

最大共通誘導部分グラフ問題の MAX SNP-hardness について

杉野 孔一*

正代 隆義*

* 九州大学大学院システム情報科学研究科情報理学専攻

{sugino, shoudai}@i.kyushu-u.ac.jp

1 はじめに

最大共通誘導部分グラフ問題とは、入力として2つのグラフが与えられたときに、最大で共通な頂点から誘導される部分グラフを求める問題である。2つのグラフの頂点数と辺の数が等しいとき、それらのグラフから最大共通誘導部分グラフを取り除いた残りのグラフの頂点数を、入力の2つのグラフの距離と考えることができる。グラフ理論において、グラフの距離はいくつか定義され、それらの関係について研究されている [5]。また、あるグラフの変換操作を固定し、一方のグラフから他方のグラフへ変換するその操作の回数を距離と定義したとき、その距離を計算する問題に対して計算量理論的な研究もなされている [1, 4]。V. Kann [3] は、グラフの類似性を判定する最大共通誘導部分グラフ問題が、MAX SNP-hardであることを示した。このことはすなわち、この問題が任意の $\epsilon > 0$ に対して最適解の相対誤差 ϵ の近似アルゴリズムを持ちそうにないことを示している。V. Kann はさらに入力とするグラフの次数が定数で制限されているときも、MAX SNP-hardであることを示した [3]。しかし、その次数を制限している定数は25というかなり大きなものである。この論文では、同じ数の頂点と辺を持つ2つのグラフの類似性の判定に着目し、2つのグラフの次数が3でおさえられているときでも最大共通誘導部分グラフ問題は MAX SNP-hardであることを示した。また2つのグラフがともに連結で次数が3でおさえられているときも、MAX SNP-hardであることを示した。我々は、MAX SNP-hardであることを示すにあたって MAX SNP 完全として知られている最大3次元マッチング問題から L 還元を行う。一方、この最大共通誘導部分グラフ問題は NP 完全であることも知られている。さらに本論文ではこの問題について、与えられる2つのグラフの次数が制限され、かつ頂点数と辺の数が等しい特殊な場合についての NP 完全性について議論する。

2 最大共通部分グラフ問題

$G = (V, E)$ を無向グラフとする。 E の部分集合 E' が与えられたとき、 $V|_{E'} = \{v \in V \mid v \text{ を端点とする辺が } E' \text{ に存在する}\}$ とする。このとき、 $G|_{E'} = (V|_{E'}, E')$ を E' の辺誘導部分グラフ、略して、部分グラフという。また、 V の部分集合 V' に対して、 $E|_{V'} = \{\{u, v\} \in E \mid u, v \text{ は } V' \text{ に含まれる頂点}\}$ とする。このとき、 $G|_{V'} = (V', E|_{V'})$ を V' の頂点誘導部分グラフ、略して、誘導部分グラフという。また、 $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$ を無向グラフとする。無向グラフ $G = (V, E)$ が G_1 , G_2 の共通誘導部分グラフであるとは、 $V'_1 \subseteq V_1$ と $V'_2 \subseteq V_2$ が存在し $G_1|_{V'_1}$ と $G_2|_{V'_2}$ が共に G に同型であるときをいう。頂点数最大の G_1 と G_2 の共通誘導部分グラフを G_1 と G_2 の最大共通誘導部分グラフという。本論文では、次のような問題を考える。

定義 1. 次数制限最大共通誘導部分グラフ問題 (MAXIMUM BOUNDED COMMON INDUCED SUBGRAPH (MAX CIS-B))

入力: 無向グラフ $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$.

ただし、 G_1, G_2 の次数はともに高々定数 $B \geq 0$ である。

問題: G_1, G_2 の最大共通誘導部分グラフを求めよ.

定理 1. (V. Kann [3]) MAX CIS-B は $B \geq 25$ のとき MAX SNP-hard である.

MAX CIS-B の入力として与えられる 2 つのグラフがともに連結であるとき, 次数制限連結最大共通誘導部分グラフ問題 (MAXIMUM CONNECTED BOUNDED COMMON INDUCED SUBGRAPH (MAX CIS-CB)) と呼ぶ.

