Hardness of Approximation of Graph Partitioning into Balanced Complete Bipartite Subgraphs

Hideaki OTSUKI

Abstract

For a graph G, a biclique edge partition $\mathcal{S}_{BP}(G)$ is a collection of complete bipartite subgraphs $\{S_1, S_2, \ldots, S_q\}$ such that each edge of G is contained in exactly one S_i. This paper proves that the Minimum Balanced Complete Bipartite Partitioning Problem (BCBP) is NP-hard to approximate within a factor $(1 + \epsilon_B)$ where $\epsilon_B = 1/34544$. BCBP seeks for $\mathcal{S}_{BP}(G)$ such that each S_i is a balanced complete bipartite graph. A balanced complete bipartite graph is a bipartite graph $G(U, V, E)$ such that $|U| = |V|$ and for all vertices $u \in U$ and $v \in V$ there is an edge $uv \in E$.

1 Introduction

For a graph G, a biclique edge partition $\mathcal{S}_{BP}(G)$ is a collection of complete bipartite subgraphs $\{S_1, S_2, \ldots, S_q\}$ such that each edge of G is contained in exactly one S_i. This paper proves that the Minimum Balanced Complete Bipartite Partitioning Problem (BCBP) is NP-hard to approximate within a factor $(1 + \epsilon_B)$ where $\epsilon_B = 1/34544$. BCBP seeks for $\mathcal{S}_{BP}(G)$ such that each S_i is a balanced complete bipartite graph. A balanced complete bipartite graph is a bipartite graph $G(U, V, E)$ such that $|U| = |V|$ and for all vertices $u \in U$ and $v \in V$ there is an edge $uv \in E$.

The biclique edge partition problem (BPP) has been studied in relation to data mining and clustering[1]. It is known that BPP is NP-hard [2]. Though the covering version of BPP, the Minimum Biclique Cover Problem (BCP), has been extensively studied [3] [4], only a few results are known for approximation hardness of BPP [5]. Feige and Kogan [6] discussed the hardness of the problem of finding the maximum size of a balanced complete bipartite subgraph in a general graph.

2 Construction of an instance of BCBP

A Boolean expression φ is in the Conjunctive Normal Form (CNF) if φ is a conjunction of clauses and each clause is a disjunction of literals of positive or negated variables. For a given φ in CNF, the Maximum Satisfiability Problem (MAX-SAT) is the problem of finding a truth assignment to φ that maximizes the number of satisfied clauses. MAX-3-SAT is MAX-SAT in which each clause has at most three literals. MAX-E3-SAT is MAX-3-SAT in which each clause has exactly three literals of different variables. MAX-(3,Bk)-SAT is MAX-E3-SAT in which every literal occurs exactly k times. Berman et al. [7] proved the next lemma.

Lemma 2.1. *Let M be a positive integer. For every $0 < \epsilon < 1$, it is NP-hard to decide whether an instance of MAX-(3,B2)-SAT with $1016M$ clauses has a truth assignment that satisfies at least $(1016 - \epsilon)M$ clauses or at most $(1015 + \epsilon)M$ clauses.*

Thus they have the next theorem.

Theorem 2.2 (Theorem 2. of [7])**.** *For every $0 < \epsilon < 1$, it is NP-hard to approximate MAX-(3,B2)-SAT to within an approximation ratio smaller than $(1016 - \epsilon)/1015$.*

Let φ be an E3-CNF formula with n variables $x_i\,(i = 1, \ldots, n)$ and m clauses $C_j\,(j = 1, \ldots, m)$ and the number of occurrences of x_i is four. Since each clause has exactly three literals of different variables in φ, $4n = 3m$ holds. We transform φ into an instance $G = (V, E)$ of BCBP as follows. Our transformation is based on the reduction given in [8].

2.1 A Graph G_{xi} for variable x_i

Our discussion is for simple graphs and thus graphs have no loops or multiple edges. First we construct a torus for each variable x_i. For each occurrence $s = 1, \ldots, 4$, we consider a 5×5 lattice graph G^s_{xi} shown in Figure 1.

