

大規模グラフを対象とした部分グラフ同型判定における Interlace 定理の利用

長屋 未来[†] 片山 薫^{††} 石川 博^{††}

[†] 東京都立大学工学部電子情報工学科 〒192-0397 東京都八王子市南王沢 1-1

^{††} 首都大学東京システムデザイン学部 〒192-0397 東京都八王子市南王沢 1-1

E-mail: [†]mirai_n@hotmail.com, ^{††}{katayama,ishikawa}@eei.metro-u.ac.jp

あらまし 画像における特徴量抽出, 化学式における類似構造の検出に, 部分グラフ同型判定は応用できる. しかし部分グラフ同型判定問題は NP 完全であり, 大規模なグラフを扱う場合, 実用時間内に解くことが困難である. そこで, 多くの計算を必要とする判定を行う前に, より少ない計算により部分グラフを含まないものを除くことが考えられる. 対称行列とその主部分行列 (Principal Submatrix) の固有値の間には, コーシーの Interlace 定理として知られる関係がある. 隣接行列で表されたグラフの主部分行列は, 元のグラフの誘導部分グラフを表しているため, この定理を利用することで誘導部分グラフでないものを見つけることができる. さらに, グラフを接続行列を利用した行列で表現すると, 誘導部分グラフではない一般の部分グラフについて同様のことが可能となる. 我々は, これらの事実を部分グラフ同型判定処理の前処理として利用することを提案する. 本稿では, 様々なグラフデータを対象として本提案の有効性を検証した. さらに, Haemers らの提案したグラフの行列表現を拡張して, ラベル付きグラフに利用できるようにする方法と, 部分グラフのラベルを利用してグラフの行列表現を縮小し計算を効率化する方法を提案する. キーワード 部分グラフ同型, インターレース定理, 固有値

Application of Cauchy's Interlace Theorem to Subgraph Isomorphism Detection for Large Graphs

Hideki NAGAYA[†], Kaoru KATAYAMA^{††}, and Hiroshi ISHIKAWA^{††}

[†] Tokyo Metropolitan University, Electronics and Information Engineering Minamiosawa 1-1, Hachiojishi, Tokyo, 192-0397 Japan

^{††} Tokyo Metropolitan University, System Design Minamiosawa 1-1, Hachiojishi, Tokyo, 192-0397 Japan

E-mail: [†]mirai_n@hotmail.com, ^{††}{katayama,ishikawa}@eei.metro-u.ac.jp

Abstract It is often necessary to solve the subgraph isomorphism problem in various application such as feature extraction from pictures, retrieval for a chemical formula. This problem is NP-complete and require large computation cost to solve it. It is effective to reduce graphs that do not include a subgraph. There is a relation which is known as Cauchy's Interlace theorem between a symmetric matrix and its principal submatrix. We use eigenvalues of adjacency matrix and the matrix using incidence matrix. According to this theorem we can find graphs which are not included in. We use these facts for efficient subgraph isomorphism detection and evaluate our proposed method. We also propose how to express labeled graphs as matrices and how to reduce the size of the matrix for the efficient computation.

Key words Subgraph Isomorphism, Interlace Theorem, Eigenvalue

1. はじめに

化合物の検索 [7] や画像検索 [8] など様々な分野にグラフは利用され, あるグラフに他のグラフが含まれるかどうかを調べる (部分グラフ同型判定問題) ことは重要な課題である. しかし,

部分グラフ同型判定問題は NP 完全であり, 数多くのグラフや大きなグラフを扱う場合, 膨大な計算コストが必要となるため, より計算量の少ないアプローチで部分グラフ同型判定を補助する手法が有効であると考えられる.

グラフは, 隣接行列など行列として表すことが可能である.

また、対称行列とその主部分行列 (Principal Submatrix) の固有値には、コーシーの Interlace 定理と呼ばれる関係 [1] があることが知られている。グラフを隣接行列で表すと、主部分行列は、誘導部分グラフを表しているため、Interlace 定理を利用することで誘導部分グラフでないものを見つけることができる。誘導部分グラフではない一般の部分グラフは、主部分行列とはならないため、このままでは Interlace 定理を利用できない。しかし、Haemers ら [1] は、グラフを隣接行列ではなく、接続行列を利用した行列として表現すると、部分グラフを主部分行列とすることができることを指摘している。

我々は、これらの事実を部分グラフ同型判定処理の前処理として利用することを提案する。Interlace 定理を使って検索対象のグラフをフィルタリングすることで部分グラフ同型判定処理を効率化できると共に、固有値の計算に関する様々な知見を利用することができる。本稿では、様々なグラフデータを対象として本提案の有効性を検証した。さらに、Haemers らの提案したグラフの行列表現を拡張して、ラベル付きグラフに利用できるようにする方法と、部分グラフのラベルを利用してグラフの行列表現を縮小し計算を効率化する方法を提案する。

実験から、以下の場合について、本手法は有効である事が確認できた。

- 誘導部分グラフの判定について、2つのグラフの頂点数の差が小さく、密なグラフの中に疎なグラフが含まれるかの判定を繰り返し行う場合
- 部分グラフの判定について、2つのグラフの頂点数の差が小さく、枝数が同程度、ラベルが無い場合

2. 関連研究

部分グラフ同型判定問題における組み合わせ的なアルゴリズムとしては、VF2 [2] や Ullmann による手法 [3] がある。Ullmann の手法は、検索範囲を減少させる手続きであるバックトラック法を使ってグラフ同型判定と部分グラフ同型判定を行う手法である。VF2 は、深さ優先探索に基づく手法で、大規模で複雑なグラフに対してグラフ同型判定、部分グラフ同型判定を行うことができる。Interlace 定理は、行列で表現されたグラフの固有値を使って、部分グラフではないものを見つける手法である。本手法は、部分グラフであるか分からないものについて、組み合わせ的なアルゴリズムを用いて、部分グラフ同型判定の処理をより高速にする。

