



Realizing Memory-Optimized Distributed Graph Processing

Panagiotis Liakos[†] – [Katia Papakonstantinou](#)[‡] – Alex Delis[†]

[†]University of Athens

[‡]Athens University of Economics and Business





Our contribution*

- ◇ **Space-efficient** representations of out-edges
- ◇ **Fast mining** of graph elements
- ◇ Execution in **memory-constrained settings**
- ◇ **Faster execution** due to improved memory management

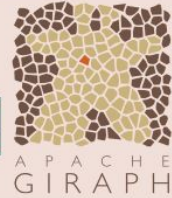
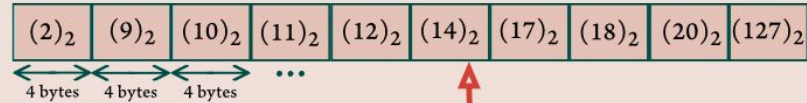
* To be included in **Apache Giraph**



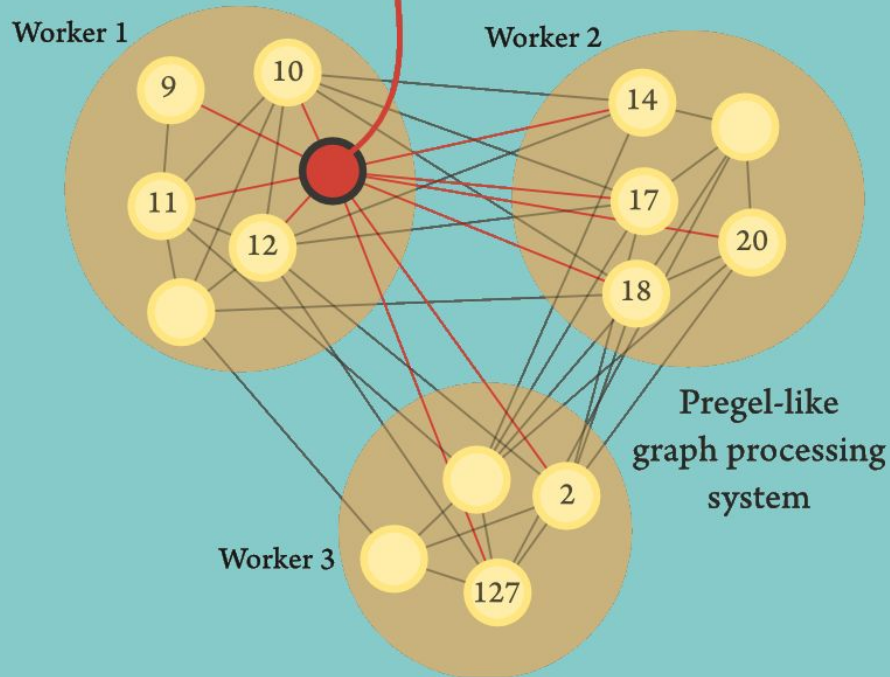
ByteArrayEdges

Apache Giraph's *default* out-edge representation:

A *byte-array* is used, in which target vertex ids are held *consecutively*.



- Ineffective memory usage
- Partitioning hardens the task of compression





Properties of real-world graphs

LOCALITY OF REFERENCE:

The majority of the edges of a graph link vertices that are close to each other in the order.

SIMILARITY (COPY PROPERTY):

Vertices that are close to each other in the order tend to have many common out-neighbors.

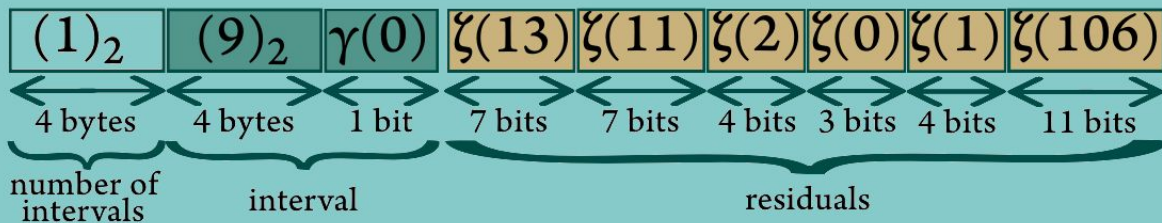
<u>label</u>	<u>URL</u>
n	http://www.aueb.gr/ http://www.aueb.gr/css/foostyle.css http://www.aueb.gr/images/logo.gif http://www.aueb.gr/about/ http://www.aueb.gr/staff/
n + 1	http://www.aueb.gr/about/ http://www.aueb.gr/css/foostyle.css http://www.aueb.gr/images/logo.gif http://www.aueb.gr/directions.html http://www.aueb.gr/about/ http://www.aueb.gr/staff/

BVEEdges (based on BV04)

2, 9, 10, 11, 12, 14, 17, 18, 20, 127



9 - 12, 2, 14, 17, 18, 20, 127



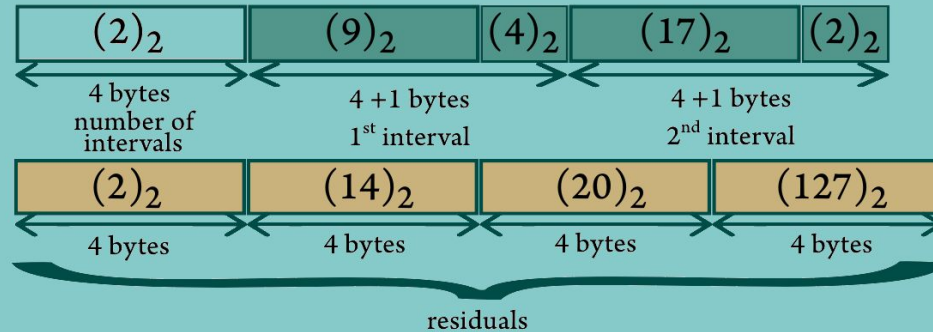
- bit encoding involves significant overhead

