

Teseo and the Analysis of Structural Dynamic Graphs

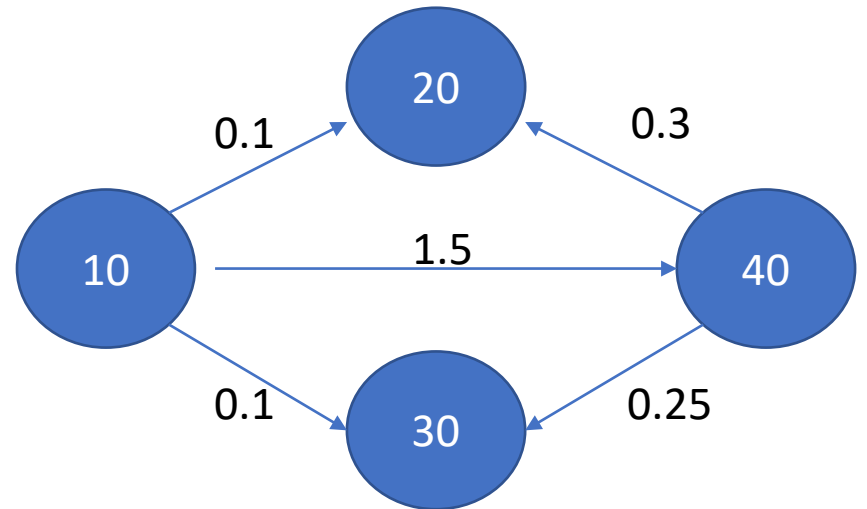
Dean De Leo

Amsterdam

18/06/2021

Context

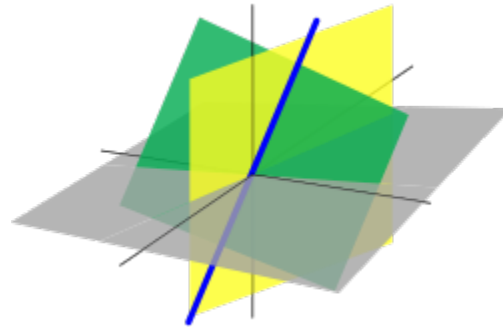
- **Graph analysis:** processing each vertex over multiple iterations, until a certain condition is satisfied (e.g. PageRank, SSSP, community detection, triangle counting).
- **Structural graphs:** homogenous graphs, potentially with weights, but no arbitrary properties.
- **Dynamic graphs:** they can change nodes, edges and weights.



Paradigms



Algorithmic approach



Linear algebra

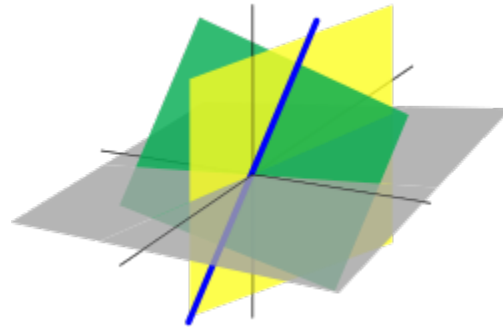


"Think like a vertex"

Paradigms



Algorithmic approach

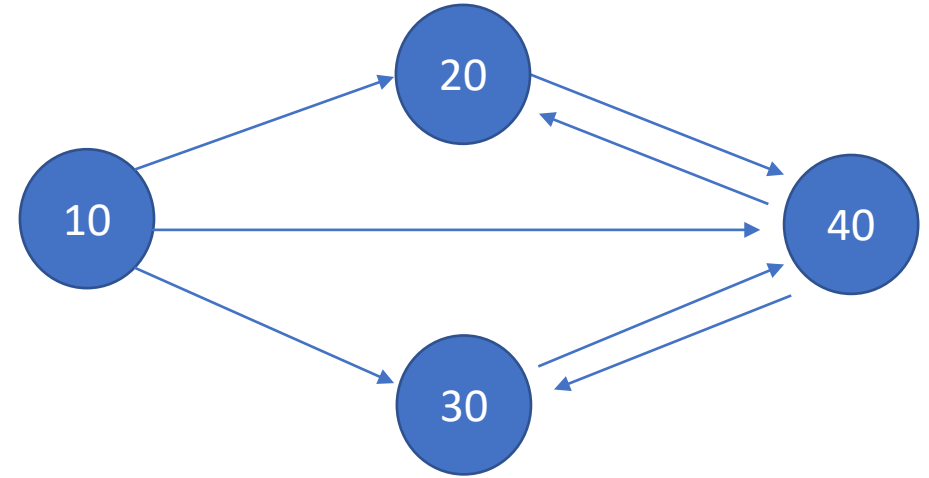


Linear algebra



"Think like a vertex"