3 最大共通部分グラフ問題への L 還元

この章では, 入力される 2 つのグラフの頂点の次数が高々 3 のときに, 次数制限最大共通誘導部分グラフ問題が MAX SNP-hard であることを示す.

定義 2. Π_1 と Π_2 を 2 つの最適化問題とする. $f: \Pi_1 \rightarrow \Pi_2$ を, 多項式時間計算可能であるような関数とする. Π_1 の任意の入力を I とし, 入力 I の最適解のコストを $OPT(I)$ とする. Π_1 の任意の入力 I に対して以下のような正定数 α, β が存在するとき, f は L 還元であると呼ぶ.

1. $OPT(f(I)) \leq \alpha OPT(I)$.
2. 値 c_2 をもつ $f(I)$ のすべての解に対して, $|OPT(I) - c_1| \leq \beta |OPT(f(I)) - c_2|$ となる値 c_1 をもつ I の解を多項式時間で見つけることができる.

本章では, 次の問題から次数制限最大共通誘導部分グラフ問題への L 還元を考える.

最大 3 次元マッチング問題 (MAXIMUM BOUNDED THREE DIMENSIONAL MATCHING (MAX 3DM-B))

入力: $X \times Y \times Z$ の部分集合 M .

ただし, X, Y, Z は互いに交わりのない 3 つの集合である.

これらの集合の要素を,

$$\begin{aligned} X &= \{x_1, x_2, x_3, \dots, x_p\}, \\ Y &= \{y_1, y_2, y_3, \dots, y_p\}, \\ Z &= \{z_1, z_2, z_3, \dots, z_p\}, \\ M &= \{m_1, m_2, m_3, \dots, m_q\} \end{aligned}$$

とする.

また, M 中の X, Y, Z の各々の要素の出現回数は, 最大 $B \geq 0$ 回であるとする.

問題: 最大マッチングを求めよ. ただし, マッチングとは, X, Y, Z のどの要素も高々 1 度しかあらわれないような M の部分集合である.

我々は, MAX 3DM-B の入力 M に対して,

$$\begin{aligned} X_M &= \{x \in X \mid (x, y, z) \in M \text{ となるような } y \in Y \text{ と } z \in Z \text{ が存在する}\}, \\ Y_M &= \{y \in Y \mid (x, y, z) \in M \text{ となるような } x \in X \text{ と } z \in Z \text{ が存在する}\}, \\ Z_M &= \{z \in Z \mid (x, y, z) \in M \text{ となるような } x \in X \text{ と } y \in Y \text{ が存在する}\} \end{aligned}$$

とするとき,

$$\begin{aligned} X_M &= \{x_1, x_2, x_3, \dots, x_{|X_M|}\}, \\ Y_M &= \{y_1, y_2, y_3, \dots, y_{|Y_M|}\}, \\ Z_M &= \{z_1, z_2, z_3, \dots, z_{|Z_M|}\} \end{aligned}$$

と仮定する. すなわち, X, Y, Z の要素で実際に M に出現するのは, インデックスの小さい方からそれぞれ $|X_M|, |Y_M|, |Z_M|$ 番目までである. また, このとき,

$$p = |X_M| \geq |Y_M| \geq |Z_M|$$

と仮定する. このように仮定しても一般性を失わない.

定理 2. (V. Kann [2]) MAX 3DM-B は $B \geq 3$ のとき MAX SNP 完全である.

定理 3. MAX CIS-B は $B \geq 3$ のとき MAX SNP-hard である.

