.
- **Share boundary operator**: Does region r_1 shares an edge with region r_2?. Operation supported in $O(d_{r_1})$.
- **Contained operator**: List all regions at level L_j contained in region r_1. Operation supported in $O(n_j)$.

Dataset

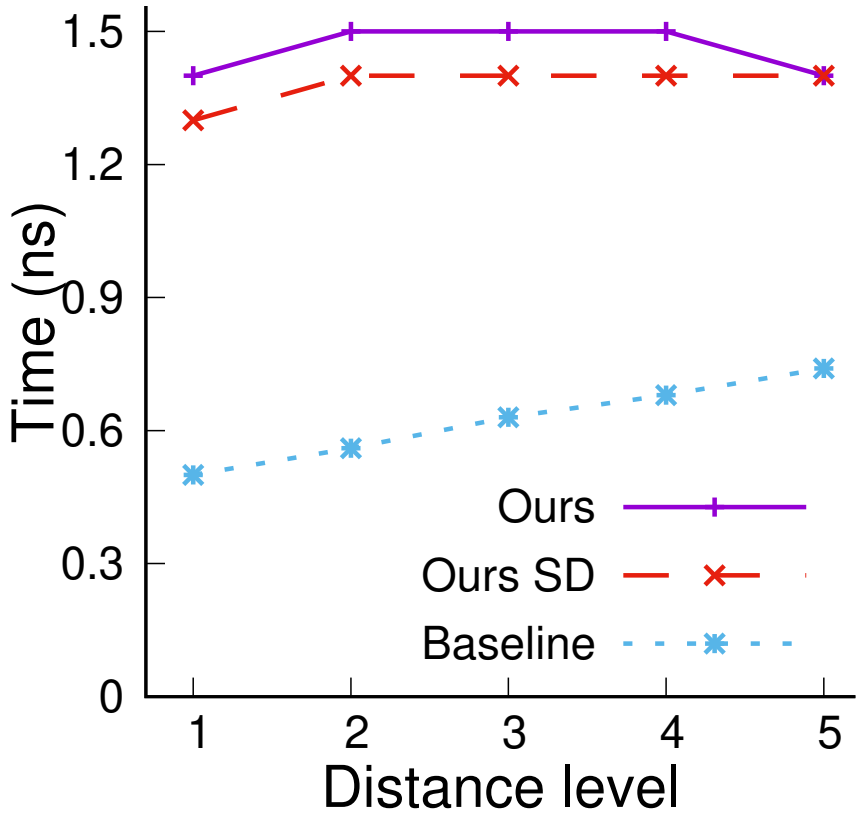
Dataset	Level	Vertex (n)	Edges (m)
`tiger_usa`	L_1: States	50	140
	L_2: Counties	3,110	9,095
	L_3: Census tracts	72,512	201,631
	L_4: Census block groups	216,243	597,784
	L_5: Census blocks	11,004,160	26,732,935
	L_6: Faces	19,735,874	43,837,150
`tiger_8s`	L_1: States	9	20
	L_2: Counties	595	1,730
	L_3: Census tracts	11,626	31,412
	L_4: Census block groups	33,804	91,891
	L_5: Census blocks	2,233,031	5,429,483
	L_6: Faces	4,761,354	10,326,904
`tiger_az`	L_1: States	2	1
	L_2: Counties	16	42
	L_3: Census tracts	1,572	3,999
	L_4: Census block groups	4,179	11,071
	L_5: Census blocks	241,651	598,898
	L_6: Faces	498,645	1,165,717

Dataset

Dataset	Structure	Embedding	Hierarchy	Total
`tiger_usa`	Baseline	50.94	259.68	310.02
	Ours	50.94	37.54	88.48
	Ours SD	50.94	6.96	57.90
`tiger_8s`	Baseline	11.45	55.51	66.96
	Ours	11.45	12.21	23.66
	Ours SD	11.45	1.63	13.08
`tiger_az`	Baseline	1.32	5.51	6.83
	Ours	1.32	0.91	2.23
	Ours SD	1.32	0.19	1.51