Vertex table



30	
10	
20	
40	

Vertex identifiers
(hash table)

40	
20	
10	
30	

Vertex table
(array)

3	1	
0		
1	0	3
0		

Edge lists
(vectors)

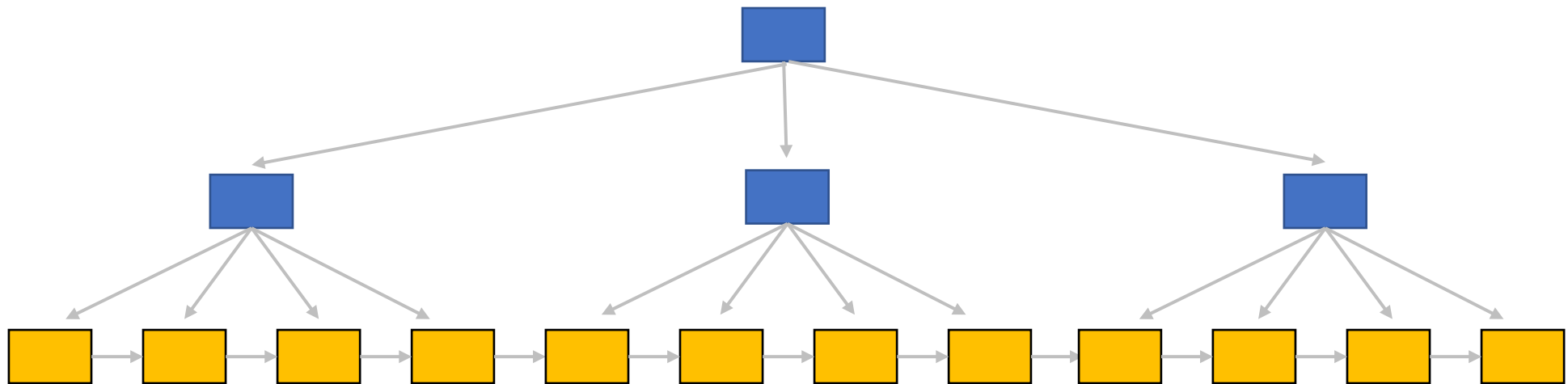
Opportunities

```
while (condition) {  
  ... // preamble  
  
  for all v in V // for all vertices  
    for all e in edges(v){  
      Pnext[i] = f(e, P[i], ...);  
    }  
  end for  
end for  
  
  ... // epilogue  
}
```

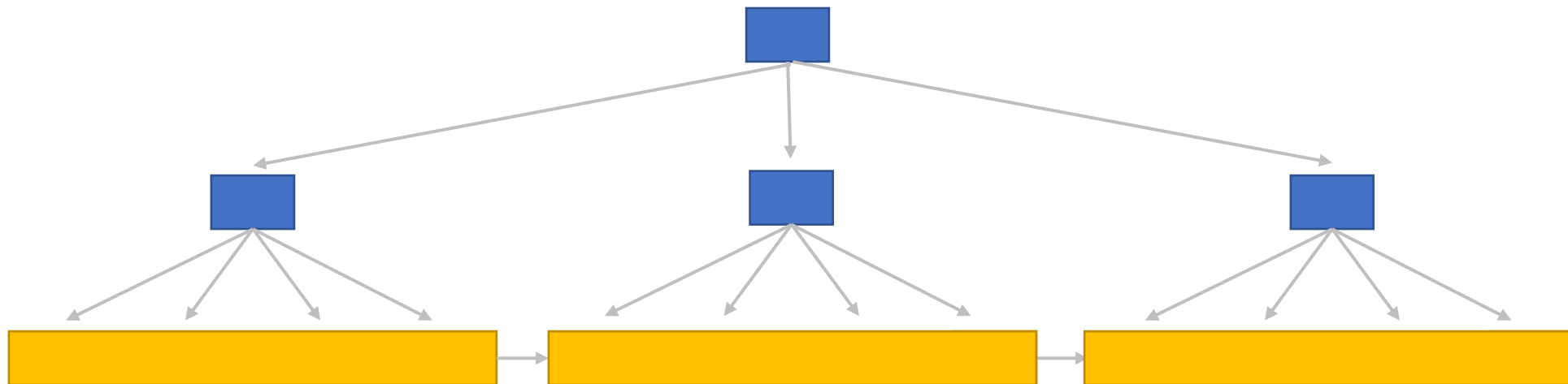
- **Sequential pattern**
- **Random pattern**

