

報告番号	※ 乙 第	号
------	-------	---

主 論 文 の 要 旨

論文題目 A Study of the Biclique Edge Partition and Cover Problems
 (2部クリーク辺被覆問題および2部クリーク辺分割問題の研究)

氏 名 大月 英明

論 文 内 容 の 要 旨

This thesis addresses the biclique edge partition (cover) problem, the problem of finding the minimum number of bicliques to partition (cover) the edge set of a given input bipartite graph.

The biclique edge partition problem is known to be NP-hard, and few results are known about its approximation hardness. In this thesis, a lower bound for the approximation ratio is demonstrated. Specifically, it is shown that we cannot have an approximate solution that is less than $6053/6052$ times the optimal solution unless $P=NP$. To obtain this hardness result, a gap-preserving reduction from the satisfiability problem to the biclique edge partition problem is given.

The biclique edge cover problem is also known to be NP-hard, and a polynomial time algorithm has been proposed for some restricted graph classes. For C_4 -free graphs and distance-hereditary bipartite graphs, a polynomial time algorithm was given in 1996. For domino-free graphs, which properly contain C_4 -free and distance-hereditary bipartite graphs, a polynomial time algorithm was given in 1999. In this thesis, a polynomial time algorithm for a new graph class is proposed. This graph class properly contains the class of domino-free bipartite graphs. Specifically, for a given bipartite graph B , a parameter $R(B)$ is defined using the modified Galois lattice, $G_m(B)$. This thesis shows that B is a domino-free graph if and only if $R(B) = 0$. The graph class defined by $R(B) \leq 1$ properly includes the domino-free graphs. A polynomial time algorithm is proposed for this new graph class.

It has been shown that if B is a domino-free graph then the size of the modified Galois lattice $G_m(B)$ is $O(nm)$, where n and m are the numbers of vertices and edges of B , respectively. In this thesis, it is shown that the size of $G_m(B)$ is actually $O(n+m)$ for graphs with $R(B) = O(1)$. This improves the time complexity of the algorithm proposed in this thesis.

The structure of this thesis is as follows. Chapter 1 is an introduction of this study. It is reviewed that graphs are not only mathematical objects but also very useful tools for modeling real-world problems. In particular, bipartite graphs are often used to model real-world problems. For example, in data mining, relations between two disjoint sets - such as "customers" and "products" or "phrases" and "documents" - are modeled by bipartite graphs. Also, in formal concept analysis, "objects" and "attributes" are well-modeled by bipartite graphs. In these applications, bipartite graphs are often required to be covered with bicliques.

In Chapter 2, definitions and notations used in this study are given. In order to introduce the concept of computational complexity class. Basic terms (alphabet, language, Turing machines, and so on) are defined in Sections 2.1, 2.2, and 2.3. Also, basic terms in the graph theory are introduced in Section 2.4. The satisfiability problem is noticed since it has important role in the theory of computational complexity and approximation algorithms. Polynomial time algorithms for the biclique edge cover and partition problems for restricted graph classes are reviewed.

In Chapter 3, the approximation hardness result for the biclique edge partition problem is shown. Specifically, it is shown that the biclique edge partition problem cannot be approximated within a ratio of $6053/6052$ unless $P=NP$. In order to obtain this approximation hardness, an approximation-preserving (gap-preserving) reduction from 3-OCC-MAX 2-SAT to the biclique edge partition problem is presented. 3-OCC-MAX 2-SAT is the satisfiability problem such that each clause has exactly two literals, and every variable occurs exactly three times. It has been shown that 2012/2011-approximation for 3-OCC-MAX 2-SAT is NP-hard. Our reduction implies that if there is a $6053/6052$ -approximation algorithm for the biclique edge partition problem then $P=NP$. Therefore, under the assumption that $P \neq NP$, the biclique edge partition problem does not have a $6053/6052$ -approximation algorithm. It follows that the biclique edge partition problem has no polynomial-time approximation scheme unless $P=NP$. To the author's knowledge, the ratio of $6053/6052$ is the first explicit lower bound on the approximation hardness of the biclique edge partition problem.