証明. MAX 3DM-B からの L 還元を示す. ただし, MAX 3DM-B の入力 M 中の X, Y, Z の各々の要素の出現回数は高々 3 回とする. 一般に問題の入力を I で表す. MAX 3DM-B の場合, 入力 I は $M \subseteq X \times Y \times Z$ のことである. ただし, M は $p \leq q$ を考えれば十分であり, かつ $q \leq 3p$ である. MAX 3DM-B の任意の入力から MAX CIS-B のそれに対応する入力へ還元する多項式時間アルゴリズムを f とし, 3 つの頂点からなる完全グラフを K_3 と表すことにする. MAX CIS-B の一方のグラフである $G_1 = (V_1, E_1)$ の構成法を以下に示す. 添字 i ($1 \leq i \leq |Z_M|$) に対して, 図 1 に示すように $G[i]$ をつくる. $G[i]$ は,

$$\begin{aligned} V[i] &= \{v_j^k[i] \mid 1 \leq j \leq 3, 1 \leq k \leq 3\} \cup \{a[i]\}, \\ E[i] &= \{\{v_j^1[i], v_j^2[i]\}, \{v_j^2[i], v_j^3[i]\}, \{v_j^3[i], v_j^1[i]\} \mid 1 \leq j \leq 3\} \\ &\quad \cup \{\{a[i], v_j^1[i]\} \mid 1 \leq j \leq 3\} \end{aligned}$$

からなる (図 1 参照).

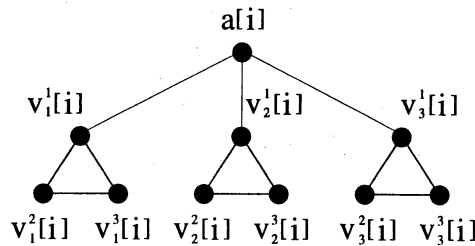


図 1: $G[i]$

このとき, $G_1 = (V_1, E_1)$ は以下のように表される.

$$\begin{aligned} V_1 &= \bigcup_{i=1}^{|Z_M|} V[i]. \\ E_1 &= \bigcup_{i=1}^{|Z_M|} E[i]. \end{aligned}$$

以上が G_1 の構成法である (図 2 参照).

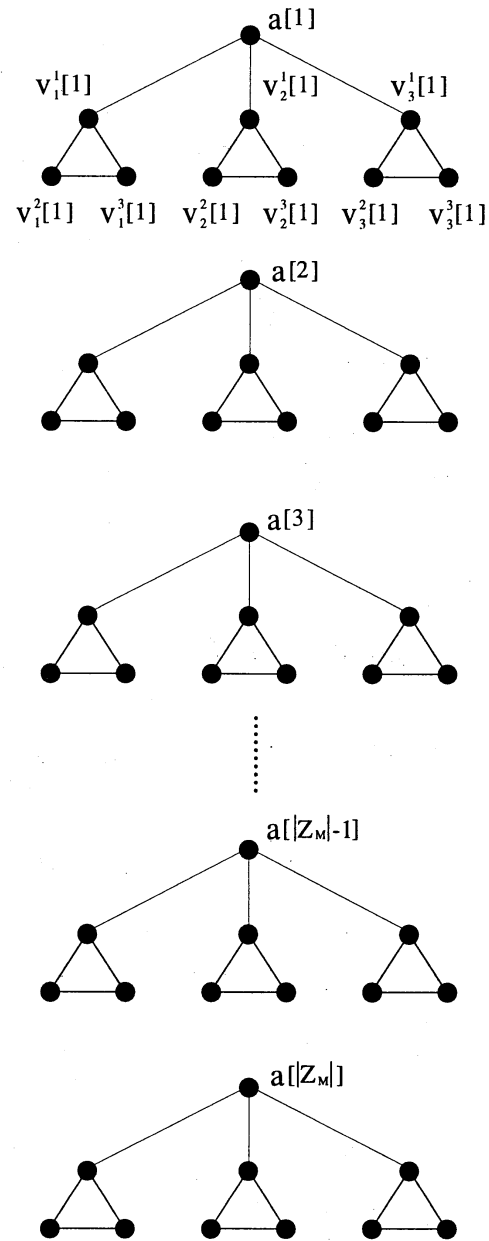


図 2: MAX 3DM のある入力に対応するグラフ G_1

つぎに, $G_2 = (V_2, E_2)$ の構成法を示す. X_M の i 番目の要素 x_i を表すために, 図 3 に示すように, $G_x[i]$ ($1 \leq i \leq |X_M|$) をつくる. $G_x[i]$ は,

$$\begin{aligned} V_x[i] &= \{u_x^k[i] \mid 1 \leq k \leq 3\}, \\ E_x[i] &= \{\{u_x^1[i], u_x^2[i]\}, \{u_x^2[i], u_x^3[i]\}, \{u_x^3[i], u_x^1[i]\}\} \end{aligned}$$

からなる (図 3 参照).