Teseo

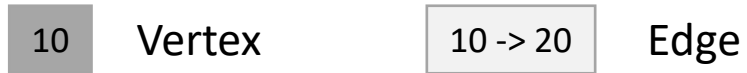
B⁺ trees



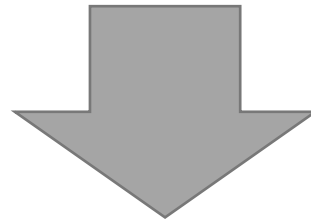
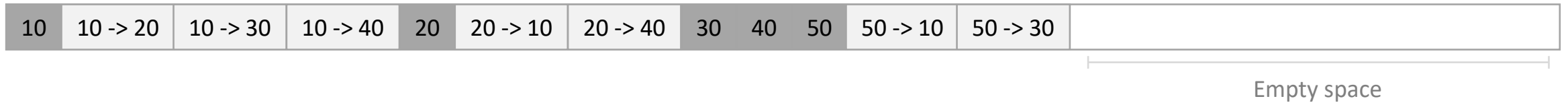
Fat tree



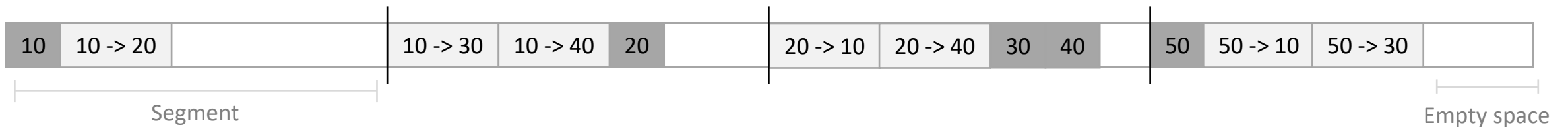
Leaf layout



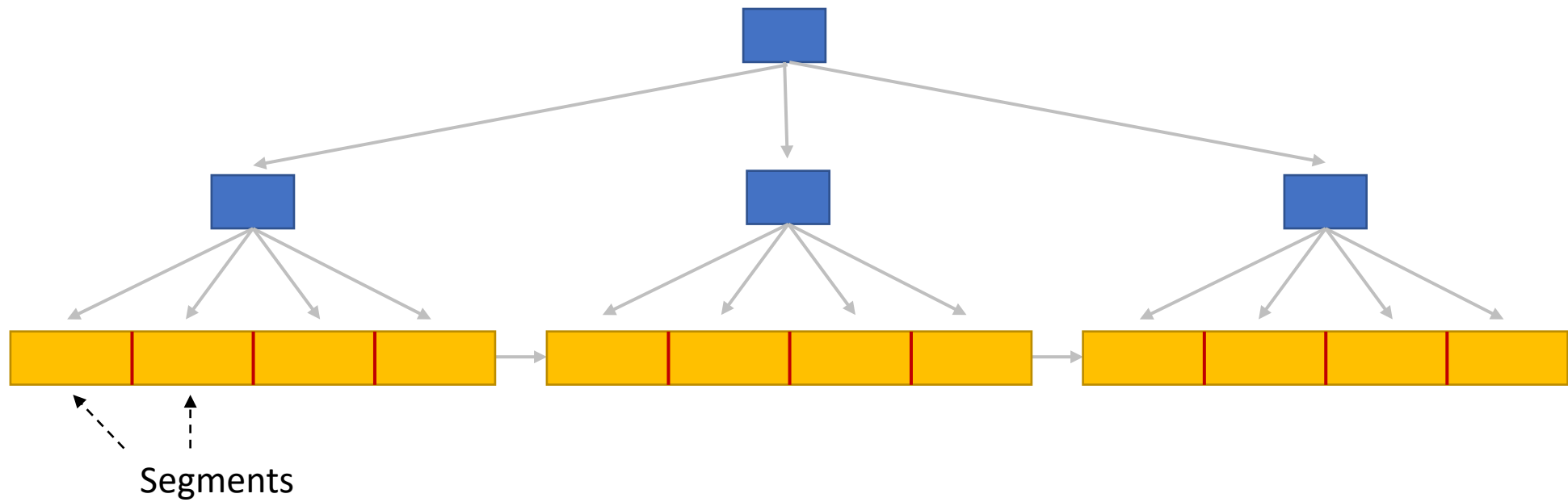
Dense array:



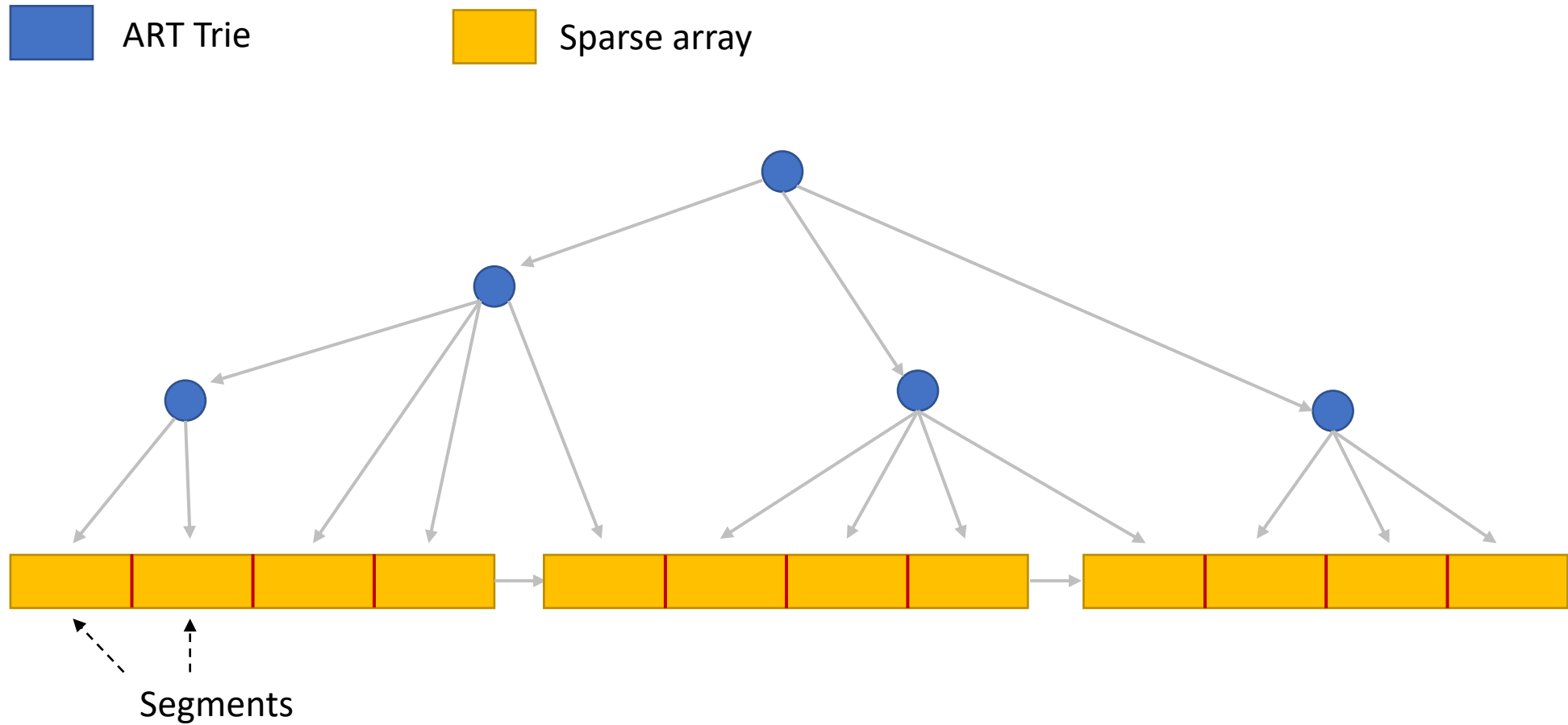
Sparse array or Packed Memory Array (PMA):