In Chapter 4, a new graph class for which the biclique edge cover problem can be solved in polynomial time is presented. For general bipartite graphs, the biclique edge cover problem cannot be solved in polynomial time if $P \neq NP$. However, it can be solved in polynomial time if we restrict the class of input graphs. It has been shown that there is a polynomial time algorithm for C_4 -free bipartite graphs and for distance-hereditary bipartite graphs. It has been shown that there is a polynomial time algorithm for domino-free bipartite graphs. The class of domino-free graphs properly contains both the C_4 -free and the distance-hereditary bipartite graph classes. In this chapter, we extend the graph class for which the minimum biclique edge cover problem can be solved in polynomial time. For a bipartite graph B , the modified Galois lattice $Gm(B)$ is defined. Investigating the structure of $Gm(B)$, the redundant parameter $R(B)$ is introduced. It is proved that for a bipartite graph B with $R(B) \leq 1$, the biclique edge cover problem can be solved in polynomial time.

In Chapter 5, the size of the modified Galois lattice is investigated. In general, the size of the modified Galois lattice can be very large for a general bipartite graph. It is shown that the size of $G_m(B)$ is at most $2n + 1$ for a distance hereditary bipartite graph. For general bipartite graphs, it is shown that the size of the modified Galois lattice is at most $(R(B)/2 + 1)n + m + 2$, where n and m are the numbers of vertices and edges of B , respectively. This improves the time complexity of the algorithm proposed in this thesis.

Chapter 6 concludes this thesis.

報告番号	※ 乙 第	号
------	-------	---

主 論 文 の 要 旨

論文題目 A Study of the Biclique Edge Partition and Cover Problems
 (2部クリーク辺被覆問題および2部クリーク辺分割問題の研究)

氏 名 大月 英明

論 文 内 容 の 要 旨

本論文は、2部グラフの辺集合に対する2部クリーク辺分割問題、および2部クリーク辺被覆問題についての研究である。2部クリーク辺分割問題とは、与えられたグラフのすべての辺からなる集合を分割する最小2部クリーク集合を求める問題であり、2部クリーク辺被覆問題とは、与えられたグラフのすべての辺からなる集合を被覆する最小2部クリーク集合を求める問題である。

2部クリーク辺分割問題はNP困難問題であることが知られている。しかし近似の困難さに関する研究結果は少ない。本論文では、2部クリーク辺分割問題の近似比率に対して、ひとつの下界を与える。具体的には、 $P=NP$ でない限り、2部クリーク辺分割問題の6053/6052より良い近似比率をもつ多項式時間近似アルゴリズムは存在しないことを示す。この近似困難性を得るために、充足可能性問題から2部クリーク辺分割問題へのギャップ保存帰着を構成する。

2部クリーク辺被覆問題もNP困難であることが知られているが、ある制限されたグラフクラスに対しては多項式時間アルゴリズムが与えられている。1996年に、 C_4 フリー2部グラフと距離遺伝2部グラフに対して多項式時間アルゴリズムが与えられた。さらに1999年には、これら2つのグラフクラスを真に含むドミノフリー2部グラフに対して多項式時間アルゴリズムが与えられた。本論文では、新しいグラフクラスに対して多項式時間アルゴリズムを提案する。具体的には、まず与えられた2部グラフ B に対して修正ガロア束 $Gm(B)$ を定義する。 $Gm(B)$ の上で、あるパラメータ $R(B)$ を定義し、 B がドミノフリー2部グラフであるとき、そしてそのときに限り $R(B) = 0$ であることを示す。 $R(B) \leq 1$ で定義されるグラフクラスはドミノフリー2部グラフを真に含む。このグラフクラスに対し多項式時間アルゴリズムを与える。更に、 B がドミノフリー2部グラフであれば、修正ガロア束 $Gm(B)$ のサイズは $O(nm)$ であることを示し、更に $R(B) = O(1)$ のグラフに対し、 $Gm(B)$ のサイズは $O(n+m)$ であることを示す。ただし n と m はそれぞれ B の頂点数と辺数である。この結果は、本論文で与えるアルゴリズムの計算時間を改善する。

