

論文の内容の要旨

論文題目 Pruned Labeling Algorithms: Unified Indexing Scheme for Graph Query Processing
(枝刈りラベリング法による大規模グラフ上の体系的なクエリ処理)

氏名 秋葉 拓哉

物事の関係が現れるほぼあらゆる場面で、データはグラフとして表現され処理される。特に近年では、インターネット及びワールド・ワイド・ウェブの普及に伴い、ソーシャルネットワークやウェブグラフを始めとする非常に大規模なグラフデータが偏在している。そのため、大規模グラフデータから有用な情報を効率的に引き出すことは現代社会の様々な場面において重要な役割を担っている。そのような大規模グラフの処理の根幹を支える重要な部品の 1 つが、最短経路及び関連問題に対する索引付け手法である。それらの手法は、グラフから予め索引と呼ばれるデータ構造を前計算し、そのデータ構造を用いて 2 点間の最短経路などの問合せに効率的に応答する。スケーラビリティ（索引サイズや索引構築時間）と応答性能（応答時間や精度）の良好なトレードオフを達成することが索引付け手法の目的である。

しかし、最短経路関連問題に対する索引付け手法の研究は未だに未成熟な状況にあった。理論的な結果として、任意のグラフにおける漸近的保証を持つ手法の開発が長らく取り組まれてきているものの、現実的な性能は実用に足るものになっていない。一方、現実的なグラフにおいて高い性能を達成する手法は、対象とする現実的なグラフの性質に依存したヒューリスティクスとして独立に開発されてきた。従って、複雑ネットワーク上での最短経路クエリ、有向無閉路グラフ上での到達可能性クエリ、道路ネットワーク上での最短経路クエリというような異なる状況が、完全に異なる問題として扱われ、個別にアプローチが開発されており、各個撃破の状態にあった。

そこで、本論文では統一的なアルゴリズムの枠組みである枝刈りラベリング法の提案を行う。提案手法は距離ラベルと呼ばれるデータ構造を前計算し索引として保存する。距離ラベルに基づく手法は一般にメモリ参照の局所性により応答性能が極めて優れている。しかし、既存手法はいずれも距離ラベルの計

算を異なる最適化問題への定式化を通じて間接的に行うため、スケーラビリティに問題があった。一方、枝刈りラベリング法は巧妙な枝刈りにより最短経路探索と同時に直接的に距離ラベルの計算を行うことができ、応答性能を犠牲にすることなくスケーラビリティを大幅に向上する。そして、最短経路探索を行う順番の変更により性質の異なるグラフの構造を活用でき、また異なる種類の距離ラベルに対しても同じアプローチでアルゴリズムを設計できるため、上記のように今まで異なる問題として扱われてきた問題に対して体系的にアルゴリズムを与えることができる。具体的な問題として、複雑ネットワーク上での最短経路クエリ、有向無閉路グラフ上での到達可能性クエリ、道路ネットワーク上での最短経路クエリ、複雑ネットワーク上での **Top-k** 最短経路クエリ、動的ネットワークにおける最短経路変化履歴クエリを扱う。実験によりそれぞれの問題における最新の手法との比較を行い、一部の問題では大きな性能改善を達成し、残りの問題でも少なくとも同程度の性能を達成することができることを示す。

さらに、提案手法を含むグラフ索引付け手法の性能を左右するグラフの性質を探る。例えほぼ同じサイズのグラフであっても、索引構築時間や索引サイズといった索引付け手法の性能がグラフによって大きく異なることがあり、その原因は今まで分かっていなかった。そこで、木分解という道具を用いることにより、このようなグラフに潜むサイズ以外の「難しさ」をある程度捉えることができることを理論的及び実験的に示す。