Fat tree



Fat tree



Fat tree

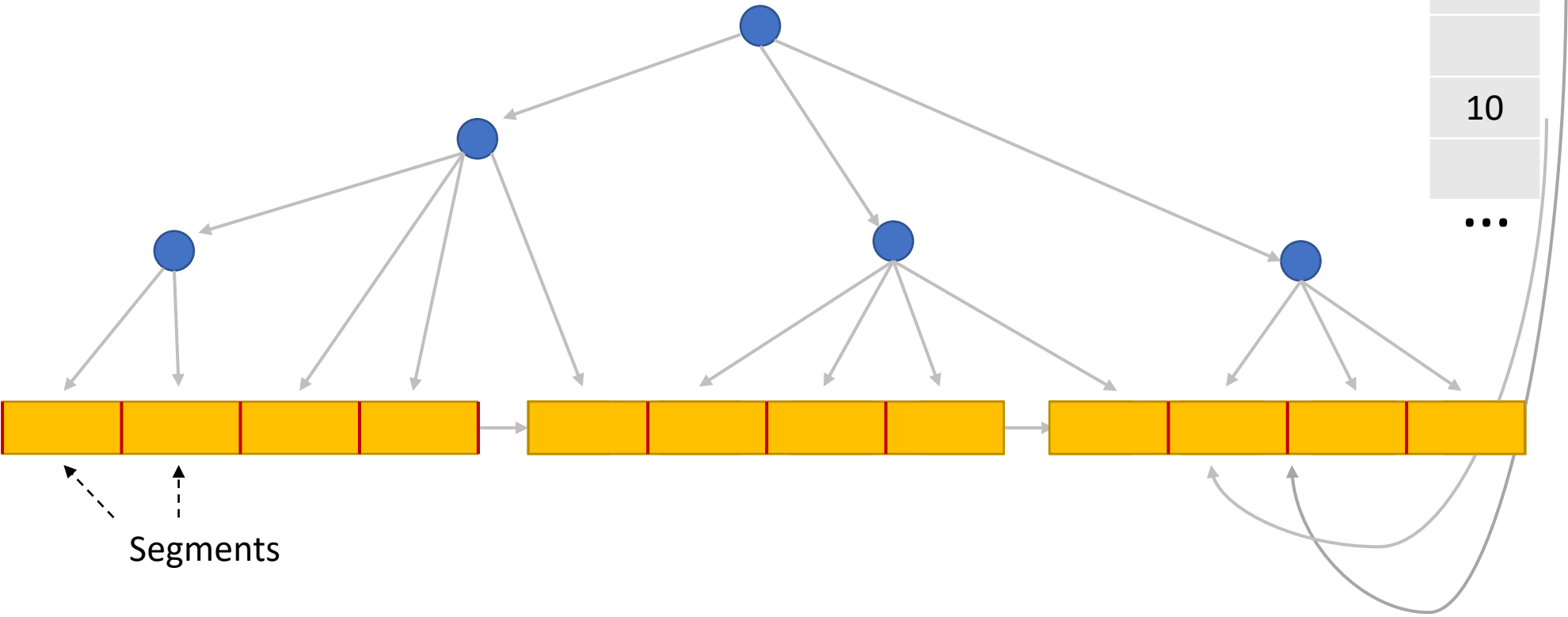


ART Trie



Sparse array

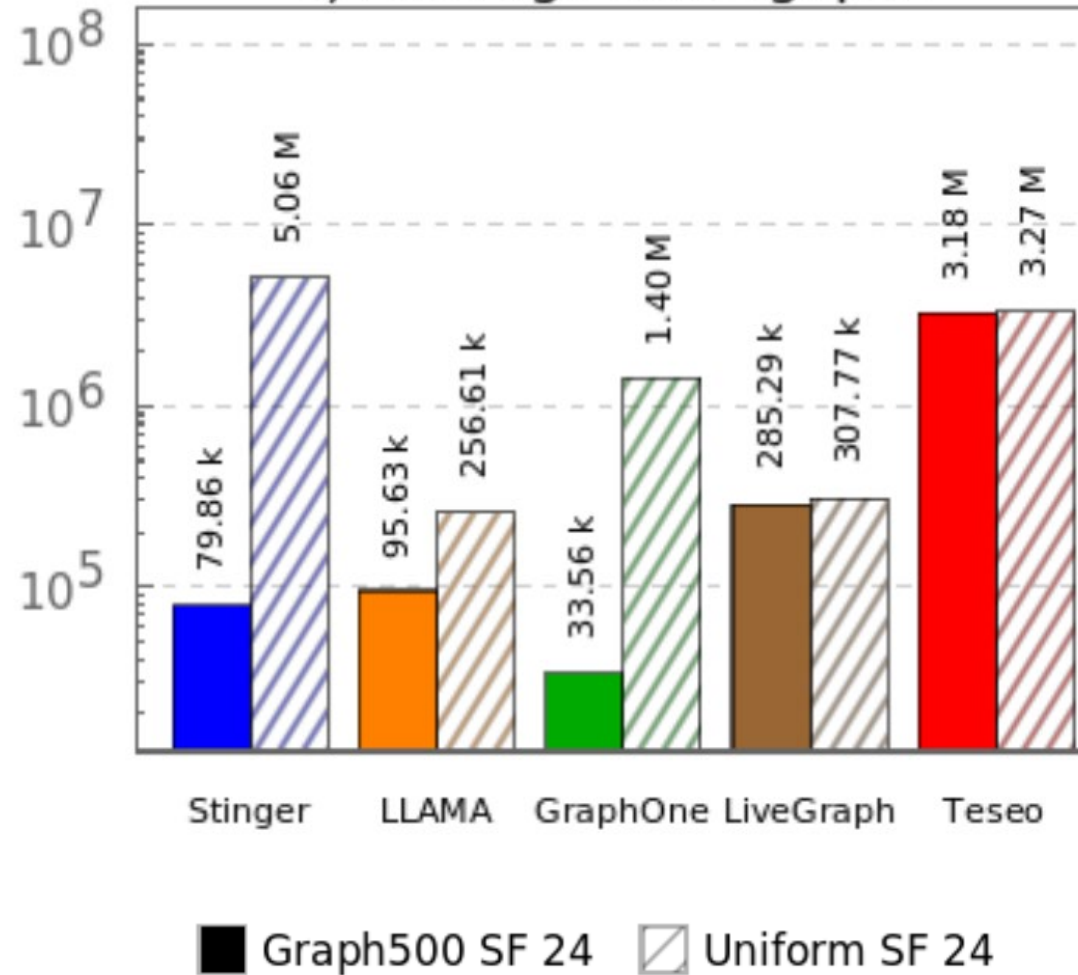
Hash table



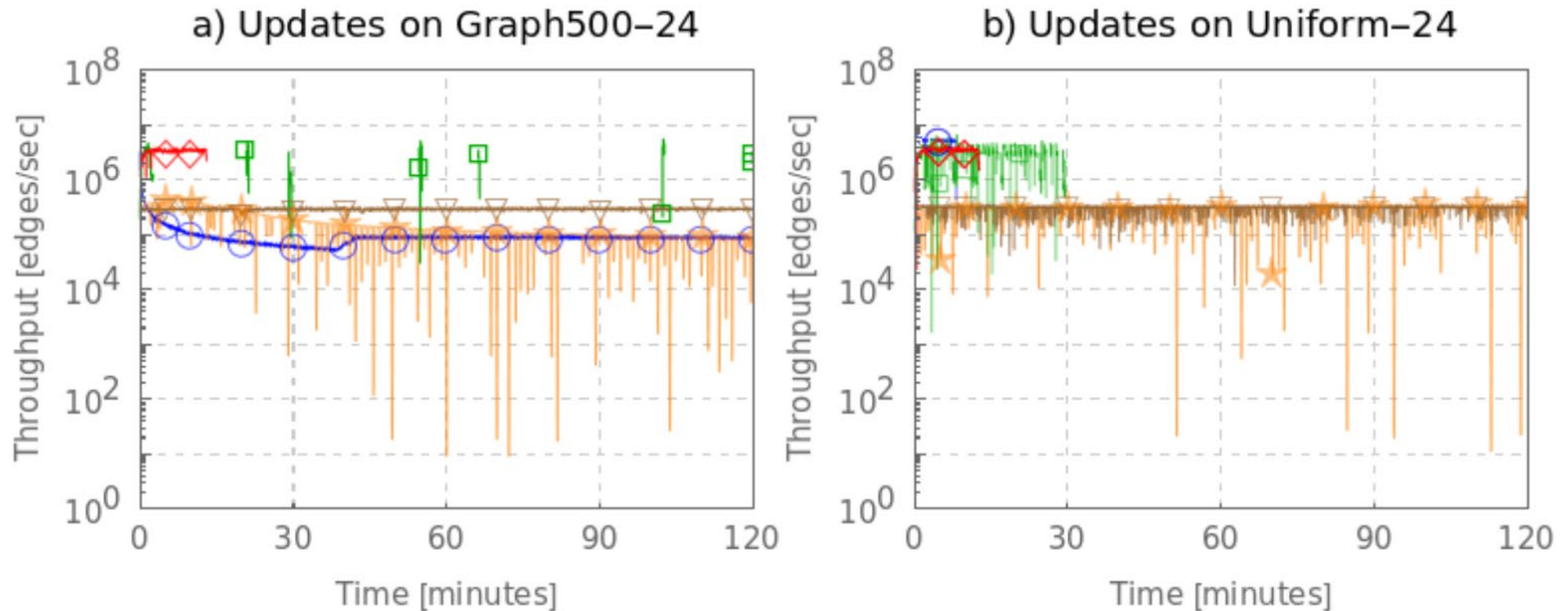
Experiments

Updates

c) Average throughput



Updates



○ ■ Stinger ★ ■ LLAMA □ ■ GraphOne ▽ ■ LiveGraph ◇ ■ Teseo

Graph analysis

- Evaluation on LDBC Graphalytics
- Same algorithm implementation for all systems

	System	BFS	CDLP	LCC	LCC (opt)	PageRank	SSSP	WCC
