

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 6 月 3 日現在

機関番号：12612

研究種目：基盤研究(C)

研究期間：2010～2012

課題番号：22500009

研究課題名（和文）効率的な最大クリーク抽出アルゴリズムの開発と理論的・実験的評価および応用

研究課題名（英文）Development of efficient algorithms for finding a maximum clique with theoretical and experimental evaluations and their applications

研究代表者

富田 悦次 (TOMITA ETSUJI)

電気通信大学・名誉教授

研究者番号：40016598

研究成果の概要（和文）：一般グラフの最大クリーク問題を多項式時間的可解とする条件とアルゴリズムを開発し、評価結果を示した。本アルゴリズムは、当条件にかかわらず、解自体は常に正しく出力する。また、一般の場合に対してこれまで新たに開発した最大クリーク抽出アルゴリズムの動作を内部にわたって詳細に実験的評価を行い、その高効率性を明らかにした。更に、これらのアルゴリズムを基として、バイオインフォマティクス等の問題に対しての有効な応用を示した。

研究成果の概要（英文）：We developed some sufficient conditions and algorithms for arbitrary graphs by which the maximum clique problem can be solved in polynomial time. These algorithms can find an exact maximum clique in any arbitrary graph without any condition. We also confirmed experimentally that our newly developed another maximum-clique-finding algorithm works very efficiently. These algorithms were effectively applied for some problems as in bioinformatics.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	1,100,000	330,000	1,430,000
2011年度	1,000,000	300,000	1,300,000
2012年度	1,100,000	330,000	1,430,000
年度			
年度			
総計	3,200,000	960,000	4,160,000

研究分野：総合領域

科研費の分科・細目：情報学・情報学基礎

キーワード：最大クリーク，極大クリーク，NP完全問題，時間計算量，多項式時間的可解性，分枝限定法，バイオインフォマティクス

## 1. 研究開始当初の背景

最大クリーク問題は典型的な NP 完全問題という難問であるが、バイオインフォマティクス、画像処理、符号理論等々多くの問題に応用出来ることが分かってきて、そのために数多くの最大クリーク抽出アルゴリズムが我々の研究グループを初めとして提案さ

れてきた。ここで、応用が広がるに従って、より大規模なグラフに対しても最大クリーク抽出を有効に実行できることが要請されている。しかし、大規模グラフに対するアルゴリズムの高速化は未だ十分とは言えず、一層の高効率化が強く求められてきた。

最大クリーク問題を一般化した、極大クリ

ーク全列挙問題に関しても同様の状況であった。その中で、我々のグループがこれまでに開発した極大クリーク全列挙アルゴリズム CLIQUES (Tomita et al., Theoretical Computer Science 2006) は、ある評価において理論的最適 (optimal) 性を達成したアルゴリズムであるが、その一層の発展や応用が期待されていた。

## 2. 研究の目的

これまでに開発してきている最大クリーク抽出アルゴリズムおよび極大クリーク全列挙アルゴリズムに対して詳細な解析、評価を行い、その上に立って、より効率化したアルゴリズムを開発し、理論的、実験的評価によりその高効率性を明らかにする。

また、それらのアルゴリズムを基として、具体的に新たな応用問題への適用の有効性を示す。

## 3. 研究の方法

アルゴリズムの高効率性を理論保証するためには、そのアルゴリズムが多項式時間計算量となることを示すことが一番明確である。ところが、最大クリーク問題は典型的な NP 完全問題であり、最大クリーク問題が一般的に多項式時間計算量で解けることはほぼ不可能であろうと予測されている。従って、先ず最初に、いかなる緩やかな条件付けをすれば最大クリーク問題が多項式時間計算量で解けるようになるかを、明確な定量的条件とアルゴリズムの下に明らかにする。ここにおいて、条件付けは出来るだけ単純明快な解析によって多項式時間計算量性が証明できる場合から始め、逐次条件をより緩和する方向で進めていく。

