

## ハイパーグラフを対象とした最大クリーク抽出の分枝限定アルゴリズム

須谷 洋一<sup>†</sup> 富田 悦次<sup>†</sup> 小林 聡<sup>†</sup>

あらまし 本稿では、一様なハイパーグラフから最大クリークを抽出する効率の良い分枝限定アルゴリズムを提案する。また、この問題の応用例として挙げられる DNA 配列設計に関する問題を含め、いくつかの問題に対しアルゴリズムの実験的評価を行った。その結果、本稿で提案するアルゴリズムの有効性を確認した。

### A Branch-and-Bound Algorithm for Finding a Maximum Clique in a Uniform Hypergraph

Yoichi SUTANI<sup>†</sup> Etsuji TOMITA<sup>†</sup> Satoshi KOBAYASHI<sup>†</sup>

**Abstract.** In this article, we present an efficient branch-and-bound algorithm for finding a maximum clique in a uniform hypergraph. We evaluated this algorithm by computational experiments on some problems which include a problem for designing DNA sequences for biomolecular computation. From these experimental results, we confirmed the effectiveness of our algorithm.

## 1 はじめに

無向グラフから最大クリークを見つける問題は、多くの数理論題に応用できることから重要であり、これまで数多くの研究が行われてきた [1, 2, 3]。一方で近年、通常のグラフでなく、その概念を拡張したハイパーグラフを対象とした最大クリーク問題についての応用 [4, 5, 6] がみられるようになった。この問題は NP 困難のクラスに属す問題であり、また、通常のグラフのそれよりも難しい問題であるが、昨今の計算機の高性能化により、ある程度の規模ならば扱えるため、今後はさらに利用される場面が増える可能性がある。しかし、現時点ではこの問題を効率よく解くアルゴリズムについての研究はそれほど多くなく、有効なアルゴリズムの開発は重要な問題である。

そこで、本稿では先行研究となる文献 [6] 等に基づき、一様なハイパーグラフから最大クリークを抽出する分枝限定アルゴリズムを提案する。また、応用例として挙げられる DNA 配列設計問題を含め、いくつかの問題に対しアルゴリズムの実験的評価を行った。

## 2 諸定義

### (a) ハイパーグラフ

ハイパーグラフ  $H = (V, E)$  は節点集合  $V$  とハイパー辺集合  $E$  からなる。ハイパー辺  $e \in E$  は空でない  $V$  の部分集合である。本稿では、ハイパー辺  $e$  が向きを持たない、無向ハイパーグラフを扱う。また、節点集合は順序付きであるものとし、 $V$  の  $i$  番目の要素を  $V[i]$  として表す。

### (b) ランク

$V$  の部分集合  $X$  に対し、 $r(X) = \max\{|X \cap e| \mid e \in E\}$  を  $X$  のランクという。 $|X|$  は、 $X$  の要素数を表すものとする。また、 $\forall e \in E$  において  $|e| = k$  であるグラフを  $k$ -一様グラフという。2-一様グラフは通常のグラフである。

### (c) クリーク

$V$  の部分集合  $C$  が、 $|C| < r$  または、 $|C| \geq r$  かつ  $C$  の要素数  $r$  の任意の部分集合が少なくとも 1 つの  $H$  の辺に含まれるとき、 $C$  をランク  $r$  のクリークであるという。 $H$  の中で最も節点数の大きいランク  $r$  のクリークの節点数、すなわち、最大クリークのサイズを  $\omega_r$  と表す。

<sup>†</sup>電気通信大学大学院 電気通信学研究科  
Graduate School of Electro-Communications, The University of Electro-Communications

### 3 分枝限定アルゴリズム

本稿では、ランク  $r$  の一様なハイパーグラフ  $H$  が与えられたとき、そこからランク  $r$  の最大クリークを1つ見つけるアルゴリズムを対象とする。

#### 3.1 基本アルゴリズム

アルゴリズムの基本としては、通常のグラフに対するいくつかの最大クリーク抽出アルゴリズムでも用いられている、単純な深さ優先探索を用いる手法が提案されている [4, 6]。具体的には、ある時点で保持しているクリーク  $Q$  (初期状態は空集合を与える) について、サイズを1つずつ大きくするよう分枝操作により部分問題を作成し、深さ優先でクリークの探索を進めていく。

探索の全過程は、全節点集合  $V$  を根、各部分問題以降でクリークの要素となり得る節点集合を  $R$  としたとき、この節点集合  $R$  について親子関係にあるものを枝で結んだ探索木として表現することができる。この探索木における枝を分枝、その総数を分枝数という。