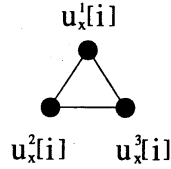


図 3: $G_x[i]$

Y_M, Z_M についても同様に, それぞれ $G_y[i], G_z[i]$ をつくる. $G_y[i]$ は,

$$\begin{aligned} V_y[i] &= \{u_y^k[i] \mid 1 \leq k \leq 3\}, \\ E_y[i] &= \{\{u_y^1[i], u_y^2[i]\}, \{u_y^2[i], u_y^3[i]\}, \{u_y^3[i], u_y^1[i]\}\} \end{aligned}$$

からなる. $G_z[i]$ は,

$$\begin{aligned} V_z[i] &= \{u_z^k[i] \mid 1 \leq k \leq 3\}, \\ E_z[i] &= \{\{u_z^1[i], u_z^2[i]\}, \{u_z^2[i], u_z^3[i]\}, \{u_z^3[i], u_z^1[i]\}\} \end{aligned}$$

からなる.

さらに, $V_M = \{c_{ijk} \mid (x_i, y_j, z_k) \in M\}$ とし, x_i が M に $B_{x_i}(\leq 3)$ 回出現するとき, x_i が出現する M の元を

$$m_{x_i}^\ell = (x_i, y_{j_\ell}, z_{k_\ell}) \quad (1 \leq \ell \leq B_{x_i})$$

とする. このとき,

$$E'_x[i] = \{\{c_{i,j_\ell,k_\ell}, u_x^\ell[i]\} \mid 1 \leq \ell \leq B_{x_i}\}$$

とする. 同様に, y_j と z_k が M にそれぞれ $B_{y_j}, B_{z_k}(\leq 3)$ 回出現するとし, y_j, z_k が出現する M のそれぞれの元を

$$\begin{aligned} m_{y_j}^\ell &= (x_{i_\ell}, y_j, z_{k_\ell}) & (1 \leq \ell \leq B_{y_j}), \\ m_{z_k}^\ell &= (x_{i_\ell}, y_{j_\ell}, z_k) & (1 \leq \ell \leq B_{z_k}) \end{aligned}$$

とする. このとき,

$$\begin{aligned} E'_y[j] &= \{\{c_{i_\ell,j,k_\ell}, u_y^\ell[j]\} \mid 1 \leq \ell \leq B_{y_j}\}, \\ E'_z[k] &= \{\{c_{i_\ell,j_\ell,k}, u_z^\ell[k]\} \mid 1 \leq \ell \leq B_{z_k}\} \end{aligned}$$

とする. このとき, $G_2 = (V_2, E_2)$ は以下のように表される.

$$\begin{aligned} V_2 &= V_M \cup \left(\bigcup_{(x_i, y_j, z_k) \in M} (V_x[i] \cup V_y[j] \cup V_z[k]) \right), \\ E_2 &= \bigcup_{(x_i, y_j, z_k) \in M} (E_x[i] \cup E_y[j] \cup E_z[k] \cup E'_x[i] \cup E'_y[j] \cup E'_z[k]). \end{aligned}$$

