

ネットワークデータベースにおける同型部分グラフの問合せ方式

嘉本 貴彰[†]

北上 始[†]

森 康真[†]

広島市立大学 情報科学研究科[†]

1. はじめに

近年、ビッグデータの1つの表現形式としてネットワークデータが注目されてきている。ネットワークデータは Web ページや Twitter などのインターネット上に構築された社会的構造を表現したソーシャルネットワークなどに見られる。これらのネットワークデータをデータベースとして蓄積したものは、ネットワークデータベースと呼ばれている。

ネットワークデータベースに注目した応用分野では、同型部分グラフの問合せ処理やグラフマイニングによる支援が大変重要となっている。同型部分グラフの問合せ処理は、問合せグラフと同型な部分グラフを検索する処理を意味する。ネットワークデータベースの大規模化に伴い、問合せ処理の高速化が求められている。同型部分グラフの問合せ処理には Ullmann の同型性判定アルゴリズム^[1]が利用可能である。しかし、大規模なネットワークに対しては探索が不要な部分も参照し、より多くの時間を要してしまうという問題点がある。

そこで、本論文ではこの問題点を解決するために問合せグラフの半径に着目した高速な問合せ処理方式を提案する。

2. 従来手法

本章では提案手法の問合せ処理で同型性判定に用いた Ullmann のアルゴリズムについて説明する。Ullmann のアルゴリズムとはラベル無し無向グラフについて一方のグラフと同型な部分を他方のグラフから検出し同型性判定を行うアルゴリズムである。この手法について図1のグラフ例で説明する。なお、例では説明のためにラベルを付与している。図1の G_b から G_a と同型な部分を検出していくものとする。

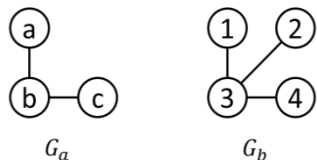


図1：同型性判定のグラフ例

まず、事前段階としてすべての頂点同士の対応の組合せを木構造で列挙する。この列挙の様子を図2に示す。図2の木において1段目が頂点 a,

2段目が頂点 b, 3段目が頂点 c に対応付けしたものである。頂点 b より高い次数を持っているか判断しながらこの木を探索していくことで同型性判定を行う。図2の反転している頂点は頂点 b より低い次数を持つため、木のこれらの頂点以下の部分については探索を回避し、探索空間を削減することができる。木を探索し終わった後に全ての頂点に対して次数の対応がとれているものが最終的に同型グラフとして検出される。

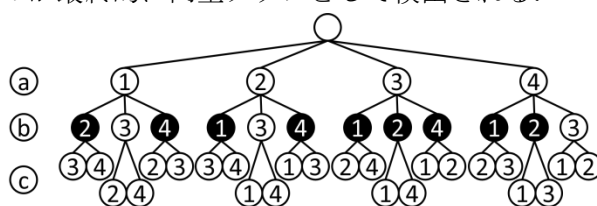


図2：組合せの列挙

3. 提案手法

本章では提案手法について説明する。提案手法は問合せグラフの半径に注目し、従来手法に参照範囲を限定する前処理を加えることで高速化を実現するものである。

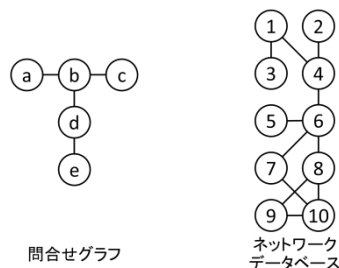


図3：提案手法におけるグラフ例

提案手法の例を図3のグラフ例を用いて説明する。なお、例では説明のために頂点にラベルを付与しているが、ラベル無し無向グラフについて考えるものとする。

(Step1)問合せグラフの中心、半径の決定

問合せグラフの中心と半径を決定する。問合せグラフの中心の決定にはグラフセンターと次数センターを併用して用いる。ダイクストラ法を用いて離心数を求め、グラフセンターと半径を決定する。グラフセンターによって複数の頂点为中心の候補とされた場合には、次数センターの考え方をを用いてそれらのうちより高い次数を持つ頂点を中心と決定する。図3の例ではグラフセンターによって問合せグラフの中心の候補として頂点 b と頂点 d が選択され、頂点 b の方が高い次数を持つことから最終的に頂点 b となる。

Isomorphic Graph Query Processing Method in Network Databases

[†]Takaaki Kamoto, Hajime Kitakami, Yasuma Mori, Graduate School of Information Sciences, Hiroshima City University