グラフに索引付けを行う手法としては、Yan ら [4] の、グラフデータベース内の頻出するグラフや特徴のあるグラフを検索し、索引として使用する gIndex と呼ばれる手法がある。

グラフの類似構造検索の手法としては、Yan ら [5] は、頻出するグラフや特徴のあるグラフを使ってフィルタリングを行い、グラフの類似検索を行う手法を提案している。

3. 対称行列における固有値の特性

3.1 主部分行列

主部分行列は、対称行列 A の内部より得られる行列であり、対称行列 A から同じ番号の行と列を抜き出すことで構成する。

主部分行列は、対称行列である。対称行列 A が図 1 であるとする。このような行列で主部分行列を構成する場合、主部分行列が対称行列であるためには、抜き出す行と列を常に同じにする。行列 A の 1,2,4 番目の行を抜き出すなら、列も 1,2,4 番目を抜き出す。行列 A から行と列の 1,2,4 番目を抜き出して作られた主部分行列 H を図 1 に示す。同じ行と列を抜き出すならば、1,4 番目や 2,3,4 番目のように抜き出しても主部分行列となる。

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12} & a_{22} & a_{23} & a_{24} \\ a_{13} & a_{23} & a_{33} & a_{34} \\ a_{14} & a_{24} & a_{34} & a_{44} \end{pmatrix} \quad H = \begin{pmatrix} a_{11} & a_{12} & a_{14} \\ a_{12} & a_{22} & a_{24} \\ a_{14} & a_{24} & a_{44} \end{pmatrix}$$

図 1 主部分行列の例

3.2 コーシーの Interlace 定理

式 (1) で表される大きさが n の対称行列 A を考える。 H は、 A の主部分行列であり、その大きさは $m (n > m)$ であるとする。

$$A = \begin{pmatrix} H & B \\ B^* & C \end{pmatrix} \quad (1)$$

対称行列 A の固有値を $\lambda_i (i = 1, 2, \dots, n, \lambda_1 < \lambda_2 < \dots < \lambda_n)$ 、主部分行列 H の固有値を $\mu_j (j = 1, 2, \dots, m, \mu_1 < \mu_2 < \dots < \mu_m)$ とするとき、 H が対称行列 A の主部分行列であれば、全ての $k (k = 1, 2, \dots, m)$ について次式で表される不等式が成り立つ。以降、本稿において、全ての k について式 (2) の不等式が成り立つ事を「Interlace 定理を満たす」といい、1 つでも不等式が成り立たない場合を「Interlace 定理を満たさない」という。

$$\lambda_k \leq \mu_k \leq \lambda_{k+(n-m)} \quad (k = 1, 2, \dots, m) \quad (2)$$

対称行列 A の主部分行列 H であれば、必ず Interlace 定理を満たす。しかし、 H が A の主部分行列では無い場合であっても、定理を満たす事があるので、定理を満たす場合でも H は対称行列 A の主部分行列であるとは限らないが、Interlace 定理を満たさない場合、 H は対称行列 A の主部分行列では無い。

4. ラベル付きグラフの行列表現

4.1 誘導部分グラフと部分グラフ

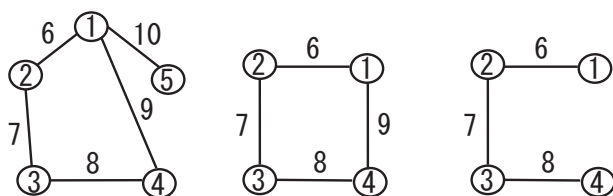


図 2 誘導部分グラフと部分グラフの例

グラフの頂点の集合を V 、枝の集合を E とする。 $A(V, E)$

と $H(V', E')$ の2つのグラフがあるとき、頂点を x, y とし、2つの頂点を結ぶ辺を xy とする。全ての $x, y \in V$ に対して $E \in E'$ を満たすとき、グラフ A とグラフ H を同型という。ここで、 $G \cup G' := (V \cup V', E \cup E')$, $G \cap G' := (V \cap V', E \cap E')$ とおく。 $G \cap G' = \emptyset$ のとき、 G と G' は交わらない。 $V \subseteq V'$ かつ $E \subseteq E'$ ならば、グラフ H をグラフ A の部分グラフといい、 $H \subseteq A$ と書く。 $H \subseteq A$ であり、 H が $x, y \in V'$ となる辺 $xy \in E$ を全て含むとき、 H を A の誘導部分グラフという。部分グラフと誘導部分グラフの例を図2に示す。

4.2 グラフの隣接行列表現

隣接行列は、以下のように定義される。 $V = \{v_1, \dots, v_n\}$ と $E = \{e_1, \dots, e_m\}$ をもつグラフ $A = (V, E)$ に対して、隣接行列 $A = (a_{ij})_{n \times n}$ は、

$$A_{ij} = \begin{cases} 1 & (v_i v_j \in E \text{ のとき}) \\ 0 & (\text{そうでないとき}) \end{cases} \quad (3)$$

枝にラベル存在する場合、非対角要素にラベルを入れ、ラベルが無い場合は、1を入れる。隣接行列は、対称行列になる。このように定義された行列を、本稿では隣接行列と呼ぶ。

図3を隣接行列表現を図4に示す。頂点、枝の両方にラベルが存在する場合、行列の対角要素に頂点のラベルを入れる。頂点1と頂点2が枝で繋がっているなら、(1, 2)要素と(2, 1)要素に値を入れる。

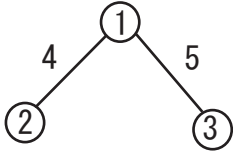


図3 グラフ例

	1	2	3
1	1	4	5
2	4	2	0
3	5	0	3

図4 非対角要素に枝を入れる

$$A = \begin{pmatrix} 1 & 6 & 0 & 9 & 0 \\ 6 & 2 & 7 & 0 & 0 \\ 0 & 7 & 3 & 8 & 0 \\ 9 & 0 & 8 & 4 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 6 & 0 & 9 \\ 6 & 2 & 7 & 0 \\ 0 & 7 & 3 & 8 \\ 9 & 0 & 8 & 4 \end{pmatrix}$$

