

Link Discovery with Guaranteed Reduction Ratio in Affine Spaces with Minkowski Measures

Axel-Cyrille Ngonga Ngomo
ngonga@informatik.uni-leipzig.de

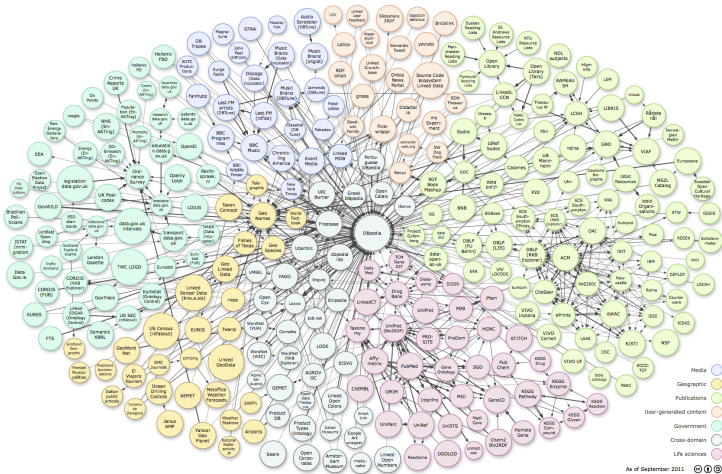
University of Leipzig
AKSW Research Group

November 13th, 2012
ISWC 2012, Boston MA, USA



Introduction

- 31+ billion triples
- Only ≈ 0.5 billion links



Question

- Can we devise lossless approaches with guaranteed RR?
- Advantages
 - Space management
 - Runtime prediction
 - Resource scheduling

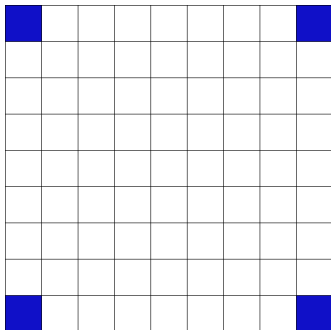
Yes! \mathcal{HR}^3

- Guaranteed reduction ratio
- Time-efficient
- No loss of recall
- $\lim_{\alpha \rightarrow \infty} RR(\mathcal{HR}^3(\alpha)) = 1$

Theorem

$$\forall \alpha > 1 \quad RR(\mathcal{HR}^3(2\alpha)) > RR(\mathcal{HR}^3(\alpha))$$

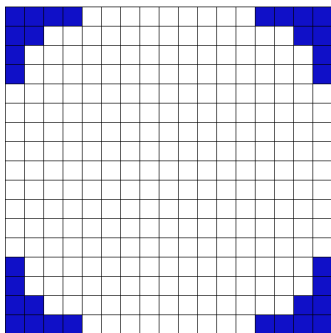
- $p = 2, \alpha = 4$



Theorem

$$\forall \alpha > 1 \quad RR(\mathcal{HR}^3(2\alpha)) > RR(\mathcal{HR}^3(\alpha))$$

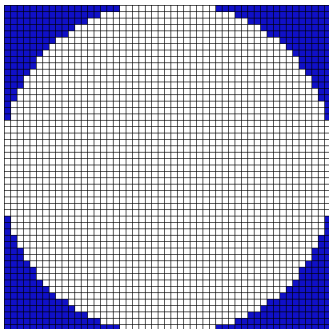
- $p = 2, \alpha = 8$



Theorem

$$\forall \alpha > 1 \quad RR(\mathcal{HR}^3(2\alpha)) > RR(\mathcal{HR}^3(\alpha))$$

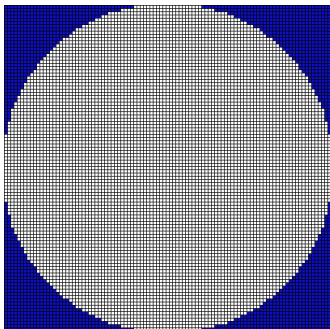
- $p = 2, \alpha = 25$



Theorem

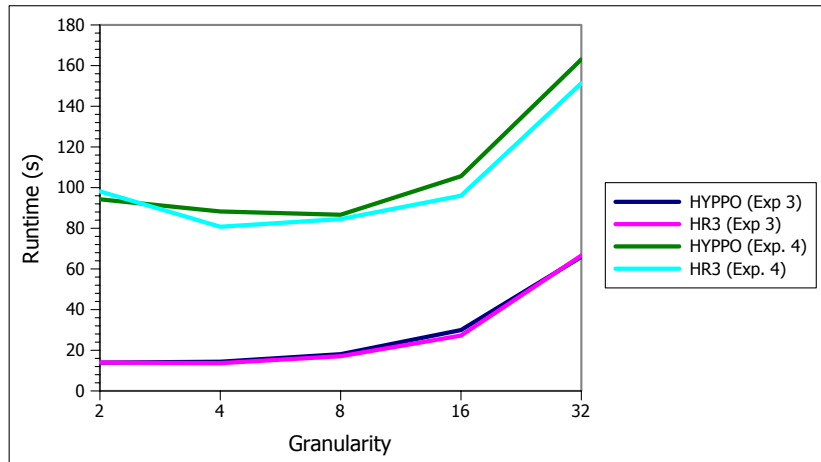
$$\forall \alpha > 1 \quad RR(\mathcal{HR}^3(2\alpha)) > RR(\mathcal{HR}^3(\alpha))$$

- $p = 2, \alpha = 50$



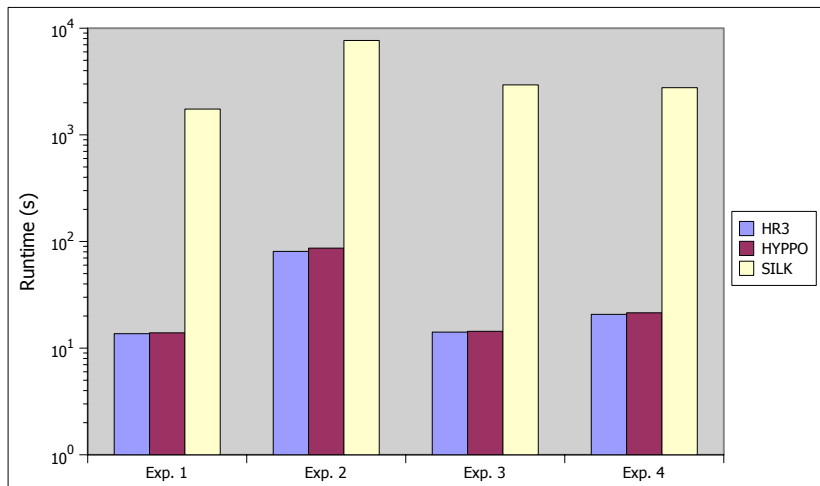
HR³: Experiments (Runtime)

- Experiment 3,4: Geonames and LGD, $\theta = 1, 9^\circ$



HR³: Experiments (Runtime)

- Experiment 1, 2: DBpedia, $\theta = 49,99m$
- Experiment 3, 4: Geonames and LGD, $\theta = 1,9^\circ$



Mission

- New category of algorithms for link discovery

Future Work

- Combine \mathcal{HR}^3 with multi-indexing approach
- Devise resource management approach
- Develop other algorithms (esp. for strings) with the same/similar theoretical guarantees

Thank You!



Axel Ngonga
Augustusplatz 10
D-04109 Leipzig
ngonga@informatik.uni-leipzig.de
<http://bis.uni-leipzig.de/AxelNgonga>
<http://limes.sf.net>