本論文は以下の構成である。第 1 章は本研究の導入である。最初に、2 部グラフは現実世界の問題のモデル化に用いられることを述べる。そのような例として、データマイニングにおける「顧客と商品」、または「単語と文書」など、互いに素な集合間の関係が 2 部グラフで自然にモデル化できることを説明する。このような応用例では、与えられた 2 部グラフを 2 部クリークで分割、もしくは被覆することが要求されることが多い。この章ではまず、そのような要求を満たす問題は、2 部クリーク辺分割問題もしくは 2 部クリーク辺被覆問題として定式化できることを述べる。その応用例の一つとして、マーケットバスケット分析の例を取り上げて概略を述べる。次に、NP 困難問題に対する近似アルゴリズム研究の背景を概観する。まず PCP 定理を紹介し、2 部クリーク辺分割問題の近似不可能性を得るために用いたギャップ保存帰着の手法について解説する。次に、2 部クリーク辺被覆問題に対する多項式時間アルゴリズムの先行研究を紹介する。この問題に対して、多項式時間アルゴリズムが存在する新しいグラフクラスに関して説明する。

第 2 章では本研究を述べるのに必要な用語の導入と、諸定義を述べる。この章ではまず、計算複雑性のクラス P 、 NP などを定義するために、言語、チューリング機械などを定義する。次に、この論文で用いられるグラフ理論の用語を定義し、この論文で触れるいくつかのグラフ理論の問題を定義する。次に、計算複雑性の理論に重要な役割を持つ、充足可能性問題について説明する。さらに、近似困難性の理論を大きく発展させた PCP 定理の説明をする。PCP 定理は対話証明系を用いた NP の特徴づけである。PCP 定理の最初の証明は、符号理論などを用いた複雑なものであるが、ここでは、近年与えられた、制約グラフを用いた PCP 定理の別証明を概観する。次に、NP 困難問題に対するアルゴリズムの近似比率の定義をする。最後に、2 部クリーク辺分割問題と 2 部クリーク辺被覆問題が多項式時間で解けるグラフクラスについて述べる。

第 3 章では、2 部クリーク辺分割問題の近似比率に対し、ひとつの下界を与える。具体的には、 $P=NP$ でない限り、2 部クリーク辺分割問題は $6053/6052$ より良い近似解を多項式時間で得ることは出来ないことを証明する。この章では、NP 困難問題である 3-OCC-MAX-2-SAT から 2 部クリーク辺分割問題への近似保存帰着（ギャップ保存帰着）を構成する。3-OCC-MAX-2-SAT は充足可能性問題のひとつである。この 3-OCC-MAX-2-SAT に関しては、先行研究により、任意の正の ϵ ($< 1/2$) に対し、 $(2012 - \epsilon)N$ 個以上の節を充足できる問題例と、 $(2011 + \epsilon)N$ 個以下の節しか充足できない問題例とを区別することは NP 困難であることが証明されている。ただし、 N は自然数である。この章では、この結果を用い、もし 2 部クリーク辺分割問題が $6053/6052$ より良い近似比率の近似アルゴリズムが存在するならば、上述の 3-OCC-MAX-2-SAT の 2 つの問題例が区別可能であることを証明する。すなわち $P \neq NP$ の仮定の下では、2 部クリーク辺分割問題の $6053/6052$ より良い近似比率をもつ近似アルゴリズムは存在しないことを証明する。このことから直ちに 2 部クリーク辺分割問題は多項式時間近似スキーム (PTAS) を持たないこともわかる。