図5 グラフ A とその誘導部分グラフ B (図2) の隣接行列

図2のグラフ A の隣接行列 A と誘導部分グラフの隣接行列 B を図5に示す。誘導部分グラフの隣接行列表現では、図5の行列 B のように、行列 A の主部分行列となる。

4.3 グラフの接続行列表現

接続行列は、以下のように定義される。 $V = \{v_1, \dots, v_m\}$ と $E = \{e_1, \dots, e_n\}$ をもつグラフ $A = (V, E)$ に対して、接続行列 $B = (b_{ij})_{m \times n}$ は、

$$B_{ij} = \begin{cases} 1 & (v_i \in e_j \text{ のとき}) \\ 0 & (\text{そうでないとき}) \end{cases} \quad (4)$$

図6に図3の接続行列を示す。頂点と枝に接続関係があれば1, 無ければ0が、各要素に入る。

	4	5
1	1	1
2	1	0
3	0	1

図6 図3の接続行列表現

5. 提案手法

5.1 Interlace 定理の利用

本研究では、組み合わせ的なアルゴリズムと Interlace 定理を組み合わせることで、大量のグラフに対する部分グラフ同型判定処理を高速化する手法を提案する。本手法による手順を図7に示す。与えられた二つのグラフを行列にし、Interlace 定理による判定を行う。Interlace 定理を満たさないと判定されたものは、部分グラフでは無いので、組み合わせ的なアルゴリズムによる部分グラフ同型判定を行う必要が無い。Interlace 定理を満たすものについては、組み合わせ的なアルゴリズムを使用して部分グラフの判定を行う。本研究では、効率的な組み合わせアルゴリズムとして知られる VF2 を使用する。

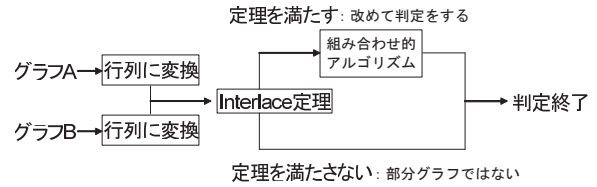


図7 提案手法

5.2 ラベル付グラフの接続行列を利用した表現

A, H の二つのグラフがあるとき、 H が A の誘導部分グラフなら、隣接行列 H は、隣接行列 A の主部分行列となるが、部分グラフの隣接行列は、主部分行列にはならない。図2のグラフ A の隣接行列 A と部分グラフの隣接行列 C を、図8に示す。部分グラフの隣接行列表現では、図8の行列 C と A のように H が A の主部分行列にならない。頂点1,4の枝9が無くなる事で、行列の非対角要素要素が変化したので、主部分行列にならなくなったためである。部分グラフについて、Interlace 定理を利用する場合、隣接行列とは別の行列表現が必要になる。

$$A = \begin{pmatrix} 1 & 6 & 0 & 9 & 10 \\ 6 & 2 & 7 & 0 & 0 \\ 0 & 7 & 3 & 8 & 0 \\ 9 & 0 & 8 & 4 & 0 \\ 10 & 0 & 0 & 0 & 5 \end{pmatrix} \quad C = \begin{pmatrix} 1 & 6 & 0 & 0 \\ 6 & 2 & 7 & 0 \\ 0 & 7 & 3 & 8 \\ 0 & 0 & 8 & 4 \end{pmatrix}$$

図8 グラフ A とその部分グラフ C (図2) の隣接行列

そこで、部分グラフでも Interlace 定理を利用できるようにするため、Haemers [1] は、接続行列 N を用いた式(5)の行列を考えることを提案した。

$$M = \begin{pmatrix} 0 & N \\ N^t & 0 \end{pmatrix} \quad (5)$$

N^t は、接続行列 N の転置行列である。

接続行列と式 (5) をを使うことで、対称行列として表現できるが、この行列表現では、頂点、枝のラベルが表現できない。そこで、本研究では、この行列表現を応用し、対角要素に頂点と枝のラベルを入れてラベル付グラフを表現する手法を提案する。図 6 の接続行列 N と式 (5) を使用することで、図 3 のグラフは、図 9 の行列で表現できる。

	1	2	3	4	5
1	1	0	0	1	0
2	0	2	0	1	1
3	0	0	3	0	1
4	1	1	0	4	0
5	0	1	1	0	5

図 9 接続行列を用いた対称行列表現

接続行列による表現を用いると部分グラフにおいても主部分行列を作ることが可能である。図 2 のグラフ A と部分グラフ C を接続行列を用いた行列で表すと、図 10 になる。接続行列を用いた表現は枝の数だけ行列が大きくなるため、枝が密であるようなグラフを表現するとき、行列が非常に大きくなる。また、非対角要素に入る 1 の数は枝数の 4 倍になる。

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 2 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 6 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 7 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 8 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 9 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 10 \end{pmatrix}$$

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 3 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 4 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 6 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 7 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 8 \end{pmatrix}$$

図 10 グラフ A とその部分グラフ C (図 2) の接続行列を用いた行列表現

5.3 ラベルに基づく行列の縮小

ラベルを使用しているグラフについて、ラベルが一致しない部分は、部分グラフの判定をしなくて良い。行列についてもラベルの情報を利用する方法は有効である。本研究では、ラベルにより行列を縮小し、固有値の計算量を減らす手法を提案する。ラベルの情報を利用して、小さいグラフに含まれるラベル以外の行と列を大きいグラフの行列から消去し、行列を縮小する。手順を以下に示す。