We cascade these G^s_{xi}'s as follows. We denote by x^s_i the s-th occurrence of variable x_i. Let $P^s_{i\,tu}$ $(1 \leq t \leq 5,\ 1 \leq u \leq 5)$ be vertices of G^s_{xi}. We will omit index i unless this causes confusion. We identify P^s_{1u} and P^s_{5u} for each u. Then we have four cylinder-like graphs. We cascade these four graphs, that is, four each $s\,(1 \leq s \leq 3)$, we identify P^s_{t5} and P^{s+1}_{t1} for each t. Finally, we identify P^4_{t5} and P^1_{t1} for each t. Thus we have a torus for each variable x_i. We denote this torus by G_{xi}.

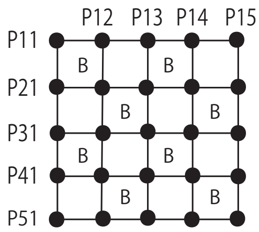

Figure 1: A lattice graph for x_i (s omitted). The black cycles are labeled with B.

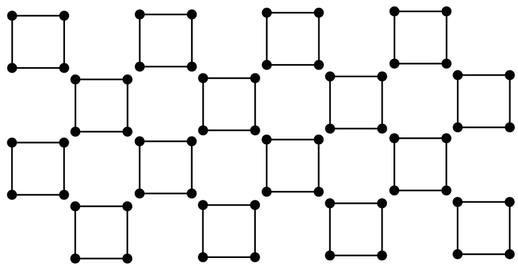

Figure 2: A Lattice graph partitioned into C_4's.

Apparently, balanced complete bipartite subgraphs of G_{xi} are $K_{2,2}(= C_4)$ and $K_{1,1}$ (an edge). G_{xi} has 64 C_4's.

Observe that G_{xi} can be edge-partitioned into 32 bicliequs (actually C_4's) in exactly two ways. One of which is shown in Figure 2. We define two sets of C_4's $\mathcal{B}$ and $\mathcal{W}$. $\mathcal{B}$ and $\mathcal{W}$ make a checkerwise pattern on G_{xi} and we assume cycle $(P^i_{11}, P^i_{12}, P^i_{21}, P^i_{22})$ is in $\mathcal{B}$. In Figure 1, C_4's in $\mathcal{B}$ are labeled with B. The remaning C_4's belong to $\mathcal{W}$. If a C_4 is in $\mathcal{B}$, we call it a black cycle, otherwise we call it a white cycle. We use this observation for a switch of an assignment for x_i being FALSE or TURE.

Next we define two subgraphs of G^s_{xi}, a T-patch and an F-patch. T-patch is a subgraph of G^s_{xi} induced by P^s_{22}, P^s_{23}, P^s_{32}, P^s_{33} and their adjacent vertices. We call cycle $(P^s_{22}, P^s_{23}, P^s_{32}, P^s_{33})$ the center of this T-patch. F-patch is a subgraph of G^s_{xi} induced by P^s_{32}, P^s_{33}, P^s_{42}, P^s_{43} and their adjacent vertices. We call cycle $(P^s_{32}, P^s_{33}, P^s_{42}, P^s_{43})$ the center of this F-patch. Note that the center of a T-patch is a black cycle and the center of an F-patch is a white cycle.

2.2 Graph G_{Cj} for clause C_j

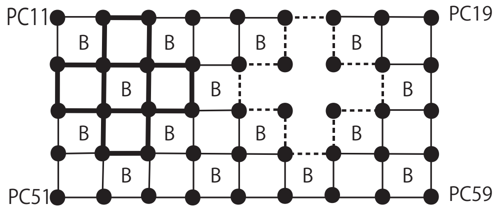

Figure 3: Lattice graph G^i_{Cj} and its two patches: an L-patch (thick line) and a C-patch (dotted line).

