

ポストペタスケールシステムにおける超大規模グラフ最適化基盤

Advanced Computing and Optimization Infrastructure for Extremely Large-Scale Graphs on Post Peta-Scale Supercomputers

中央大学 & JST CREST 藤澤 克樹

今後は最適化問題の応用分野が非常に広範化（企業、社会、公共政策）すると共に、巨大なセンサーデータによる最適化問題の複雑&巨大化が予想される。問題サイズを n とした場合では、計算量に応じて $O(n^3) \rightarrow$ 数百万程度, $O(n \log n) \rightarrow 100$ 億以上, $O(n) \rightarrow 100$ 兆程度の問題規模を 2015 年頃に登場する予定のポストペタスパコンで扱うことができるようになって予想されている。我々が推進している JST CREST プロジェクト¹では、今後5年以内にスーパーコンピュータ等で以下の規模の問題に対するアルゴリズムとソフトウェアの開発を目標としている。

1. 超大規模ネットワークに対する最適化、探索及びクラスタリングアルゴリズム

- (a) 最短路 (中心性), 最大フロー, PageRank, グラフ分割, 中心性, その他
- (b) 数理計画問題 (SDP, MIP): 汎用ソルバの必要性 (個別の問題に対する仮定やチューニングは実用性が低い)
- (c) グラフクラスタリング, 高速グラフレイアウト

2. 数百万頂点～数兆頂点, 数億枝～数百兆枝からなる超大規模なグラフ解析 (図 1, 2)

- (a) 2^{42} 頂点 数 PB 以上のワーキングデータセット
- (b) グラフ解析: 例: 890 億個のニューロンとその接続 100 兆枝 ペタバイト級のストア領域: (2023 年:エクサスパコン)
- (c) 5 億頂点 250 億辺のグラフのインタラクティブな操作のための計算性能

グラフ解析に関しては数年以内に数千万規模の並列性を備えたポストペタスパコン上での高性能な超大規模グラフ処理技術が確立され、ストレージの階層性が深化したポストペタスパコン上での高性能な超大規模グラフ処理技術が達成されると予想されている。以下の分野に対してグラフ解析を適用することを想定している (図 2)。

- 1. 交通データに対する経路探索: 動的に変化する交通量等から高速な重要度判定を行うことによって、交通管制等に活用する。

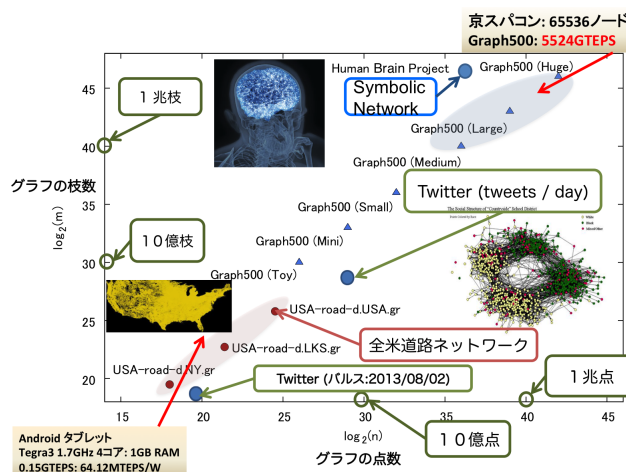


図 1: 今後想定されるグラフ解析の応用と問題規模

- 2. ソーシャルネットワークデータ (マイクロブログや SNS など) やウェブデータに対する動的な重要度, 影響度の判定. 各点の周辺, 及び広域内における影響 (情報の伝播力) を推定する.
- 3. その他: 疫病の拡散, 人口の増減, 経済動向等の分析. ライフライン等の基盤計画 (電力, 水, 食料). 生命科学系 (創薬, 遺伝子). ビジネス系 (金融, データマイニング). 安全保障分野 (組織構成の解明, 事件事故の事前予測).

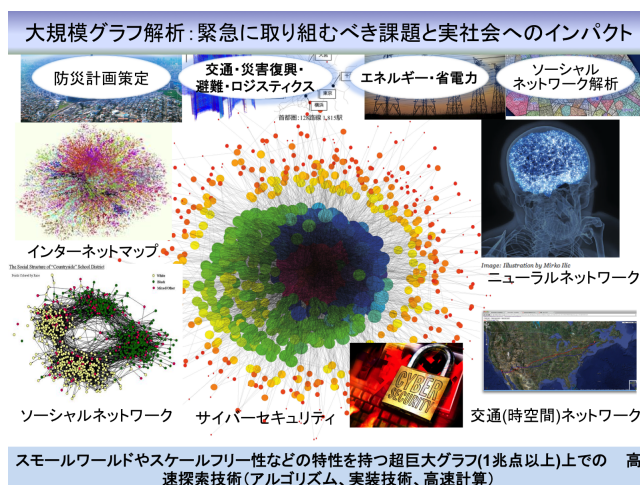


図 2: 大規模グラフ解析とその応用

¹<http://www.graphcrest.jp/jp/>