次に、最大クリーク抽出アルゴリズムの実験的効率性をより高めるために、効率的アルゴリズムの各部動作の効果および全体を実験的に徹底的に検証評価する。

その結果から、優れた評価確認の得られた最大クリーク抽出および極大クリーク全列挙アルゴリズムを基として、具体的な応用に対する有効性を明らかにする。

## 4. 研究成果

[1] 最大クリーク問題の多項式時間的可解性を保証する条件、アルゴリズムの開発と理論的評価

1) 節点数  $n$  のグラフ中の節点の最大次数  $\Delta \leq 2.495d \lg n$  ( $d \geq 1$  は任意の定数) である時、最大クリーク問題を  $O(n^{2+d})$  なる多項式時間で可解とした。

これは、節点数  $n$  のグラフ中の極大クリークを  $O(3^{n/3})$  時間で全列挙するアルゴリズム CLIQUES を基とし、列挙対象を最大なクリーク 1 個だけと限定することにより余分な

探索を削除して効率化することにより得た結果で、アルゴリズム、計算量解析共に単純、簡潔なものとして与えた。

2) 節点数  $n$  のグラフ中の節点の最大次数  $\Delta \leq 2.6135d \lg n$  ( $d \geq 1$  は任意の定数) である時、最大クリーク問題を  $n^{1+d}$  なる多項式時間で可解とした。

前記アルゴリズムと計算量解析を更に詳細化することにより、次数に対する定数部分をより大きくし、多項式時間計算量における指数部分を 1 だけ小さくした。

3) 節点数  $n$  のグラフ中の節点の最大次数  $\Delta \leq 2.773d \lg n$  ( $d \geq 0$  は任意の定数) である時、最大クリーク問題を  $n^{1+d}$  なる多項式時間で可解とした。

更にアルゴリズムと計算量解析を詳細化し、かつグラフを表すデータ構造として隣接リストを用いることにより、前記の様に、定数  $d$  に関する条件を  $d \geq 0$  へと拡張した。これにより、対象グラフが非常に疎となるに従い、多項式時間計算量は特に線形時間に近づくものとした。

4) 節点数  $n$  のグラフ中の節点の最大次数  $\Delta \leq 2.994d \lg n$  ( $d \geq 0$  は任意の定数) である時、最大クリーク問題を  $n^{1+d}$  なる多項式時間で可解とした。更に、節点次数に対する条件は、必ずしも全節点に対してではなく、隣接 2 節点において次数の小さい方に対してだけ上記の次数条件が成立していれば、次数の大きい方の隣接節点の次数は任意であっても同様の多項式時間計算量性を保証出来ることを明らかにした。

これまでの結果をより詳細化することにより得られた本結果は、多項式時間計算量性を保証出来る条件をより緩やかなものとした。

上記のいずれのアルゴリズムにおいても、多項式時間的可解性のため条件が成立していない場合でも、多項式時間性は保証されないが、最大クリーク自体は常に正しく抽出できる。

## [2] 最大クリーク抽出アルゴリズムの効率性の実験的評価

先に確立した最大クリーク抽出のための分枝限定アルゴリズム MCR (Tomita et al., Journal of Global Optimization 2007) を基礎として開発してきた最大クリーク抽出アルゴリズム MCS に対して、アルゴリズムの各構成要素における探索領域 (分枝限定効果) と探索時間の関係について、より詳細化した実験を行い、開発アルゴリズムの各要素の効果を明らかにし、それらが巧妙に相互作用を行った上で全体として非常に効率的なアルゴリズムとなっていることを、明確な実験的評価結果として与えた。これによって、更なる高効率化のための方向性を明らかに

出来た。

具体的結果の一部を、表 1 (探索領域比較)、表 2 (実行時間比較) に示す。ここで、dfmax はベンチマーク用の基準アルゴリズム、その他 MCR-R 等はアルゴリズム MCS を構成する要素となるアルゴリズムである。これにより、探索領域の削減が小さいオーバーヘッドで実現出来、結果的に MCS の高効率性につながっていることが明らかにされた。