(1) 小さい行列の対角要素に使用されているラベルを保存する。

(2) 大きい行列の対角要素の中に、小さい行列で使用されていないラベルがあれば、そのラベルの対角要素を含む行と列を除く。

これにより固有値を計算するコストを削減することができる。2つの隣接行列を挙げて説明する。

$$A = \begin{pmatrix} 4 & 1 & 3 & 0 & 0 \\ 1 & 2 & 0 & 4 & 0 \\ 3 & 0 & 3 & 0 & 4 \\ 0 & 4 & 0 & 3 & 2 \\ 0 & 0 & 4 & 2 & 5 \end{pmatrix} \quad H = \begin{pmatrix} 4 & 3 & 0 \\ 3 & 3 & 4 \\ 0 & 4 & 5 \end{pmatrix}$$

図 11 行列の例

$$A' = \begin{pmatrix} 4 & 3 & 0 & 0 \\ 3 & 3 & 0 & 4 \\ 0 & 0 & 3 & 2 \\ 0 & 4 & 2 & 5 \end{pmatrix}$$

図 12 図 11A の縮小行列 A'

行列 H に存在しているラベルは、対角要素の情報と同じである。行列 H の場合、3, 4, 5 がグラフに含まれる頂点のラベルとなる。3, 4, 5 以外のラベル情報を対角要素として持つ(図 11 では、2) 行と列を削除すると、行列 A の主部分行列 A' を形成する。行列 A' は、行列 A の主部分行列であるが、行列 H の主部分行列でもある事がわかる。数値計算において、行列の大きさが小さいほうが計算量は少なくなる。

6. 評価実験

6.1 実験の概要

次の 2 つについて評価した。

(1) Interlace 定理の判定精度

グラフを隣接行列で表現すると、誘導部分グラフについて Interlace 定理を利用することが出来る。誘導部分グラフであれば、Interlace 定理を満たす。しかし、誘導部分グラフでは無い場合、Interlace 定理を満たさないと限らない。そこで、誘導部分グラフでは無いものが、Interlace 定理をどの程度満たさないのか実験を行った。

(2) 計算時間

計算時間を計測する実験として、部分グラフではないグラフと、部分グラフを混ぜて実験を行った。部分グラフと部分グラフでは無いものについて、誘導部分グラフの場合と部分グラフの場合の実験を行った。以下の 4 つの計算時間を計測した。

(a) VF2 のみで同型判定を行った処理時間

(b) 本提案手法の処理時間

(c) Interlace 定理による判定を行った処理時間

(d) ラベルによる行列の縮小を行った場合の Interlace 定理の処理時間

本研究の対称行列における固有値計算は、線形代数計算パッケージ LAPACK [9] のライブラリを使用して行った。実験環境は、CPU:Pentium4 2.2GHz、メモリ:1GB OS:WindowsXP である。本実験でグラフにラベルを使用する場合、頂点に使用した。実験は、単純グラフを用いて行う。単純グラフは、自己ループ（枝の両端の頂点と同じ）を含まず、2 頂点に複数の枝が存在しないものをいう。単純グラフでは、頂点数 N に対して使用できる枝の最大数は、 $\frac{N}{2}(N-1)$ となる。実験で使用するグラフは、頂点数と枝数は指定し、頂点と枝の接続関係や頂点へのラベルは無作ために生成した。

6.2 Interlace 定理の判定精度

この実験では、誘導部分グラフでは無いものを使用した。誘導部分グラフについての判定なので、グラフを隣接行列で表現した。誘導部分グラフでは無いものであっても、Interlace 定理を満たす場合、組み合わせアルゴリズムで誘導部分グラフを調べなければならない、Interlace 定理の計算時間と VF2 の計算時間がかかるため、提案手法は有効ではない。

隣接行列の場合、グラフの頂点数と枝数が行列の大きさと非対角要素の要素数になるため、グラフの頂点数と枝を変えて Interlace 定理を判定する事は、行列の大きさや非対角要素の要素数を変えて Interlace 定理を判定する事と同じである。また、誘導部分グラフでは無いものを使用したのは、Interlace 定理を使用する場合、必ず主部分行列にならないものを作るためである。この実験によって、主部分行列では無いものを Interlace 定理で判定したとき、行列の大きさや非対角要素の要素数の変化によって、どの程度 Interlace 定理を満たさないのかを検証できる。

実験では、頂点数の差を 50,30 のように固定し、大きいグラフの枝数も固定する。小さいグラフの枝数を変えたとき、どの程度 Interlace 定理を満たさないのかを検証した。小さいグラフの頂点数を変えて、同様の実験を行うことで、頂点数を変化させた場合にどの程度 Interlace 定理を満たさないのかを検証した。また、頂点数 100 と 80, 100 と 60 のグラフについて、同様の実験を行うことで、2 つのグラフのスケールを大きくすることによって、どの程度 Interlace 定理を満たさないのかを検証した。

最後に、誘導部分グラフであるものについて、Interlace 定理の誤差実験をした。

6.2.1 頂点数の差が 50 と 30 の場合

頂点数 50 枝数 500 のグラフと頂点数 30 のグラフを使って誘導部分グラフでは無いグラフがどの程度 Interlace 定理を満たさないのか実験を行った。結果を図 13 に示す。同様の実験を頂点数 50 枝数 250 のグラフと頂点数 30 について行った。結果を図 14 に示す。

図 13 により、頂点数 50 枝数 500 のグラフの隣接行列と頂点数 30 枝数 250 のグラフの隣接行列は、Interlace 定理を全て満たす結果となった。頂点数 30 枝数 250 以下のグラフも全て Interlace 定理を満たす結果となった（ラベルを頂点に 25 種類使用した結果を除く）頂点数 50 枝数 500 のグラフと頂点数 30 枝数 250 以下のグラフについて、誘導部分グラフではないもの