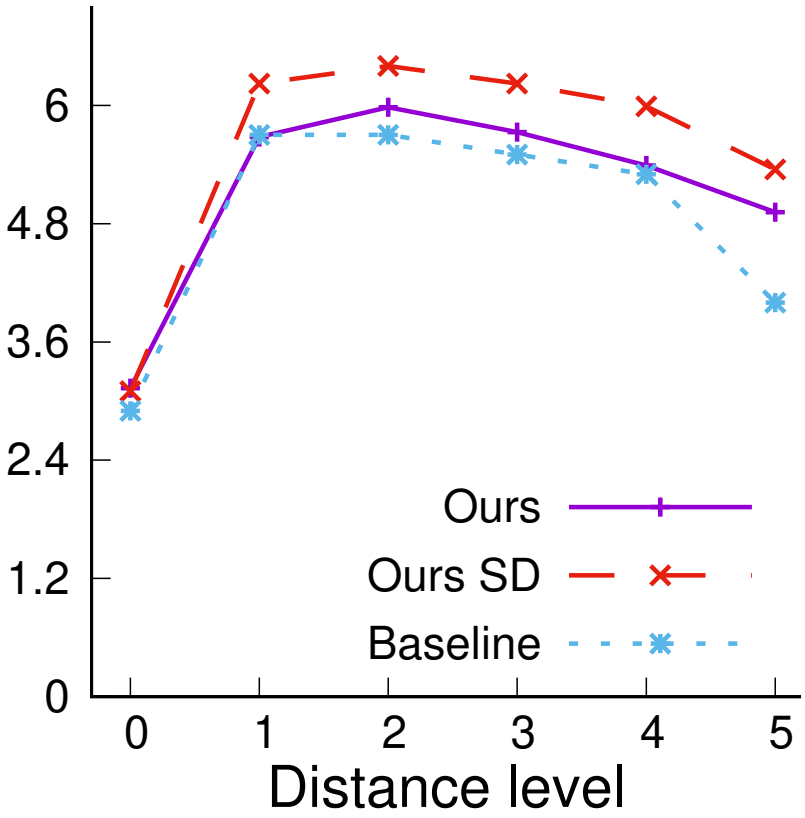
Space occupied by each implementation in MB

Experimental evaluation



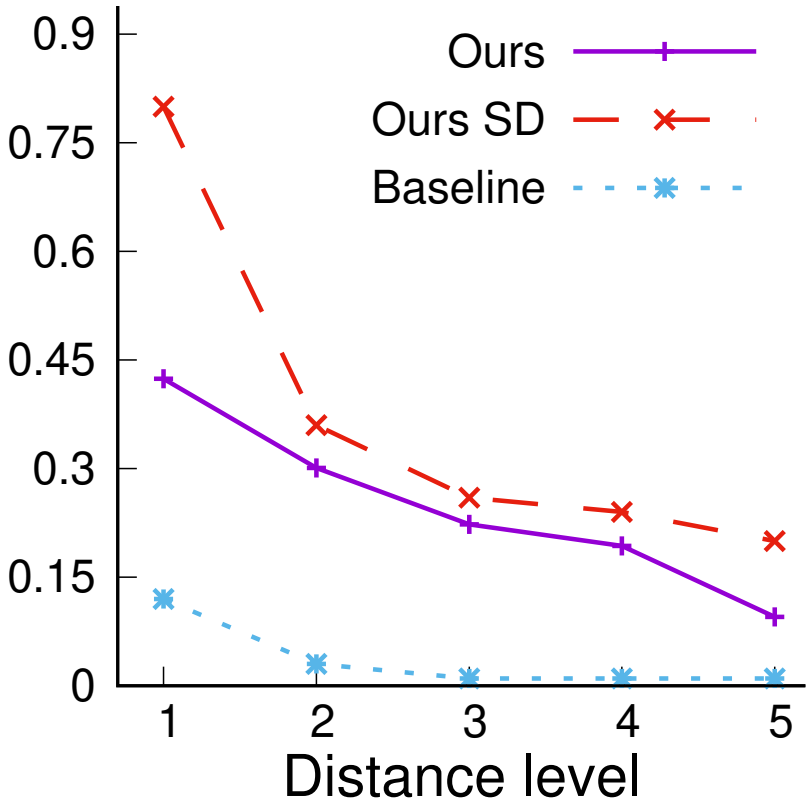
Contains operator

Experimental evaluation



Shareboundary operator

Experimental evaluation



Contained operator

Conclusions and future work

- The proposed structure uses only $4\sum_{i=1}^{h} m_i + 2n_h(h-1) + o(hn_h)$ bits.
- The use of compressed bitmaps does not drastically change the performance of the developed algorithms.
- In practice, the proposed hierarchy representation uses about 20% of the space needed by the baseline, and less than 5% when using compressed bitmaps.
- Our model assumes that all regions composing a region at a lower level of detail are contiguous.

Compact Representation of Multi-Granular Topological Hierarchies

José Fuentes Sepúlveda, **Diego Gatica**, Gonzalo Navarro, Andrea Rodríguez, Diego Seco

March 24th, 2021