[3] 最大クリーク抽出アルゴリズム、極大クリーク全列挙アルゴリズムの応用

1) 最大クリーク抽出アルゴリズムの応用

2 対象構造間におけるマッチングの抽出を 2 対象構造の直積空間内の最大クリーク抽出によって実現することが出来る。これを基本として、糖鎖を表す木構造間の編集距離計算

というバイオインフォマティクス上の問題に対し、これまでに開発した効率的な最大クリーク抽出アルゴリズムが効果的に応用出来ることを明らかにした。

2) 極大クリーク全列挙アルゴリズムの応用

極大クリーク全列挙アルゴリズム CLIQUES の手法を一般化した、疑似極大クリーク全列挙アルゴリズムにより、ソーシャルネットワークの構造変化を的確に抽出して、有効な情報を提示出来ることを明らかにした。

[4] その他

上記に関連して、5. 主な発表論文等に記載した様に、画像処理、学習アルゴリズムについての有効な研究成果を得た。

表 1 分枝数  $\times 10^{-3}$

Graph	$\omega$	dfmax	MCR	MCR-R	MCR*	MCR*-R	MCS
r200.8	24-27	542,315	3,266	1,206	1,619	703	703
r200.9	40-44	$>4.29 \times 10^6$	97,627	18,401	20,660	6,310	6,310
brock200_1	21	38,043	482	218	296	144	144
brock400_1	27	$>4.29 \times 10^6$	329,599	139,299	182,920	89,389	89,389
MANN_a45	345	$>4.29 \times 10^6$	2,952	$>4.29 \times 10^6$	2,952	225	225
p_hat500-2	36	292,275	408	139	124	61	61
p_hat700-3	62	$>4.29 \times 10^6$	3,733,665	629,368	201,949	88,168	88,168
san400_0.9_1	100	$>4.29 \times 10^6$	74.0	2.4	20.0	2.1	2.1
sanr400_0.7	21	$>4.29 \times 10^6$	89,124	43,206	54,622	28,513	28,513

表 2 実行時間 [sec]

Graph	$\omega$	dfmax	MCR	MCR-R	MCR*	MCR*-R	MCS
r200.8	24-27	192.7	12.3	7.5	6.8	4.8	4.5
r200.9	40-44	$>10^5$	647	197	158	82	74
brock200_1	21	14.53	1.72	1.26	1.16	0.90	0.86
brock400_1	27	22,051	1,771	1,157	1,057	748	693
MANN_a45	345	$>10^5$	3,090	$>10^5$	3,089	314	281
p_hat500-2	36	132.9	3.1	1.6	1.2	0.8	0.7
p_hat700-3	62	$>10^5$	68,187	18,401	4,374	2,504	2,392
san400_0.9_1	100	$>10^5$	3.43	0.14	1.00	0.12	0.12
sanr400_0.7	21	2,426	379	284	248	197	181

5. 主な発表論文等  
(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 20 件)