アルゴリズムの効率化は、分枝操作により作成された無駄な部分問題を限定操作によって省くことにより行う。 $Q_{max}$  をこれまでに見つかっている最大のクリークとしたとき、ごく基本的な限定操作として以下の式を利用できる。

$$|Q| + |R| \leq |Q_{max}| \quad \dots (1)$$

この条件を満たすとき、その部分問題における以降の探索を不要とみなし、探索を省くことが可能となる。

探索の効率化のため、通常のグラフの場合にも有効な手法として、探索開始前に節点のソートを行う操作がある。この操作が、ハイパーグラフにも有効であることが実験的に示されている [6]。具体的には、探索の前処理として節点を次数の降順にソートし、次数の小さい節点から探索を行うようにする。節点の次数は、その節点を含むハイパー辺の個数と定義される。

以上のようなアルゴリズムを本稿では、基本アルゴリズムと呼ぶ。

#### 3.2 節点彩色に基づく上界 [6]

通常のグラフに対する最大クリーク問題においては、節点彩色によりグラフ中のクリークサイズの上界を得ることができ、これを限定操作に利用できる [2, 3]。ハイパーグラフを対象とする問題においても、この手法を拡張した分枝限定手法が提案されている [6]。彩色の条件は以下の通りである。

#### 彩色条件

任意のハイパー辺  $e \in E$  を構成するランク  $r$  個の節点を最低2色以上で彩色する。

この彩色条件において、 $C$  をランク  $r$  のクリーク、彩色数を  $\chi_r(C)$  とするとき、 $|C| \leq (r-1) \times \chi_r(C)$  の関係が成り立つ [6]。この関係をクリークサイズの上界として利用できる。

上記の彩色条件を満たすように、彩色番号“1”から順に逐次的な番号付けを行い、全節点の番号付け終了後、この番号順に節点を並べ替える。この部分問題において、彩色番号の大きいものから探索を行えば、ある節点  $p$  の彩色番号を  $Col[p]$  とするとき、クリークサイズの上界を  $No[p] = Col[p] \times (r-1)$  とできる。よって、以下の条件を限定操作に利用できる。

$$|Q| + No[p] \leq |Q_{max}| \quad \dots (2)$$

以上の節点彩色操作に掛かる計算量は  $O(n^r)$  である。

#### 3.3 提案手法

ここで、3.2 節の節点彩色とは異なる番号付けを基にして上界を得る手法を提案する。前述の節点彩色による手法では、以後クリークの要素となり得る節点集合  $R$  に関してのみ考慮して、クリークサイズの上界を得ていた。一方、本手法においては、 $R$  に加え、拡大中のクリーク  $Q$  も含め、上界を求める操作に用いる。

#### 上界となる番号付けの条件

ある時点において、深さ  $l$  の部分問題を探索途中 (このとき、 $|Q| = l$ ) であるとする。また、 $Q_{sub} \subseteq Q$  かつ  $|Q_{sub}| = r-2$  とする。このとき、 $\forall v_a, v_b \in R$  (ただし、 $v_a \neq v_b$ ) が、任意の  $Q_{sub}$  に対して、 $Q_{sub} \cup \{v_a\} \cup \{v_b\} \in E$  を満たすとき、 $v_a$  と  $v_b$  に異なる番号を付与する。

$Q_{sub} \cup \{v_a\} \cup \{v_b\} \notin E$  であれば、明らかに節点  $v_a$  と節点  $v_b$  の2節点を含むクリークがその深さ以降の探索において発見され得ない。そのため、上記条件を満たすように全節点番号を付与すると、その最大番号は節点集合  $R$  から今後クリーク  $Q$  の要素として追加される節点数の上界となる。よって、この条件を満たすように、前述の彩色による番号付けと同様の手順によって、節点の番号付けと並べ替えを行うことで、ある節点  $p$  に番号  $No[p]$  が付与されたとき、(2) 式を用いて限定操作が可能となる (図1)。