(Step2)参照グラフの作成

問合せグラフの半径に基づいて同型性判定を行う範囲を決定し、その範囲を参照グラフとする。Ullmann のアルゴリズムにおける次数での判断と同様に、問合せグラフの中心より高い次数を持つネットワークデータベースの頂点を探索し、その頂点を参照グラフの中心とする。次に参照グラフの各頂点に隣接している頂点を参照グラフに追加する操作を参照グラフの半径が問合せグラフと等しくなるまで繰り返す。図3の例では頂点bの次数は3であるので例として頂点4を参照グラフの中心とした場合を考える。問合せグラフの半径は2であるので、参照グラフの半径が2となるまで隣接頂点を追加し、参照グラフの頂点は{1,2,3,4,5,6,7,8}となる。参照グラフ作成の様子を図4に示す。

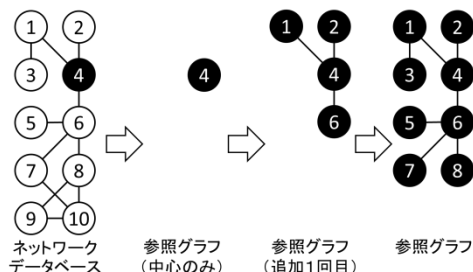


図4：参照グラフ作成の様子

(Step3)同型性判定

Ullmann の同型性判定アルゴリズムを用いて (Step2)で行った問合せグラフの中心と参照グラフの中心との対応付けを保持したまま、問合せグラフと参照グラフの同型性判定を行う。2回目以降は(Step2), (Step3)を繰り返すことによりすべての同型部分グラフを検出することができる。また、参照範囲の限定によって同型可能性が高い部分のみを探索し、探索が不要な部分の参照を回避することで高速化を行っている。

4. 評価実験

提案手法の有効性、性質を確かめるために行った評価実験について説明する。

(1) 実験方法

頂点数5の問合せグラフと頂点数100から500のネットワークデータベースを用いて従来手法、提案手法それぞれの同型部分グラフ問合せ処理の実行時間を計測した。問合せグラフ、ネットワークデータベース共にランダムに作成したものを使用し、ネットワークデータベースについてはエッジ数が頂点数の3倍、5倍、10倍のものをそれぞれ使用した。実際に使用した問合せグラフを図5に示す。

(2) 実験結果

実験結果を図6に示す。図6は縦軸を性能向上比、横軸をネットワークデータベースの頂点数として頂点数に対するエッジ数の比率が3倍、5

倍、10倍の時の結果を示したものである。縦軸については数値の差が大きいため対数表示で示している。

(3) 考察

図6から、ネットワークデータベースの頂点数が増加すると従来手法に対する提案手法の性能向上比も上昇していることがわかる。このことから、提案手法が従来手法に比べ問合せ処理が高速化されていることがわかる。また、頂点数に対してエッジ数が大きくなると性能向上比が低くなっていることから、提案手法は頂点数に対してエッジ数が小さいネットワークデータベースに対してより有効であることがわかる。逆に、エッジ数が頂点数に対して大きいグラフに関しては十分な性能が発揮できていないと言える。

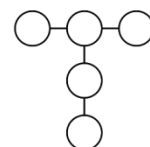


図5：評価実験に用いた問合せグラフ

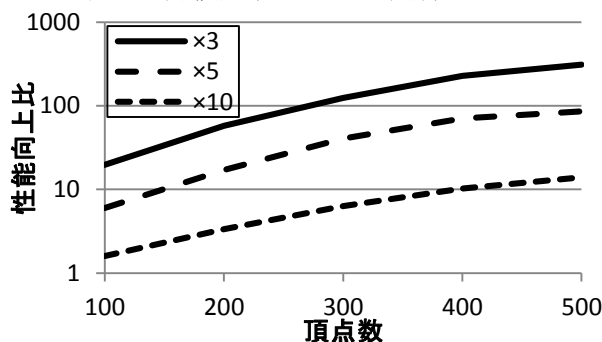


図6：評価実験結果

5. まとめ

ネットワークデータベースにおける同型部分グラフの問合せ処理に対して、問合せグラフの半径に着目し、参照範囲を限定することで高速化を行った。その結果、提案手法では従来手法よりも問合せ処理を高速化できるとわかった。

また、ネットワークデータベースの頂点数が増加するに連れて性能向上比が上昇することから大規模なネットワークデータベースに対しても有効であると考えられる。

しかし、頂点数に対してエッジ数が大きなネットワークデータベースに関しては十分に高速化できていないという問題点がある。この問題を解決するための新たな手法、アプローチを取り入れていくことが今後の課題として挙げられる。

参考文献

[1] J.R.Ullmann, "An algorithm for subgraph isomorphism," Journal of the ACM, vol.23, pp.31-42, 1976