Next we construct a graph for each clause C_j ($j = 1, \ldots, m$). Let $x_{i_1}, x_{i_2}, x_{i_3}$ be variables appearing in C_j. For each $i \in \{i_1, i_2, i_3\}$, we construct a 5×9 lattice graph. We denote this graph by G^i_{Cj}. Let PC^i_{tu} ($1 \leq t \leq 5$, $1 \leq u \leq 9$) be vertices of G^i_{Cj}. We transform G^i_{Cj} into a torus by identifying PC^i_{1u} and PC^i_{5u} for each $u = 1, \ldots, 9$ and identifying PC^i_{t9} and PC^i_{t1} for each $t = 1, \ldots, 5$.

We define black cycles and white cycles in the same manner as in G_{xi}. We assume that cycle (PC^i_{11}, PC^i_{12}, PC^i_{21}, PC^i_{22}) is a black cycle. We delete the edges of black cycle ($PC^i_{26}, PC^i_{27}, PC^i_{37}, PC^i_{36}$). See Figure 3. We define a C-patch of G^i_{Cj} as the subgraph induced by vertices PC^i_{26}, PC^i_{27},PC^i_{37},PC^i_{36} and their adjacent vertices. Actually, C-patch is a cycle graph C_{12}. We identify C-patches of G^i_{Cj} for all $i \in \{i_1, i_2, i_3\}$. We denote the resulted graph by G_{Cj}.

For each $i \in \{i_1, i_2, i_3\}$, we define an L-patch of G^i_{Cj} as the subgraph induced by vertices PC^i_{22}, PC^i_{23}, PC^i_{33}, PC^i_{32} and their adjacent vertices. We call cycle (PC^i_{22}, PC^i_{23}, PC^i_{33}, PC^i_{32}) the center of this L-patch. Assume that x_i appears as positive literal in C_j and it is an s-th occurrence of x_i in φ. Then we identify the T-patch of G^s_{xi} and the L-patch of G^i_{Cj}. If x_i appears as negated literal in C_j, we identify the F-patch of G^s_{xi} and the L-patch of G^i_{Cj}. We denote by G the resulted graph. G consists of n torus graphs G_{xi} for variables and m graphs G_{Cj} for clauses. G has $204m$ distinct edges. Note that any balanced complete bipartite subgraphs in G is an edge or C_4. Thus, in order to minimize the size of $\mathcal{S}_{BP}(G)$, we must let C_4 be in $\mathcal{S}_{BP}(G)$ as many as possible.

2.3 How to partition G for φ

For the constructed graph G, we will show that if φ is satisfiable, G can be partitioned into $51m$ C_4's. Next we will show that for an arbitrarily small ϵ, if the number of satisfiable clauses of φ is at most $(1-\epsilon)m$, then we cannot partition G into less than $(51 + 9\epsilon/2)m$ balanced complete bipartite graphs.

Theorem 2.3. $(1+\epsilon_B)$*-approximation of BCBP for a bipartite graph is NP-hard, where* $\epsilon_B = 1/34544$.

Proof . *Assume that φ is satisfiable. We will show that all edges of G can be partitioned into only C_4's. Let π be a satisfying assignment and we will construct $\mathcal{S}_{BP}(G)$ from π where $|\mathcal{S}_{BP}(G)| = 51m$. We have observed that if we partition G_{xi} into white (or black) C_4's, all edges of G_{xi} are partitioned. If all edges of G_{xi} are partitioned into white C_4's, we call this partition W-partitioning. If all edges of G_{xi} are partitioned into black C_4's, we call this partition B-partitioning. For G^i_{Cj}, we define W-partitioning and B-partitioning in the same manner as for G_{xi}.*

In principle, if π assigns TRUE to x_i, we partition G_{xi} by W-partitioning, otherwise we partition G_{xi} by B-partitioning. We partition G_{Cj} as follows. For each clause C_j, choose one literal that is TRUE under π. If there are two or three TRUE literals, choose one literal arbitrarily. If this literal is s-occurrence of x_i, we denote it by x^s_{ij} and we call it a selected literal. We denote by $x_{i'j}$ and $x_{i''j}$ the other two literals and denote by $x_{i'}$ and $x_{i''}$ their variables. We partition G^i_{Cj} by B-partitioning. See Figure 4.