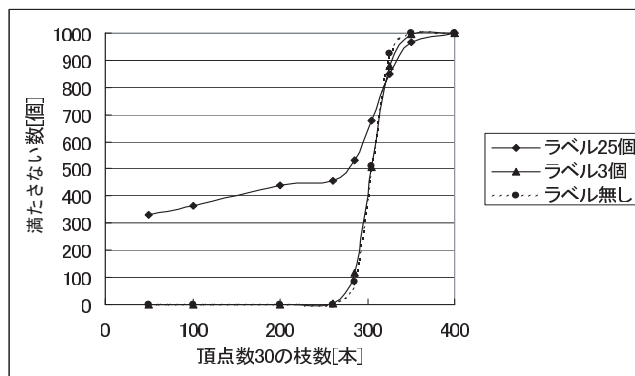


図 13 頂点数 50 枝数 500 と頂点数 30 の枝数変化の関係

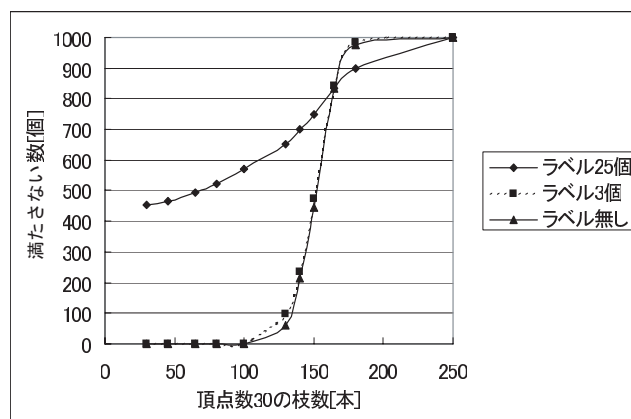


図 14 頂点数 50 枝数 250 と頂点数 30 の枝数変化の関係

が全て Interlace 定理を満たすので、全て VF2 で誘導部分グラフを判定しなければならない、Interlace 定理の使用が有効ではないと考えられる。頂点数 30 の枝の数を増やしていき、枝数 350 程度になると、Interlace 定理を満たさない。頂点数 50 のグラフに枝数に近い枝数を頂点数 30 が持つとき、Interlace 定理を満たさないものが増える。Interlace 定理を満たさない部分で、本手法は有効である。図 13 の実験では、図 14 の実験で使用したグラフの 2 倍の枝数をもつグラフを使用した。この場合でも、図 13 と図 14 Interlace 定理の判定精度は変わらない。大きいグラフの枝数と同じ割合で小さいグラフの枝数を変化させれば、Interlace 定理の判定精度は、似た特性が得られると考えられる。

6.2.2 頂点数の差が 50 と 40 の場合

頂点数 50 枝数 500 のグラフと頂点数 40 のグラフを使って誘導部分グラフでは無いグラフがどの程度 Interlace 定理を満たさないのか実験を行った。結果を図 15 に示す。同様の実験を頂点数 50 枝数 750 のグラフと頂点数 40 について行った。結果を図 16 に示す。

図 15 の結果から、頂点数 50 に枝数 500 のグラフに近い枝数を頂点数 40 が持つ場合、Interlace 定理で満たさない判定される数が増える。この部分は、頂点数 50 と頂点数 30 の結果と同じであると考えられる。しかし、頂点数 40 の場合、頂点数 40 の枝数 50 程度の少ない場合についても Interlace 定理で満たさないものが増える結果になった。図 15 で使用したグラフ

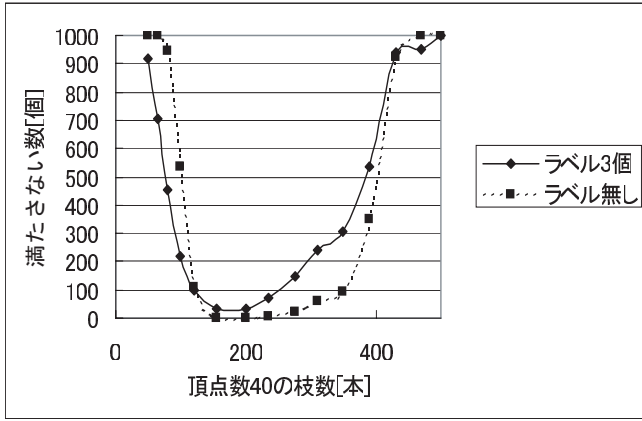


図 15 頂点数 50 枝数 500 と頂点数 40 の枝数変化の関係

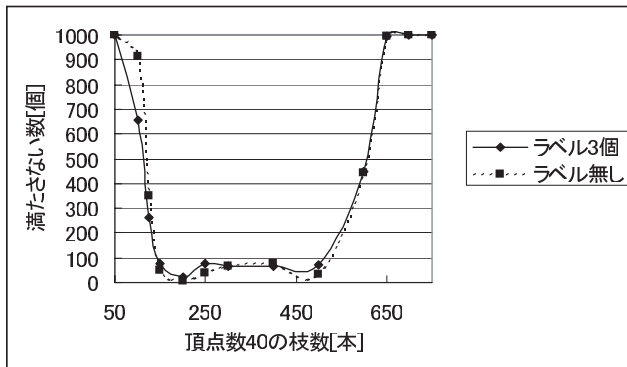


図 16 頂点数 50 枝数 750 と頂点数 40 の枝数変化の関係

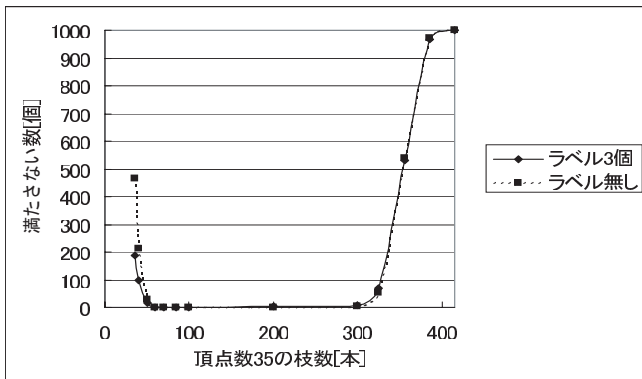


図 17 頂点数 50 枝 500 本と頂点数 35 の枝数変化の関係

の持つ枝数を 1.5 倍したのが、図 16 である。図 15 と図 16 についても、Interlace 定理の判定精度が変わらない。

6.2.3 頂点数の差を変えた場合

頂点数 50 枝数 500 のグラフと頂点数 35 のグラフを使って誘導部分グラフでは無いグラフがどの程度 Interlace 定理を満たさないのか実験を行った。結果を図 17 に示す。