ただし、 $Q_{sub}$  の定義より、 $l \leq r-2$  のときはこの手法を利用できない。そのため本稿では、 $l \leq r-2$  を

満たす深さ  $l$  の部分問題においては、単純に  $1, 2, \dots, |R|$  と順次番号を付けて終了することにする。

以上の上界を求める操作に掛かる計算量は、彩色を用いる手法と同じ  $O(n^r)$  である。

- 
- (1)  $i \leftarrow 1, \max \leftarrow 1, R_1 \leftarrow \phi$  とする
  - (2)  $p \leftarrow R[i], k \leftarrow 1$  とする
  - (3)  $\forall Q_{sub}, \forall v \in R_k$  に対し、 $Q_{sub} \cup \{v\} \cup \{p\} \in E$  なら  $k \leftarrow k + 1$  とし、再度 (3)  $\rightarrow$  (ただし、 $k > \max$  のとき  $\max \leftarrow \max + 1, R_{\max} \leftarrow \phi$  とする)。そうでないなら、(4)  $\rightarrow$
  - (4)  $R_k \leftarrow R_k \cup \{p\}, \text{No}[p] \leftarrow k$  とする
  - (5)  $i \leftarrow i + 1$  とする
  - (6)  $i \leq |R|$  なら (2)  $\rightarrow$
  - (7)  $R$  に  $R_1$  から  $R_{\max}$  の全ての要素を順に格納する
- 

図 1: 提案手法による番号付け手続き

## 4 計算機実験

### 4.1 実験対象および環境

分枝限定手法の違いによるアルゴリズムの効果を評価するため、以下の各アルゴリズムを実装し、同じグラフに対する各アルゴリズムの実行時間および分枝数を測定することで比較実験を行った。

アルゴリズムは、基本アルゴリズム (basic) と、これに彩色による限定操作を加えたもの (coloring)、提案手法による限定操作を加えたもの (numbering) の 3 つである。実験の対象としたグラフの詳細を以下に示す。なお、本稿では、ランク 3 のハイパーグラフのみ実験対象とする。

#### ランダムグラフ

節点数  $n$ 、ランク  $r$  の一様ハイパーグラフの場合、ハイパー辺が存在し得る  ${}_n C_r$  組の節点の組合せに対して、辺の存在確率  $p$  で辺の存在の有無を決定し、グラフを生成する。

#### ランク $r$ の G10 グラフ

節点数  $n/10$  箇所に辺が密在するように作成したグラフであり、通常のグラフにおける G10 グラフ [1] を拡張したグラフである。グラフの作成法としては、まず、ハイパー辺の密在箇所を中心の位置を、節点の並びを等間隔に分けるようにして  $n/10$  個求め、これを中心とした標準偏差 1.0 の正規分布に従って決める。このような密在箇所を中心に、標準偏差 2.0 の正規分布

に従って各ハイパー辺の 1 つの節点を選び、残るランク  $r-1$  個の節点をランダムに選んで、それをハイパー辺とする。これを節点数  $n$  に対し、辺の数が辺密度  $d$  となるまで繰り返す。なお、辺密度  $d$  は  $d = |E|/|V|^r$  として定義される。

### DNA 配列設計問題から還元したグラフ

本稿で扱う DNA 配列設計問題とは、与えられた DNA 配列集合  $S$  に対して、想定外の構造を形成しない要素数が最大の  $S$  の部分集合を求める問題である。

著者の一人である小林らは、この問題をハイパーグラフから最大クリークを抽出する問題に還元する方法を提案した [4, 5]。本稿では、そのデータを実験対象として用いた。具体的には、DNA 配列集合  $S$  を節点集合  $V$  とし、それぞれの DNA 配列を頂点  $v \in V$  とする。このとき、任意の要素数 3 の頂点集合に対して自由エネルギー  $F$  が与えられている。これは、各頂点に対応する 3 本の配列が想定外の構造を形成する度合い ( $F$  が小さいほど想定外の構造を形成しやすい) を表している。この  $F$  が予め与えられた閾値  $T$  よりも大きいときに、配列が想定外の構造を形成しないと考へ、その 3 頂点からなるハイパー辺が存在するものとする。閾値  $T$  は実験条件によって変わるため、応用上はさまざまな  $T$  の値に対して、ハイパーグラフを形成し、最大クリークを求めることが必要となる。詳細については、文献 [4, 5] を参照されたい。

実験は以下の環境を用いて行った。

- ・使用計算機
  - CPU : Pentium4 3.6GHz
  - Memory : 2.0GB
  - OS : Linux
- ・使用言語
  - C 言語 コンパイラ gcc -O2 (ver4.1.0)

### 4.2 実験結果および考察

図 2, 3 は、節点数 150、辺の存在確率をパラメータとしたランダムグラフおよび辺密度をパラメータとした G10 グラフの結果である。それぞれ実行時間と分枝数をあわせて記載している。これらの結果より、提案手法 (numbering) が従来手法 (basic または coloring) と比較して、短い時間で解を得ていることが分かる。特に、提案手法は実行時間の増大する辺密度が高いグラフに対して効果が顕著である。これは図 2, 3 から分かる通り、提案手法を用いたアルゴリズムは分枝数が他の手法によるものより大幅に削減できているため