Since the L-patch of this G^i_{Cj} has been partitioned, we have to change the partitioning of G^s_{xi} as follows. If x^s_{ij} is a positive literal, T-patch of G^s_{xi} was identified to the L-patch of this G^i_{Cj}. We exclude C_4's that partition the edges of this T-patch from the partition of G_{xi}. See Figure 5. If x^s_{ij} is a negated literal, we exclude C_4's that partition the edges of the F-patch of G^s_{xi} from the partition of G_{xi}. See Figure 6.

Next we partition $G^{i'}_{Cj}$ and $G^{i''}_{Cj}$ by W-partitioning as shown in Figure 7. That is, we do not touch the L-patch in this partition. Note that regardless of $x_{i'}$ ($x_{i''}$) being TRUE or FALSE, all edges of $G^{i'}_{Cj}$ ($G^{i''}_{Cj}$) can be partitioned into C_4's appropriately since the edges of T-patch/F-patch have already partitioned in the partitioning for G_{xi}.

For each clause C_j, we have partitioned all edges of three G_{Cj} into $(15{+}2{\times}8) = 31$ C_4's. Thus, all edges of G_{xi} have been partitioned into $32n - 4m$ $(= 24m - 4m)$ C_4's. Therefore, we have a biclique edge-partition $\mathcal{S}_{BP}(G)$ with $|\mathcal{S}_{BP}(G)| = 32n - 4m + 31m = 51m$. Since G has $204m$ edges, $\mathcal{S}_{BP}(G)$ is an optimal solution.

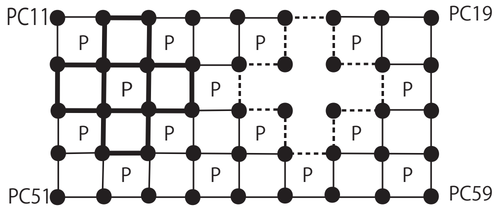

Figure 4: C_4's (indicated by label P) in $\mathcal{S}_{BP}(G)$ of G^i_{Cj} for a selected literal x^s_{ij}.

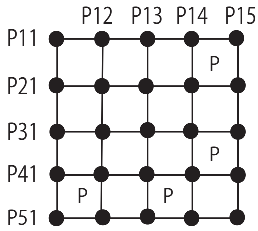

Figure 5: C_4's (indicated by label P) in $\mathcal{S}_{BP}(G)$ of G_{xi} when a selected literal x^s_{ij} is positive.

In order to prove the remaining part of Theorem 2.3, We will prove two lemmas. Let Cl be a set of clauses of φ, and C_j be a clause in Cl. We assume that C_j has variables x_a, x_b and x_c. We define G_j as the subgraph of G consisting of G_{Cj}, G_{x_a}, G_{x_b} and G_{x_c}.

Lemma 2.4. *If every G_j such that $C_j \in Cl$ can be partitioned into only C_4's, then there is an assignment that satisfies all clauses in Cl.*

Proof . *Since G_j is partitioned into C_4's, G_{Cj} is also partitioned into C_4's. Then, there is one G^i_{Cj} such that its C-patch is partitioned into C_4's. W.l.o.g., we assume that G^a_{Cj} is partitioned into C_4's. Thus, G^a_{Cj} is partitioned by B-partitioning (See Figure 4). If T-patch of G_{x_a} is identified to L-patch of G^a_{Cj}, that is, if x_a appears as a positive literal in C_j, then G_{x_a} should be partitioned by W-partitioning. If F-patch of G_{x_a} is identified to L-patch of G^a_{Cj}, then G_{x_a} should be partitioned by B-partitioning. G_{x_b} and G_{x_c} may be partitioned by either W-partitioning or B-partitioning. In this way, we decide the partition of G_{x_a} for a variable x_a of each clause in Cl. From the construction of G, it is easy to verify that these partitions are consistent each other. Since G_{x_a}, G_{x_b} and G_{x_c} are partitioned by either W-*

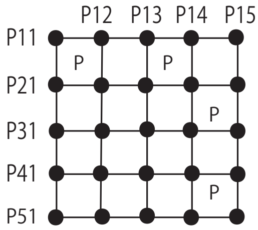

Figure 6: C_4's (indicated by label P) in $\mathcal{S}_{BP}(G)$ of G_{xi} when a selected literal x^s_{ij} is negative.

