

# データ工学 (天笠 俊之)

## Data Engineering (AMAGASA Toshiyuki)



AMAGASA Toshiyuki, Ph.D.  
Professor  
Center for Computational Sciences /  
Center for Artificial Intelligence Research  
University of Tsukuba

E-mail address: amagasa@cs.tsukuba.ac.jp  
URL: <http://kde.cs.tsukuba.ac.jp/>



### 大規模データに対する処理の高速化および知識発見

コンピュータやネットワークの急速な進歩とともに、あらゆる分野でビッグデータが生成・収集されており、ビッグデータに対する効率的な処理や、ビッグデータからの有効な知識発見が強く求められています。我々はデータ工学/データベースに関連する技術を活用して、ビッグデータに対する処理の高速化や知識の発見に関する研究や、それらの技術を活用した科学分野のデータ活用に取り組んでいます。研究テーマ：GPU, FPGA, メニーコアプロセッサ等を活用したビッグデータ処理の高速化, 非負値行列分解 (NMF) 等のデータマイニング技術を利用したビッグデータからの知識発見, Linked Open Dataやグラフ等の半構造データに対する検索・知識発見, 生物学や天文学データ等の管理・知識発見。

### Efficient processing / knowledge extraction for big data

With the rapid development of computer and network technologies, various kinds of data are being collected and accumulated, which is called "big data," and efficient processing and knowledge extraction from such big data are highly demanded in wide spectrum of domains. We have been engaged in research/development on techniques of efficient processing and knowledge extraction for big data exploiting database and data engineering technologies. Research topics: efficient processing of big data using GPU/FPGA/many-core processors, data mining and knowledge extraction from big data exploiting techniques, such as NMF (non-negative matrix factorization), search and knowledge extraction from semi-structured data, such as Linked Open Data (LOD) and graphs, management and knowledge extraction for scientific domains, such as biology and astronomy.

### GPU Acceleration of Label Propagation

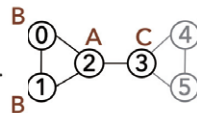
- Basic idea
  - Express the algorithm by only operations suitable for GPUs

- Data layout
  - The CSR format

ptr [ 0 | 2 | 4 | 7 | ... ]

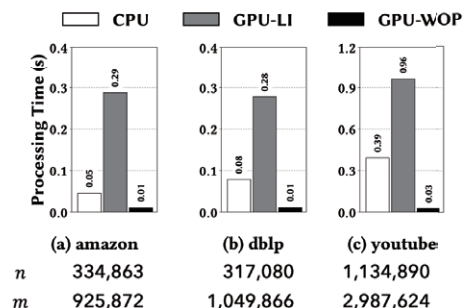
id [ 1 | 2 | 0 | 2 | 0 | 1 | 3 | ... ]

Label [ B | B | A | C | ... ]



### Experimental Evaluation

- Observations
  - GPU-WOP (without partitioning) is the fastest
  - Load balancing is important



[Yusuke Kozawa, Toshiyuki Amagasa, Hiroyuki Kitagawa:  
GPU-Accelerated Graph Clustering via Parallel Label Propagation, CIKM 2017, pp. 567-576, 2017]