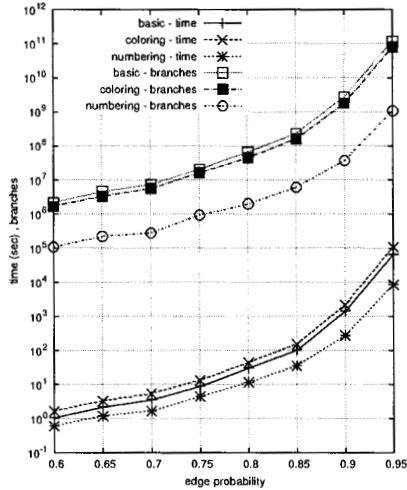


図 2: 節点数 150 のランダムグラフ

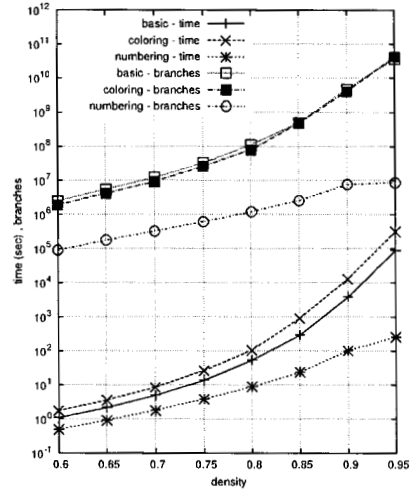


図 3: 節点数 150 の G10 グラフ

ある。分枝数の削減の割合に対し、実行時間がそれほど減少していないのは、提案手法によって上界を得る操作が、節点彩色による処理よりも計算手数が多いためと考えられる。

また、結果より、提案手法はランダムグラフに比べ G10 グラフに対しより優位である。このことから、提案手法による限定操作は、辺の存在箇所に偏りがあるグラフに対してより効果的であるといえる。現実の問題をモデル化した場合、完全にランダムなグラフとなる場合は非常に稀だと考えられることから、提案手法の有効性がうかがえる。

一方、DNA 配列設計問題に対する結果例を表 1 に示す。この結果からも、提案手法の効果が確認できる。こちらは、彩色による効果が、ランダムグラフや G10 グラフより現れているが、提案手法の効果はそれをさらに上回っている。

表 1: DNA 配列設計問題  $|S|=200$  の場合の例 - 実行時間 (sec)

$T$	$\omega_3$	basic	coloring	numbering
-8	8	0.02	0.02	0.02
-9	11	0.13	0.10	0.05
-10	15	3.3	1.9	0.6
-11	21	284	95	16
-12	31	62,546	11,164	840

## 5 おわりに

本稿では、一様なハイパーグラフから最大クリークを抽出する効率の良いアルゴリズムを提案した。ま

た、本稿で提案した分枝限定手法を用いたアルゴリズムが従来の手法によるものと比較して、応用問題を含むデータに対する実験において高速に解を得られることが分かった。なお、本稿ではランク 3 のグラフに対する実験結果のみ掲載しているが、ランク 4 に対しても各分枝限定手法の違いによるアルゴリズムの効果が同傾向にあることを確認している。

## 謝辞

本研究は科学研究費補助金基盤研究 (B), (C) の支援を受けている。

## 参考文献

- [1] 富田悦次, 藤井利昭, “最大クリーク抽出の効率化手法とその実験的評価,” 電子通信学会論文誌 (D), vol. J68-D, no.3, pp.221-228 (1985).
- [2] E. Tomita, T. Kameda, “An efficient branch-and-bound algorithm for finding a maximum clique with computational experiments,” J. Global Optim., 37, pp.95-111 (2007).
- [3] E. Tomita, Y. Sutani, T. Higashi, “A more efficient algorithm for finding a maximum clique with an improved approximate coloring,” Proc. PDPTA 2007, pp.719-725 (2007).
- [4] 奥田謙平, 小林聡, “バルジ・内部ループを形成しない DNA 配列セットの設計,” 情報処理学会研究報告, 2002-MPS-42, pp.99-102 (2002).
- [5] 小林聡, 近藤朋大, 奥田謙平, 富田悦次, “大局的に構造を取らない配列集合の抽出方法,” 電子情報通信学会技術研究報告, COMP2003-16, pp.1-6 (2003).
- [6] 奥田謙平, “超グラフの最大クリーク抽出アルゴリズムとその応用に関する研究,” 電気通信大学大学院情報工学専攻修士論文 (2004).