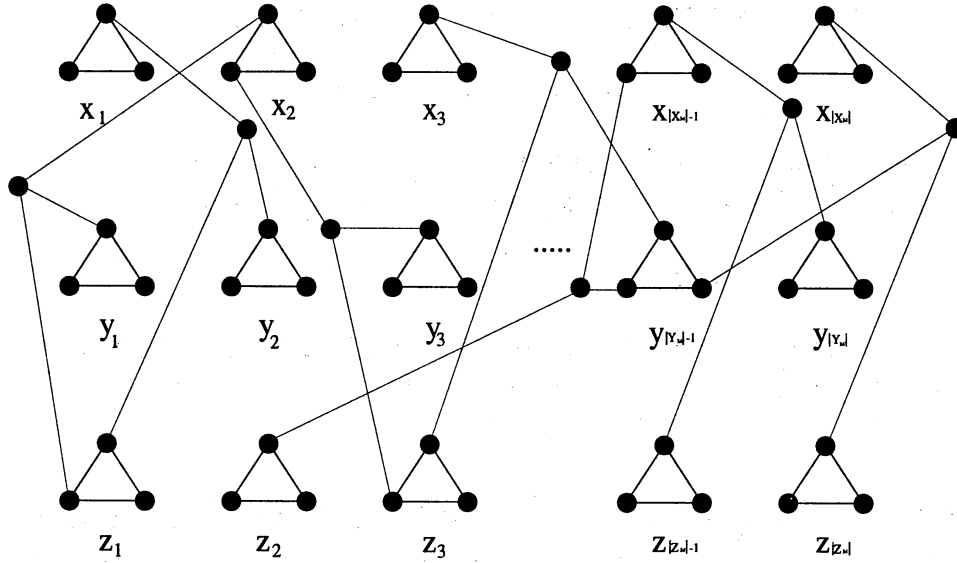


図 4: MAX 3DM のある入力に対応するグラフ G_2

以上が G_2 の構成法である (図 4 参照).

G_1 と G_2 は, 以下の性質をもっている.

性質 1. G_2 の中で, K_3 と同型なグラフは,

$$G_x[i] \ (1 \leq i \leq |X_M|), \ G_y[i] \ (1 \leq i \leq |Y_M|), \ G_z[i] \ (1 \leq i \leq |Z_M|)$$

のみである.

性質 2. 任意の $G[i] \ (1 \leq i \leq |Z_M|)$ と同型な G_2 の誘導部分グラフは,

$$G_2|_{V_x[r] \cup V_y[s] \cup V_z[t] \cup \{c_{rat}\}} \quad ((x_r, y_s, z_t) \in M)$$

のみである.

性質 1 および性質 2 より, MAX 3DM-B の最適解のコストを $OPT(I)$ としたとき, 以下の頂点集合 U の誘導部分グラフ $G|_U$ は, MAX CIS-B の入力 $f(I)$ に対する最適解の一つである.

$$U = \left(\bigcup_{i=1}^{OPT(I)} V[i] \right) \cup \left(\bigcup_{i=OPT(I)+1}^{|Z_M|} (V[i] - \{a[i]\}) \right).$$

これは,

$$G[i] \ (1 \leq i \leq OPT(I)), \ G[i]|_{\{v_j^k[i] | 1 \leq j \leq 3, 1 \leq k \leq 3\}} \ (OPT(I) + 1 \leq i \leq |Z_M|)$$

からなる G_1 の誘導部分グラフに同型なグラフである. これより,

$$\begin{aligned} OPT(f(I)) &= 10 \cdot OPT(I) + 9(|Z_M| - OPT(I)) \\ &\leq 27B \cdot OPT(I) + OPT(I) \\ &= (27B + 1) \cdot OPT(I) \end{aligned}$$

である.

さらに、コスト c_2 をもつ MAX CIS-B の入力 $f(I)$ の任意の解に対して、

$$\{c_{i,j,k}\} \cup V_x[i] \cup V_y[j] \cup V_z[k]$$

がその解に入っているときに限り、 (x_i, y_j, z_k) を MAX 3DM の入力 I のマッチングの解の一つとして選ぶものとする。そのようにして得られるマッチングの解がコスト c_1 をもつとすれば、

$$|OPT(I) - c_1| \leq |OPT(f(I)) - c_2|$$

となるコスト c_1 をもつ MAX 3DM の入力 I の解を多項式時間で見つけることができる。従って、次数制限最大共通誘導部分グラフ問題は、最大 3 次元マッチング問題から L 還元可能である。

□

入力される 2 つのグラフの頂点数と辺の数がともに等しい MAX CIS-B および MAX CIS-CB、さらに入力される 2 つのグラフの頂点数と辺の数がともに等しい MAX CIS-CB の 3 つの問題についても、MAX 3DM-B から L 還元することによって、以下の定理が証明される。上で示した証明とくらべると以下の証明で作られるグラフは、条件が加えられているほど複雑になっている。