1. Etsuji Tomita, Yoichi Sutani, Takanori Higashi, Mitsuo Wakatsuki: “A simple and faster branch-and-bound algorithm for finding a maximum clique with computational experiments,” IEICE Trans. on Information and Systems, vol.E96-D, no.6, 査読有 (2013 掲載確定)
2. 中西裕陽, 富田悦次, 若月光夫, 西野哲朗: “最大クリーク問題の多項式時間的<sup>可解性</sup>の拡張,” 電子情報通信学会論文誌 D, vol.J95-D, no.9, pp. 1716-1728, 査読有 (2012)
3. Tomoya Mori, Takeyuki Tamura, Daiji Fukagawa, Atsuhiko Takasu, Etsuji Tomita, Tatsuya Akutsu, “A clique-based method using dynamic programming for computing edit distance between unordered trees,” Journal of Computational Biology, 19, pp.1089-1104, 査読有 (2012)
4. Yoshiaki Okubo, Makoto Haraguchi, Etsuji Tomita, “Structural change pattern mining based on constrained maximal k-plex search,” Lecture Notes in Artificial Intelligence, 7569, pp.284-298, 査読有 (2012)
5. Aixiang Li, Makoto Haraguchi, Yoshiaki Okubo, Etsuji Tomita: “Finding what changes for two graphs constructed from different time intervals,” Proc. the 2012 IIAI International Conference on Advanced Applied Informatics, pp.48-53, 査読有 (2012).
6. Yoshiaki Okubo, Makoto Haraguchi, Etsuji Tomita, “Relational change pattern mining based on modularity difference,” Proc. International Workshop on Information Search, Integration and Personalization, pp.1-10, 査読有 (2012)
7. 中西裕陽, 富田悦次, 若月光夫, 西野哲朗: “最大クリーク問題の多項式時間的<sup>可解性</sup>の更なる改良結果,” 電子情報通信学会論文誌 D, J94-D, no.12, pp.2037-2046, 査読有 (2011)
8. 中西裕陽, 富田悦次: “最大クリーク問題の多項式時間的<sup>可解性</sup>の改良結果,” 電子情報通信学会論文誌 D, vol.J94-D, no.5, pp.843-851, 査読有 (2011)
9. Tatsuya Akutsu, Tomoya Mori, Daiji Fukagawa, Atsuhiko Takasu, Takeyuki Tamura, Etsuji Tomita: “An improved clique-based method for the edit distance between unordered trees with application to similar glycan structures,” Proc. International Conference on Complex, Intelligent and Software Intensive Systems, pp.536-540, 査読有 (2011)
10. Daiji Fukagawa, Takeyuki Tamura, Atsuhiko Takasu, Etsuji Tomita, Tatsuya Akutsu: “A clique-based method for the edit distance between unordered trees and its application to analysis of glycan structures,” BMC Bioinformatics, 12:S13, pp1-9, 査読有 (2011)
11. Akinobu Miyata, Jun Tarui, Etsuji Tomita: “Learning Boolean functions in AC<sup>0</sup> on attribute and classification noise - Estimating an upper bound on attribute and classification noise,” Theoretical Computer Science, 412, pp.4650-4660, 査読有 (2011)
12. 中西裕陽, 富田悦次, 若月光夫, 西野哲朗: “最大クリーク問題の多項式時間的<sup>可解性</sup>について,” 京都大学数理解析研究所講究録, vol. 1744, pp. 169-176, 査読無 (2011)
13. 若月光夫, 清野和司, 富田悦次, 西野哲朗: “実時間最終状態受理式決定性限定 1 カウンタ変換器の多項式時間等価性判定アルゴリズム,” 京都大学数理解析研究所講究録, vol. 1744, pp. 1-10, 査読無 (2011)
14. Jesus Robles-Castro, Gonzalo Duchon-Sanchez, Haruhisa Takahashi: “Improving object position estimation based on non-linear mapping using relevance vector machine,” Proc. International Conference on Electrical Communications and Computers (CONIELECOMP), pp.171-176, 査読有 (2011)
15. Rameswar Debnath, Haruhisa Takahashi, Takio Kurita: “A Comparison of SVM-based evolutionary methods for multicategory cancer diagnosis using microarray gene expression data,” Journal of Systemics, Cybernetics and Informatics, vol.6, pp.63-68, 査読有 (2011)