頂点数 10 のグラフと頂点数 50 に枝数 500 のグラフ、頂点数 50 に枝数 60 のグラフを使って誘導部分グラフでは無いグラフがどの程度 Interlace 定理を満たさないのかを実験した。結果を図 18 に示す。

図 18 の結果から、頂点数 50 に枝数 500 と比較した場合、頂点数 10 の枝数が少なくても、Interlace 定理が満たす。また、

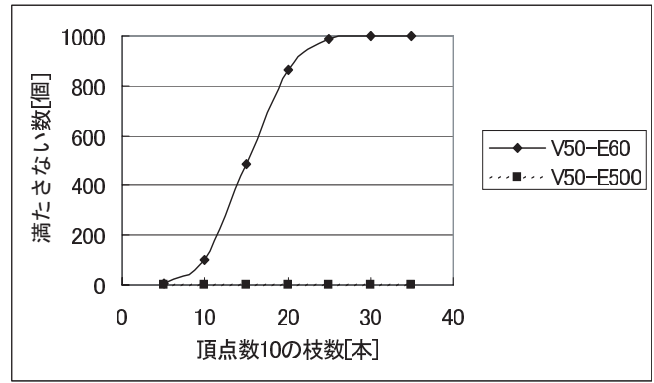


図 18 頂点数 10 の枝数の変化と頂点数 50 の枝数の関係

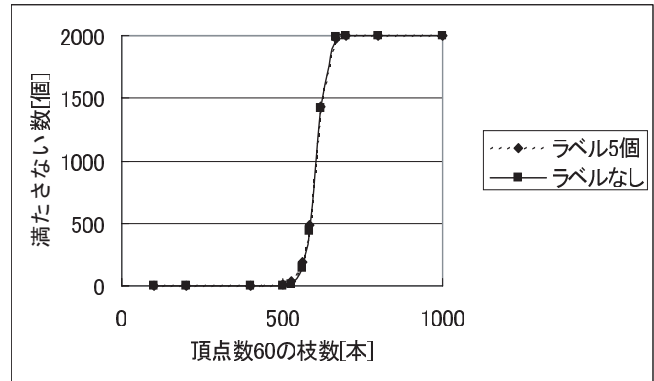


図 19 頂点数 100 枝 1000 本と頂点数 60 の枝数変化の関係

図 13 の頂点数 50 と頂点数 30 の結果でも、頂点数 30 の枝数が少ないグラフで、全て Interlace 定理を満たす。しかし、図 17 の結果では、頂点数 35 の枝数の少ないグラフで、Interlace 定理を満たさないものがある。さらに、図 15 の結果では、頂点数 40 の枝数の少ないグラフで、Interlace 定理を満たさないものがあり、頂点数 35 のグラフより数が多い。この結果から、比較する頂点数の差が 50 と 40 のように近いと、図 15 のように、小さいグラフの枝数が少ない部分で Interlace 定理の判定で満たさないものが増えると考えられる。さらに、今までの実験結果と図 18 の結果から、大きいグラフの枝数と小さいグラフの枝数が近い場合は、頂点数の差に係わらず Interlace 定理を満たさないものが増えると考えられる。

6.2.4 グラフのスケールを大きくした場合

頂点数 100 枝数 1000 のグラフと頂点数 60 のグラフを使って誘導部分グラフでは無いグラフがどの程度 Interlace 定理を満たさないのか実験を行った。結果を図 19 に示す。同様の実験を頂点数 100 枝数 2000 のグラフと頂点数 80 について行った。結果を図 20 に示す。

図 19 では、図 14 と変わらず図 20 は図 15 と変わらない Interlace 定理の判定精度となった。頂点数を変化させても、Interlace 定理の判定は、大きいグラフと小さいグラフの枝数の割合によって、似た特性を示すと考えられる。頂点数が 1000 と 800 の様な大きいグラフで Interlace 定理を使用しても、同様の特性を示す事を確認した。

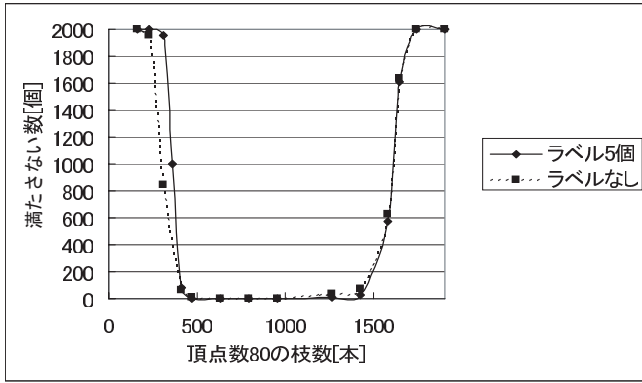


図 20 頂点数 100 枝 2000 本と頂点数 80 の枝数変化の関係

6.2.5 Interlace 定理の判定精度まとめ

小さい行列が主部分行列では無い場合について、以下の場合、Interlace 定理を使用して部分グラフでは無いものを高精度で判定できると考えられる。

(1) 大きい行列の非対角要素の要素数と小さい行列の非対角要素の要素数が近い場合について主部分行列では無いもの

(2) 大きい行列と小さい行列の大きさの差が小さく、小さい行列の非対角要素が疎であり、大きい行列の非対角要素が密であるもの

6.2.6 計算精度の影響

計算誤差によって、主部分行列である場合に Interlace 定理を満たさない状態は起こるかを検証した。使用した行列は大きさ 100 と大きさ 100 の主部分行列である大きさ 80 の行列を 5000 個、大きさ 100 と大きさ 100 の主部分行列である大きさ 60 の行列を 5000 個用意して実験を行った。結果、10000 個のデータについて、9999 個が Interlace 定理を満たす結果になった。大きさ 100 と大きさ 100 の主部分行列である大きさ 80 の行列について、Interlace 定理で比較した 1 組のみ、満たさないと判定される場合があった。このデータについて、満たされなかった部分についての固有値を調べた結果、 10^{-14} 部分の値が Interlace 定理を満たさないと判定される要因であった。本実験では、計算における誤差を考慮していないため、計算誤差により Interlace 定理の判定に間違い含まれると考えられる。