定理 4. $B \geq 3$ であり、かつ入力される 2 つのグラフの頂点数と辺の数がともに等しいとき、MAX CIS-B は MAX SNP-hard である。

定理 5. MAX CIS-CB は、 $B \geq 3$ のとき MAX SNP-hard である。

定理 6. $B \geq 3$ であり、かつ入力される 2 つのグラフの頂点数と辺の数がともに等しいとき、MAX CIS-CB は MAX SNP-hard である。

4 最大共通誘導部分グラフ決定問題の NP 完全性について

この章では、入力される 2 つのグラフの頂点の次数が高々 2 のときに、次数制限最大共通誘導部分グラフ問題 (決定問題) が NP 完全であることを示す。次数制限最大共通誘導部分グラフ問題の NP 完全性を議論するために、決定問題として次のように定義する。

次数制限最大共通誘導部分グラフ問題 (決定問題) (MAXIMUM BOUNDED COMMON INDUCED SUBGRAPH (MAX CIS-B))

入力: 無向グラフ $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$, 正定数 K .

ただし、 G_1, G_2 の次数はともに最大 $B \geq 0$ であるとする。

問題: 頂点数が K 以上の G_1, G_2 の共通誘導部分グラフが存在するか。

本章では、次の問題から MAX CIS-B への還元を考える。

3 分割問題 (3-PARTITION)

入力: 有限集合 A , 正定数 L , A から正定数への関数 s .

ただし、任意の $s(a)$ に対して、 $\frac{L}{4} < s(a) < \frac{L}{2}$ である。また、 $\sum_{a \in A} s(a) = mL$.

問題: A を互いに交わりのない m 個の集合 $S_1, S_2, \dots, S_m$ に分けることができるか。

ただし、 $1 \leq i \leq m$ に対して、 $\sum_{a \in S_i} s(a) = L$

定理 7. MAX CIS-B は $B \geq 2$ のとき NP 完全である.

証明. 3-PARTITION からの還元である. □

入力される 2 つのグラフの頂点数と辺の数がともに等しい MAX CIS-B についても, 3-PARTITION から還元することによって, 以下の定理が証明される.

定理 8. $B \geq 2$ であり, かつ入力される 2 つのグラフの頂点数と辺の数がともに等しいとき, MAX CIS-B は NP 完全である.

5 結論

本論文では, 小さな次数に制限した最大共通誘導部分グラフ問題の計算量について考察した. 結果をまとめると表 1 のようになる. 入力として与えられる 2 つのグラフの次数がともに高々 2 であるときに, PTAS があるか, それとも MAX SNP-hard であるか, また, 入力として与えられる 2 つのグラフの次数がともに高々 3 であるときに, MAX SNP 完全かという課題が残されている.

入力される 2 つのグラフの 頂点の次数	MAX CIS-B	MAX CIS-B $ V_1 = V_2 $ $ E_1 = E_2 $	MAX CIS-CB	MAX CIS-CB $ V_1 = V_2 $ $ E_1 = E_2 $
高々 2	NP 完全	NP 完全	多項式時間	G_1, G_2 は同型
高々 3	MAX SNP-hard	MAX SNP-hard	MAX SNP-hard	MAX SNP-hard

表 1: まとめ

参考文献

- [1] T. Jiang, L. Wang, and K. Zhang. Aligment of trees - an alternative to tree edit. *Theoretical Computer Science*, pp. 137–148, 1995.
- [2] V. Kann. Maximum bounded 3-dimensional matching is max snp-complete. *Imformation Processing Letters*, Vol. 37, pp. 27–35, 1991.
- [3] V. Kann. On the approximability of the maximum common subgraph problem. In *STACS 92*, pp. 377–388, 1992.
- [4] E. Kubika, G. Kubicki, and I. Valalis. Using graph distance in object recognition. In *ACM 90*, pp. 43–48.
- [5] B. Zelinka. Distances between graphs. In *4th Czechoslovakian Symposium on Combinatorics, Graphs and Complexity*, pp. 355–361, 1992.