16. 中西裕陽, 富田悦次: “最大クリーク問題の多項式時間的可解性の一結果,” 電子情報通信学会論文誌 D, vol.J93-D, no.4, pp.417-425, 査読有 (2010)
17. 中西裕陽, 富田悦次, 若月光夫: “最大クリーク抽出の単純な最大時間計算量評価と多項式時間的可解性,” 京都大学数理解析研究所講究録, vol. 1691, pp. 108-114, 査読無 (2010)
18. Mitsuo Wakatsuki, Etsuji Tomita: “Polynomial time identification of strict prefix deterministic finite state transducers,” Lecture Notes in Artificial Intelligence, 6339, pp.313-316, 査読有 (2010)
19. Atsuo Nomoto, Kazuhiro Hotta, Haruhisa Takahashi: “Asbestos detection in building materials through consolidation of similarities in color and shape features,” Journal of Robotics and Mechatronics, vol.22, no.4, pp. 496-505, 査読有 (2010)
20. Hirokazu Nagai, Haruhisa Takahashi, Kazuhiro Hotta: “Fast human action recognition using conditional random field,” Journal of Signal Processing, vol. 14, no. 6, pp499-505, 査読有 (2010)

[学会発表] (計 16 件)

1. 間澤直寛, 原口誠, 富田悦次: “疑似独立集合制約と正規化カットを用いたグラフの構造比較,” 情報処理学会数理モデル化と問題解決研究会, 2013年2月27-28日, 武雄文化会館.
2. 若月光夫, 富田悦次, 西野哲朗: “決定性限定 1 カウンタ変換器のある部分クラスに対する質問による 多項式時間学習アルゴリズム,” 冬のLAシンポジウム, 2013年 1月 28-30日, 京都大学.
3. 大久保好章, 原口誠, 富田悦次: “極大 k-plex探索による外部接続度を考慮した構造変化抽出,” ERATO湊離散構造処理系プロジェクト「2012年度秋のワークショップ」, 2012年10月15-17日, ゆうばりホテル シューパロ.
4. エラウインディサラ, 原口誠, 大久保好章, 富田悦次: “クリーク全列挙に基づく構造変化検出アルゴリズム,” 情報処理学会数理モデル化と問題解決研究会, 2012年3月1-2日, 指宿市市民会館.
5. 若月光夫, 清野和司, 富田悦次, 西野哲朗: “最終状態受理式決定性限定 1 カウンタ変換器の多項式時間等価性判定アルゴリズム,” 冬の LA シンポジウム, 2012年 1月 30日-2月 1日, 京都大学.
6. 富田悦次: “最大・極大クリーク抽出アルゴリズムとその応用,” 東京工業大学 GCOE セミナー, 2011年 12月 19日, 東京工業大学.
7. 原口誠, 大久保好章, 富田悦次, 吉岡真治: “変化検出のための極大整合連結集合,” 人工知能学会 人工知能基本問題研究会, 2011年 12月 16日, 慶応大学.
8. 中西裕陽, 富田悦次, 若月光夫, 西野哲朗: “最大クリーク問題の多項式時間的可解性の拡張,” 電子情報通信学会コンピュータ研究会, 2011年 10月 21日, 東北大学.
9. 富田悦次: “最大クリーク問題の多項式時間的可解性について,” JST ERATO 湊離散構造処理系プロジェクト セミナー, 2011年 10月 12日, 北海道大学.
10. 森 智弥, 田村武幸, 深川大路, 高須淳宏, 富田悦次, 阿久津達也: “An improved clique-based method for computing edit distance between rooted unordered trees,” 情報処理学会バイオ情報学研究会, 2011年 9月 13日, 神戸大学.
11. 中西裕陽, 富田悦次, 若月光夫, 西野哲朗: “最大クリーク問題の多項式時間的可解性の更なる改良結果,” 電子情報通信学会コンピュータ研究会, 2011年 4月 22日, 京都大学.
12. 中西裕陽, 富田悦次, 若月光夫, 西野哲朗: “最大クリーク問題の多項式時間的可解性について,” 冬の LA シンポジウム, 2011年 2月 1-3日, 京都大学.
13. 若月光夫, 清野和司, 富田悦次, 西野哲朗: “実時間最終状態受理式決定性限定 1 カウンタ変換器の多項式時間等価性判定アルゴリズム,” 冬の LA シンポジウム, 2011年 2月 1-3日, 京都大学.
14. 中西裕陽, 富田悦次: “最大クリーク問題の多項式時間的可解性の改良結果,” 電子情報通信学会コンピュータ研究会, 2010年 12月 3日, 九州工業大学 Kyutech プラザ.
15. 深川大路, 田村武幸, 高須淳宏, 富田悦次, 阿久津達也: “無順序木の編集距離を計算する実用的アルゴリズムについて,” 人工知能学会 人工知能基本問題研究会, 2010年 8月 1日, 兵庫県立大学.
16. Etsuji Tomita: “PLENARY LECTURE, The Maximum Clique Problem,” WSEAS International Conference, 2010年 7月 24日, Corfu Island, Greece.