6.3 計算時間

Interlace 定理の実験結果から、Interlace 定理の利用が有効であるグラフについて、計算速度の実験を行った。本実験は、誘導部分グラフと部分グラフの判定について行った。計算時間を計測したのは以下の 4 つである。

- (1) VF2 のみで同型判定を行った処理時間
- (2) 本提案手法の処理時間
- (3) Interlace 定理による判定を行った処理時間
- (4) ラベルによる行列の縮小を行った場合の Interlace 定理の処理時間

6.3.1 誘導部分グラフにおける結果と考察

Interlace 定理の精度実験において、比較する頂点数が近い場合、密なグラフと疎なグラフを Interlace 定理で判定すると判定精度が良い事が分かった。そこで、本実験では、頂点数が

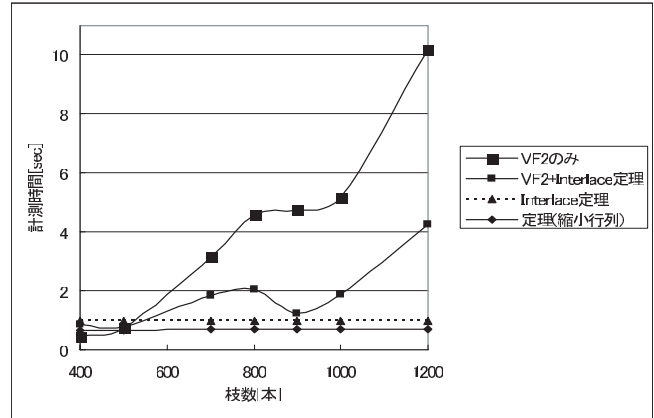


図 21 頂点数 80 に枝 100 のグラフと頂点数 100 のグラフ枝数の関係

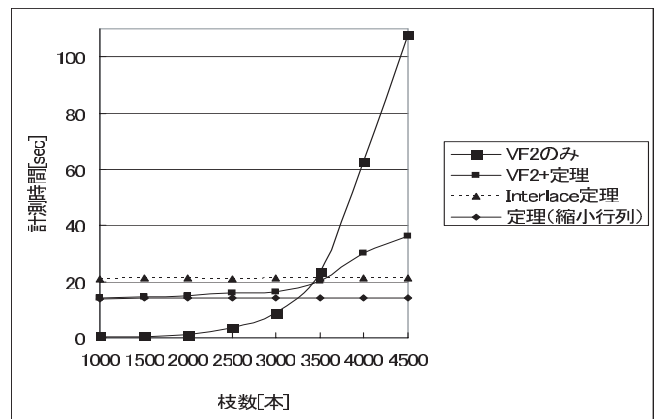


図 22 頂点数 250 枝 500 のグラフと頂点数 300 のグラフ枝数の関係

100 と 80, 300 と 250 の場合について 1 対 1 の部分グラフ同型判定を 100 回行う実験をした。

頂点数 80 で頂点に 1 から 3 のラベルを使用した枝数 100 のグラフと、頂点に 1 から 5 のラベルを使用した頂点数 100 のグラフを使って、誘導部分グラフであるものを 50 個、誘導部分グラフではないものを 50 個用意して速度計測を行った。結果を図 21 に示す。頂点数 250 で頂点に 1 から 5 のラベルを使用した枝数 500 のグラフと、頂点に 1 から 10 のラベルを使用した頂点数 300 のグラフを使って、誘導部分グラフであるものを 50 個、誘導部分グラフではないものを 50 個用意して速度計測を行った結果を図 22 に示す。

頂点数 100 のグラフの頂点には、1 から 5 までのラベルが使用され、頂点数 80 のグラフは 1 から 3 までのラベルが使用されているため、頂点数 80 のグラフは、4,5 というラベルを含まない。図 21 において、Interlace 定理の判定の計算速度は、ほぼ一定のとなり、平均計算時間は 0.68sec となる。行列の縮小を行わずに Interlace 定理の判定を行う場合もほぼ一定で、平均計算時間は、0.98sec となる。頂点数 100 のグラフの枝が多くなるに従って、VF2 による誘導部分グラフの判定に時間がかかる。隣接行列の場合、枝は非対角成分で表されているので、行列の大きさに影響が無いため、Interlace 定理による判定の速度は一定である。密なグラフの中に疎なグラフが誘導部分グラフとして含まれているかを判定する場合、全体の処理速度が速