IntervalResidualEdges

2, 9, 10, 11, 12, 14, 17, 18, 20, 127



9 - 12, 17 - 18, 2, 14, 20, 127



- Avoids expensive encodings & bit streams
- Significant compression due to locality of reference



Yet some more representations

VariableByteWeightsEdges

- ◇ edge weights exhibit right-skewed distributions
- ◇ $\log_{128} n + 1$ bytes to represent integer n

RedBlackTreeEdges

- ◇ does not waste space for empty buckets
- ◇ cost-free iterations with regards to space





Experimental evaluation

- ◇ Space-efficiency
- ◇ Performance





Space efficiency

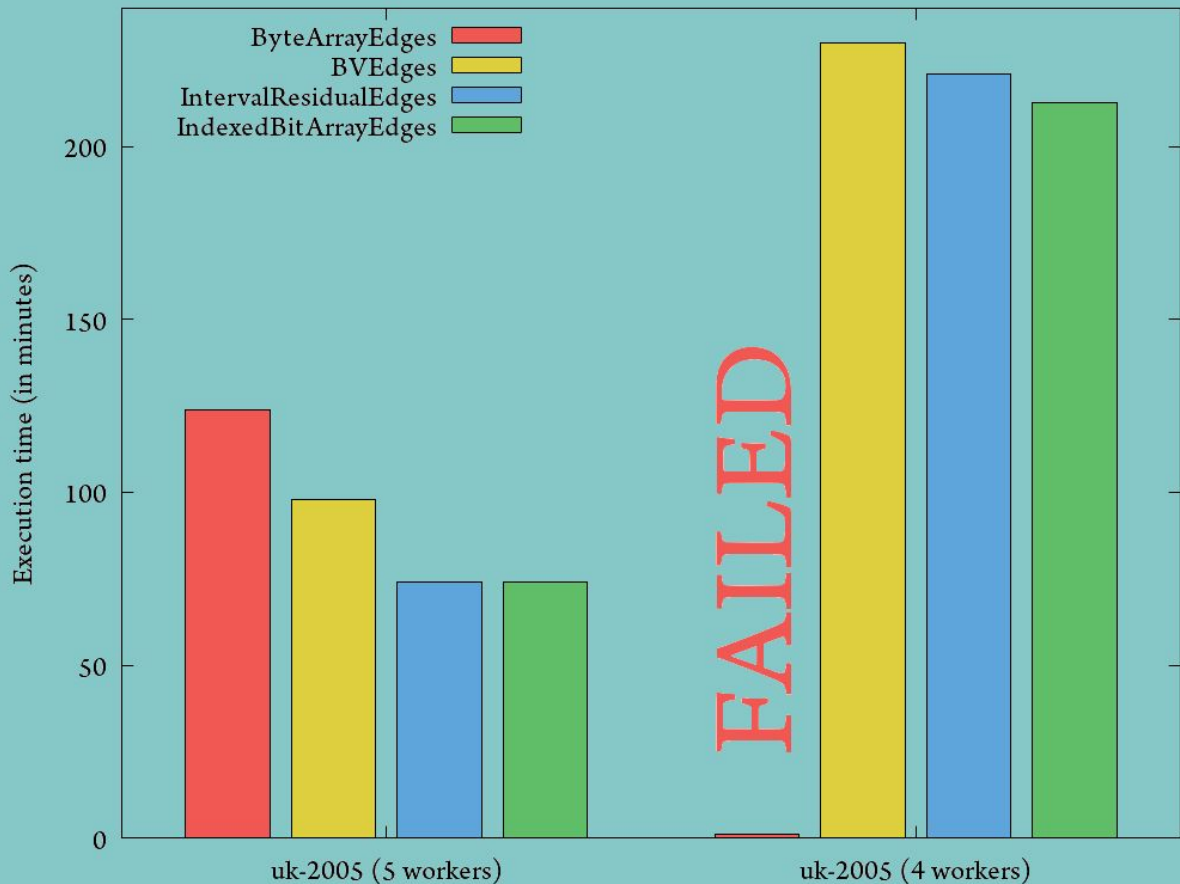
	Giraph	BV	IR	IBA
uk-2005	7.4GB	2.4GB	2.7GB	2.9GB
twitter-2010	11.2GB	4.6GB	7.1GB	8.9GB
sk-2005	14.8GB	4.9GB	5.6GB	6.4GB

- Impressive savings with all our techniques!
- 

Performance



- *IntervalResidualEdges* is faster than *BVEdges*
- *IndexedBitArrayEdges* outperforms all





Thanks!

ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ



ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS



MAD:IK

Management of Data, Information & Knowledge Group

Any questions?

You can find me at:

- ◇ e-mail: katia@aueb.gr
- ◇ Website: <http://www2.aueb.gr/users/katia/>
- ◇ Source code: <https://tinyurl.com/lpd-tkde-2018>