第 4 章では 2 部クリーク辺被覆問題が多項式時間で解ける新しいグラフクラスを提案する。ドミノフリー 2 部グラフに対しては、この問題が多項式時間で解けることが知られている。本章で提案するグラフクラスはドミノフリー 2 部グラフを真に含む。この結果を与えるために、まず、入力 of 2 部グラフ B に対して修正ガロア束 $G_m(B)$ を導入する。 $G_m(B)$ は、 B の極大 2 部グラフと B の頂点を中心とする極大星グラフの集合を、頂点集合とみなした有向グラフである。 $G_m(B)$ の辺は、これらの頂点の間のある半順序によって定義され、 $G_m(B)$ はこの半順序に基づいて構成されたハッセ図になる。 $G_m(B)$ は先行研究において構成されたガロア束 $G(B)$ といくつか異なる点を持つ。まず、 $G(B)$ はドミノフリー 2 部グラフ B に対して定義されるが、

$G_m(B)$ は一般の 2 部グラフ B に対して定義される. また, 必ずしも極大 2 部クリークではない星グラフを頂点に持つ. さらに, ガロア束 $G(B)$ の構成には, グラフの単純化の手順が必要であるが, $G_m(B)$ の構成においてはそのような手順は必要ではない. この $G_m(B)$ の構成の後, $G_m(B)$ の持つ性質について述べる. B がドミノフリー 2 部グラフであれば, $G_m(B)$ は, $G(B)$ と同様の性質を持つ. 具体的には B のそれぞれの辺は, $G_m(B)$ における路に一意に対応する. この性質を用い, B の 2 部クリーク分割のサイズ, 2 部クリーク被覆のサイズ, さらに $G_m(B)$ のカットのサイズは等しいことが証明できる. 次に $G_m(B)$ 上で定義される重複パラメータ $R(B)$ を導入する. ここでは, B がドミノフリー 2 部グラフのとき, そしてそのときに限り $R(B) = 0$ であることが示される. 次に $R(B) \leq 1$ であるグラフ B のもつ性質について述べる. グラフ B を $R(B) \leq 1$ を満たすグラフとする. B は必ずしもドミノフリー 2 部グラフではない. このとき B の 2 部クリーク辺被覆は, $G_m(B)$ のカットになっていることを証明する. $G_m(B)$ の最小カットは多項式時間で求めることができることから, グラフ B に対する 2 部クリーク辺被覆問題は多項式時間で解けることになる. すなわち $R(B) \leq 1$ で定義されるグラフクラスに対して, 2 部クリーク辺被覆問題の多項式時間アルゴリズムが存在することが示される. $R(B) \leq 1$ で定義されるグラフクラスはドミノフリー 2 部グラフを真に含むことから, この結果は 2 部クリーク辺被覆問題が多項式時間で解けるグラフクラスを拡張したと言える.

第 5 章は修正ガロア束 $G_m(B)$ のサイズ (頂点数) についての考察である. 一般の 2 部グラフに対して, $G_m(B)$ のサイズは B の頂点数 n の指数関数で増加する可能性がある. このようなグラフに対しては, 前章で提案したアルゴリズムでは 2 部クリーク被覆問題を多項式時間で解くことは不可能である. この章では, グラフクラスを限ると $G_m(B)$ のサイズが n の多項式で抑えられることを示す. まず B が距離遺伝 2 部グラフの場合, $G_m(B)$ のサイズが $2n + 1$ 以下であることを示す. この証明には, 距離遺伝 2 部グラフ B には bisimplicial edge が存在するという性質が用いられる. 次に一般の 2 部グラフの場合, $G_m(B)$ のサイズは $(R(B)/2 + 1)n + m + 2$ 以下であることを示す. ここで, m は B の辺数である. このことから, $R(B)$ が $0(1)$ のグラフに対して, $G_m(B)$ のサイズは $O(n + m)$ であることが示される. この結果は, $R(B) \leq 1$ のグラフクラスに対して, 2 部クリーク被覆問題の解が n と m の多項式時間で得られることを意味する.

第 6 章では本論文の総括であり, 今後の課題として, 2 部クリーク被覆問題, 分割問題の計算時間の改善や, 多項式時間でこれらの問題が解けるグラフクラスの拡張ができるかどうか, などの課題があることを述べる.