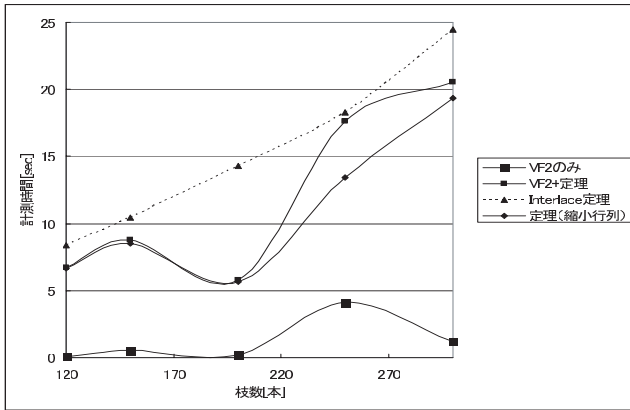


図 23 頂点数 80 に枝 100 のグラフと頂点数 100 のグラフ枝数の関係

くなり、本手法が有効であると確認できる。

疎であるグラフと疎である部分グラフの判定の場合は、Interlace 定理による前処理を行うより、VF2 で全てを計算する方が速い事が分かる。図 22 では、頂点数 300 と 250 で実験を行っているので、頂点数 80 と 100 の場合より、隣接行列が大きくなる。行列が大きくなると、固有値を計算するのに時間がかかるようになる。しかし、頂点数を変えて実験を行った結果についても、密なグラフの中に疎なグラフが誘導部分グラフとして含まれているかを判定する場合、本手法が有効である事が確認できる。

6.3.2 部分グラフにおける結果と考察

部分グラフではないグラフ 50 個と部分グラフであるグラフ 50 個を混ぜて 1 対 1 の部分グラフ同型判定を 100 回行う実験をした。

まず、頂点数 80 で頂点に 1 から 3 のラベルを使用した枝数 100 のグラフと、頂点に 1 から 5 のラベルを使用した頂点数 100 のグラフを使って部分グラフであるものを 50 個、部分グラフではないもの 50 個を用意して速度実験を行った。結果を図 23 に示す。図 23 の結果から、枝数が増えるに従って、Interlace 定理による判定時間が増えることが分かる。Interlace 定理による判定時間が増えていくのは、頂点数と枝数の和が行列の大きさである接続行列を利用した行列を使用しているためである。実験した枝数の範囲における、部分グラフの判定において、2 つのグラフの頂点数の差が小さく、ラベルを使用しているグラフでは、枝数を増やしても、VF2 のみで部分グラフの判定を行う方が処理速度が速いことが分かる。

次に、頂点数 100 に枝数 900 のグラフと、頂点数 80 に枝数 800 のグラフについて部分グラフであるものを 50 個、無いものを 50 個用意して、頂点に使用するラベルの数を変化させて計算速度の実験を行った。速度計測は、VF2 のみ、Interlace 定理の計算時間、VF2 と Interlace 定理を組み合わせたもので行った。本結果でも、頂点にラベルを使用しない場合の時間は、非常に時間が掛かったため計測不能とした。実験結果を表 1 に示す。

頂点ラベルがある場合、VF2 のみで部分グラフ同型判定を行う方が速いことが分かる。ラベルが無い場合に限り、本手法が

より高速であると分かる。本実験から、グラフの規模が大きくなってくる場合に頂点、枝にラベルを使用しないグラフの部分グラフ同型判定を行うと、判定を行うのにさらに時間がかかるため、本手法が有効になると考えられる。

ラベル数	VF2 のみ	Interlace	VF2+定理
無し	-	974.5[sec]	974.5[sec]
2	342[sec]	969.6[sec]	969.6[sec]
3	2.604[sec]	956.6[sec]	956.6[sec]
4	0.233[sec]	967.3[sec]	967.3[sec]

表 1 頂点数 100 と 80 における計測結果

7. ま と め

本論文では、部分グラフ同型判定の前処理として Interlace 定理を利用することを提案し、様々なグラフデータに対してその有効性を検証した。実験から、以下の場合に提案手法が有効であることが分かった。

- 誘導部分グラフの判定について、頂点の差が小さく、密であるグラフの中に疎であるグラフが含まれるかの判定を行う場合
- 部分グラフの判定について、頂点の差が少なく、枝の本数が近いときにラベルが無い場合

今後の課題として、次のものが挙げられる。

- Interlace 定理の効率的な判定方法の開発
- 固有値計算における計算誤差の考慮
- 提案手法の計算量の評価

謝辞 本研究の一部は、(独)日本学術振興会科学研究費補助金基盤研究(B)(2)(課題番号:16300030)による。多くの貴重なコメントを頂いた査読者の方々に深く感謝いたします

文 献

- [1] Willem H. Haemers, "Interlacing Eigenvalues and Graphs," Linear Algebra Appl. 226, pp.593-616, 1995.
- [2] L.P. Cordella, P. Foggia, C. Sansone, M. Vento, "An Improved Algorithm for Matching Large Graphs," Proc. of the 3rd IAPR TC-15 Workshop on Graph-based Representations in Pattern Recognition, Ischia, May 23-25, pp. 149-159, 2001.
- [3] J.R. Ullmann, "An algorithm for subgraph isomorphism," JAssoc. Comput. Mach, vol.23, pp.31-42,1976.
- [4] X. Yan, P. S. Yu, J. Han, "Graph Indexing: A Frequent Structure-based Approach," Proc. 2004 ACM-SIGMOD Int. Conf. on Management of Data (SIGMOD'04), Paris, France, June 2004.
- [5] X. Yan, P. S. Yu, J. Han, "Substructure Similarity Search in Graph Databases," Proc. of 2005 Int. Conf. on Management of Data (SIGMOD'05), 2005.
- [6] A. Shokoufandeh, S.j. Dickinson, K. Siddiqi, S.W. Zucker, "Indexing using a spectral encoding off topological structure," In Proc. IEEE Int'l Conf Computer Vision and Pattern Recognition (CVPR'99),Fort Collins, CO,Jun.1999
- [7] 高橋由雅, 藤島悟志, 加藤博明, "化学物質の構造類似性に基づくデータマイニング," J.Comput. Chem.Jpn.,vol.2, No.4,pp.119-126, 2003.
- [8] 下野弘貴, 佐藤正実, 北澤仁志, "特徴空間中の部分グラフ間距離の高速計算による実時間行動識別," 信学技報, vol. 104, no. 669, PRMU2004-197, pp. 109-114, 2005 年 2 月.
- [9] CLAPACK, "http://www.netlib.org/clapack/