

04

CPUを賢く使ってグラフから素早く知識を発見

～データ配置の最適化による高速なグラフ分析～

どんな研究

ビッグデータの一種であるネットワーク構造（グラフ）を表すデータの分析において、**CPU本来の性能を引き出すことでグラフ処理を高速化する研究**です。例えばSNS上のつながりから影響力が大きい人を発見したり、仲良しグループを発見したりといった多様な分析を高速化します。

どこが凄い

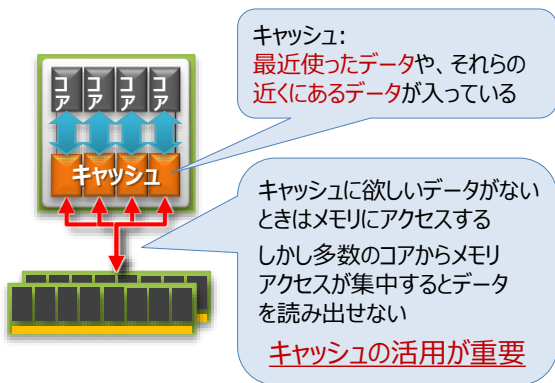
Webページのリンク構造や友人関係など「実世界のネットワーク」と呼ばれる構造の性質と、CPUが持つハードウェア的な性質を共に捉え利用することで、**従来のレイアウト手法を用いる場合と比べ最大2倍、レイアウト手法を用いない場合と比べ最大17倍**の高速化を実現します。

目指す未来

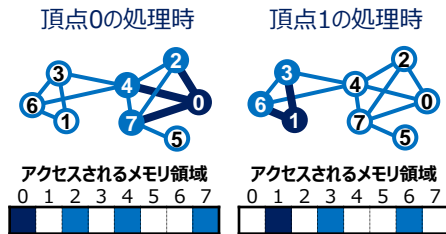
グラフ分析には「知り合いかも？」（友達検索）や商品・動画の推薦、個人の趣味・趣向を捉えたマーケティングなど幅広い応用があります。本技術は分析内容を問わず処理を高速化することで、より**大量のデータを用いた正確な知識発見や、瞬時に表示される推薦**などを可能にします。

メモリアクセスの効率化によってCPUの演算性能を引き出し、影響力分析やクラスタリングなど様々なグラフ分析処理を最大17倍高速化します

背景: CPUとメモリの構造



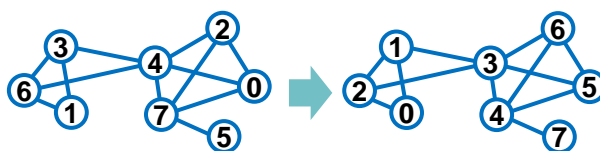
課題: グラフ分析のメモリアクセス



- アクセスされるデータの場所がバラバラ
- 同じデータの再利用が少ない

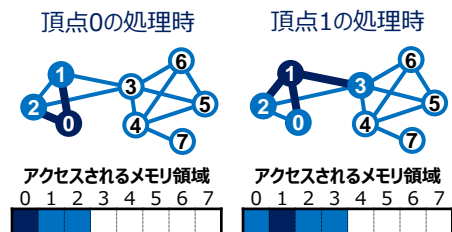
キャッシュを活用できない！

手法: 頂点順序の最適化



頂点の ID 番号を振りなおす

隣接頂点のデータがメモリ上で近くに配置されるよう並べ替える



データが密集 & データの再利用が多い

キャッシュを活用できる！

- 頂点順序の最適化により影響力・関連度分析 (Personalized PageRank) を最大12倍高速化 (既存手法比約2倍)
- さらに**グラフデータの圧縮**によるメモリとの通信量削減を組み合わせることで**最大17倍高速化**

関連文献

[1] 新井淳也, 塩川浩昭, 山室健, 鬼塚真, “頂点順序の最適化によるスケーラブルなグラフ並列処理,” in *Proc. the 6th Forum on Data Engineering and Information Management*, 2015.

連絡先

新井淳也 (Junya Arai) ソフトウェアイノベーションセンター
E-mail: arai.junya(at)lab.ntt.co.jp