Graph500 SF 26	CSR (baseline)	5.03s	110s	DNF	704s	19.20s	36s	9.66s
	Stinger	1.18x	1.15x	DNF	N/A	2.33x	1.97x	2.08x
	GraphOne	2.64x	1.73x	DNF	N/A	2.55x	2.84x	36.61x
	LiveGraph	2.76x	1.55x	DNF	N/A	2.73x	2.32x	1.32x
	Teseo, log							
	Teseo, real							

Errata for the paper pending review.

Graph analysis

- Evaluation on LDBC Graphalytics
- Same algorithm implementation for all systems

	System	BFS	CDLP	LCC	LCC (opt)	PageRank	SSSP	WCC
Graph500 SF 26	CSR (baseline)	0.50s	108s	DNF	683s	15.05s	36s	5.19s
	Stinger	2.55x	1.13x	DNF	N/A	2.49x	1.63x	2.91x
	GraphOne	10.26x	1.66x	DNF	N/A	2.64x	2.22x	62.01x
	LiveGraph	21.72x	1.53x	DNF	N/A	3.14x	2.02x	1.89x
	Teseo, log. vtx	1.34x	1.71x	DNF	DNF	3.45x	2.97x	4.24x
	Teseo, real vtx	1.03x	1.00x	DNF	2.08x	1.62x	1.76x	1.49x

Conclusions

- Hybrid design between B⁺ trees and sparse arrays
- It favours sequential scans:
 - 50% slower than CSR in the sequential pattern
- In the random pattern, comparable performance w.r.t. other systems.