[図書] (計 4 件)

1. 富田悦次, 横森貴: 「オートマトン・言語理論 (第二版)」森北出版 (印刷中)
2. 西野哲朗, 若月光夫, 後藤隆彰: 「応用オートマトン工学」コロナ社 (2012)

3. Etsuji Tomita, Tatsuya Akutsu, Tsutomu Matsunaga: “Efficient algorithms for finding maximum and maximal cliques: Effective tools for bioinformatics,” in “Biomedical Engineering, Trends in Electronics, Communications and Software,” A. N. Laskovski (Ed.), InTech, pp.625-640, (2011)  
<http://www.intechopen.com/articles/show/title/efficient-algorithms-for-finding-maximum-and-maximal-cliques-effective-tools-for-bioinformatics>
4. 中森真理雄, 富田悦次: “数理モデルの応用技術,” 情報処理学会 50 年のあゆみ, 第 3 編, 情報処理学会, pp.227-229 (2010)

[その他]

・ホームページ:

[http://www.etlab.ice.uec.ac.jp/~tomita/index\\_e.html](http://www.etlab.ice.uec.ac.jp/~tomita/index_e.html)

・本研究の重要な基礎をなす下記論文 TCS2006 は,  
「Theoretical Computer Science Top Cited Article 2005-2010」表彰を受けた。

(論文 TCS2006)

Etsuji Tomita, Akira Tanaka, Haruhisa Takahashi: “The worst-case time complexity for generating all maximal cliques and computational experiments,” Theoretical Computer Science, vol.363, pp.28-42 (2006)

・前記論文(TCS2006)は, 次の特集号記事の中においても詳しく紹介された。

[信学誌] 伊藤大雄, “クリーク列挙 — アルゴリズムと下限 —,” 電子情報通信学会誌, vol.95, no.6, pp.484-489 (小特集 「広がる列挙の技術 — 列挙による問題解決アプローチ —」) (2012年6月号)

[人工知能学誌] 宇野毅明, “高速クリーク・密部分グラフマイニングアルゴリズム,” 人工知能学会誌, vol.27, no.3, pp.246-251 (特集: 「離散構造処理系 — 知能情報処理を支えるアルゴリズムの技法」) (2012年5月号)

・上記 [図書] の 3 番目 (印刷物図書およびオンラインで出版の Book Chapter) は 8,500 回以上ダウンロードされた。

## 6. 研究組織

### (1) 研究代表者

富田 悦次 (TOMITA ETSUJI)

電気通信大学・名誉教授

研究者番号: 40016598

### (2) 研究分担者

若月 光夫 (WAKATSUKI MITSUO)

電気通信大学・大学院情報理工学研究科・助教

研究者番号: 30251705

高橋 治久 (TAKAHASHI HARUHISA)

電気通信大学・大学院情報理工学研究科・教授

研究者番号: 90135418

西野 哲朗 (NISHINO TETSURO)

電気通信大学・大学院情報理工学研究科・教授

研究者番号: 10198484