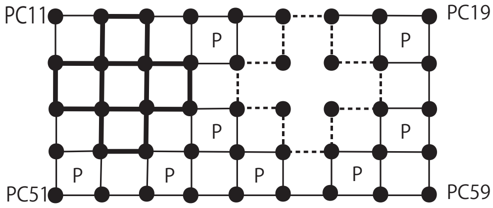

Figure 7: C_4's (indicated by label P) in $\mathcal{S}_{BP}(G)$ of G^i_{Cj} for a non-selected literal.

partitioning or B-partitioning, there is at least one consistent assignment that satisfies all clauses in Cl. □

Lemma 2.5. *Assume that for some positive constant* $\epsilon\,(< 1)$*, at most* $(1-\epsilon)m$ *clauses of* φ *are satisfiable simultaneously. Then, at least* $(51 + 3\epsilon/2)m$ *balanced complete bipartite subgraphs are necessary to partition* G*.*

Proof . *Let* $\mathcal{S}_{BP}(G)$ *be an optimal solution of BCBP.* $\mathcal{S}_{BP}(G)$ *consists of some* C_4*'s and some edges. We define* $\mathcal{S}_{C4}(G)$ *as the set of* C_4*'s in* $\mathcal{S}_{BP}(G)$*, and define* $\mathcal{S}_e(G)$ *as the set of edges in* $\mathcal{S}_{BP}(G)$*. Thus,* $\mathcal{S}_{BP}(G) = \mathcal{S}_{C4}(G) \cup \mathcal{S}_e(G)$*.*

Since the degree of each vertex of G *is even, each element (edge) of* $\mathcal{S}_e(G)$ *incidents to another element of* $\mathcal{S}_e(G)$*. Thus, edges of* $S_e(G)$ *make cycles. We denote by* $Cy(G)$ *the set of these cycles. Each cycle of* $Cy(G)$ *is even cycle, since* G *has no odd cycle. Since* $\mathcal{S}_{BP}(G)$ *is an optimal solution,* $Cy(G)$ *has no* C_4*. Thus, the size of each cycle of* $Cy(G)$ *is no less than six.*

We restrict ourselves to G_j *and define* $Cy(G_j)$*,* $\mathcal{E}(G_j)$ *and* $C_4(G_j)$ *as follows. Define* $Cy(G_j)$ *as the subset of* $Cy(G)$ *such that each element is a subgraph of* G_j*. Define* $\mathcal{E}(G_j)$ *as the subset of* $\mathcal{S}_e(G)$ *such that each element is a subgraph of* G_j*,* *and define $C_4(G_j)$ as the subset of $\mathcal{S}_{C4}(G)$ such that each element is a subgraph of G_j.*

Let π be an optimal assignment to φ. Let Cl' be the set of unsatisfied clauses of φ under π, and C_j be a clause in Cl'. From the assumption of the lemma, $|Cl'| \geq \epsilon m$ holds. We will show that $|\mathcal{S}_e(G)| \geq 2\epsilon m$ as follows.

Since one variable appears in at most four clauses, $\epsilon m/4$ variables appear at most ϵm clauses. Assume that $|\mathcal{S}_e(G)| < 8 \times \epsilon m/4 = 2\epsilon m$. Then the number of G_j such that $|\mathcal{E}(G_j)| \leq 8$ is more than $(1-\epsilon)m$. Note that $|E(G_j)| = 4|C_4(G_j)| + |\mathcal{E}(G_j)|$. Since $|E(G_j)|$ is multiple of four and $|\mathcal{E}(G_j)|$ has no C_4, it is clear that if $|Cy(G_j)| \neq 0$, then $|\mathcal{E}(G_j)| \geq 8$. Thus, the number of G_j's for which $|Cy(G_j)| = 0$ is more than $(1 - \epsilon)m$. From Lemma 2.4, there is an assignment that satisfies all C_j for which $|Cy(G_j)| = 0$. Thus, there exist an assignment that satisfies more than $(1 - \epsilon)m$ clauses. This leads to a contradiction.

Note that $204m = 4|\mathcal{S}_{C4}(G)| + |\mathcal{S}_e(G)|$. The number of C_4's in $\mathcal{S}_{BP}(G)$ is no more than $(204m - |\mathcal{S}_e(G)|)/4 = 51m - 2\epsilon m/4$. We have proved that it is not possible to partition G into less than $51m - 2\epsilon m/4 + 2\epsilon m = (51 + 3\epsilon/2)m$ balanced complete bipartite graphs. □

Let φ be a formula that can be satisfied at most $(1 - \epsilon)m$ clauses. We have proved that φ can be transformed into an instance of BCBP that cannot be partitioned into no less than $(51 + 3\epsilon/2)m$ balanced complete bipartite subgraphs. From Lemma 2.1 with $\epsilon = 1/1016$, we have Theorem 2.3. □

3 Conclusion

As we have mentioned in Section 1, only a few results are known for approximation hardness of the biclique edge partition problem (BPP). In [5], it is shown that BPP is NP-hard to approximate within a factor 6053/6052. (Note that the problem is not the balanced complete bipartite partition.)

Note that our reduction in this paper cannot be used directly to the biclique edge partition problem. The lattice graph can be partitioned not only into C_4's but into stargraphs $K_{1,t}$ $(2 \leq t)$. It is not clear that partitioning into stargraphs can be used for a switch of an assignment for x_i being FALSE or TURE. We cannot count the number of bicliques by a simple discussion, since G has $K_{1,t}$ $(2 \leq t \leq 8)$.

The Minimum Biclique Cover Problem (BCP) is a graph covering problem similar to BPP in which we seek a cover of the edge set instead of a partition. BCP arises in data mining, the set basis problem, textile designing and some other optimization problems [4][9]. Polynomial time algorithms are known for restricted cases [4]. It is known that $O(n^{1/3})$-approximation for BCP is NP-hard, where n is the number of vertices of G [3]. Such a good hardness result for BPP is challenging research.

References

[1] H. Zha, X. He, C. Ding, H. Simon, and M. Gu, "Bipartite graph partitioning and data clustering," Proceedings of the tenth international conference on Information and knowledge management, CIKM '01, New York, NY, USA, pp.25–32, ACM, 2001.

[2] T. Kratzke, B. Reznik, and D. West, "Eigensharp graphs: Decomposition into complete bipartite subgraphs," Transactions on the American Mathematical Society, vol.308 No.2, 1998.

[3] H. Gruber and M. Holzer, "Inapproximability of nondeterministic state and transition complexity assuming P ≠ NP," Lecture Notes In Computer Science, vol.4588, pp.205–216, 2007.

[4] J. Amilhastre, M.C. Vilarem, and P. Janssen, "Complexity of minimum biclique cover and minimum biclique decomposition for bipartite domino-free graphs," Discrete Appl. Math., vol.86, no.2-3, pp.125–144, 1998.

[5] H. Otsuki and T. Hirata, "Inapproximability of the minimum biclique edge partition problem," IEICE TRANSACTIONS on Information and Systems, vol.E93-D No.2, pp.290–292, 2010.

[6] U. Feige and S. Kogan, "Hardness of approximation of the balanced complete bipartite subgraph problem," Tech. Rep. MCS04-04, Department of Computer Science and Applied Math., The Weizmann Institute of Science, 2004.

[7] P. Berman, M. Karpinski, and A. Scott, "Approximation hardness of short symmetric instances of max-3sat," Tech. Rep. IHES-M-2004-28, Inst. Hautes Etud. Sci., Bures-sur-Yvette, 2004.

[8] I. Holyer, "The NP-completeness of some edge-partition problems," SIAM J. Comput., vol.10, no.4, pp.713–717, 1981.

[9] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," SIGMOD Rec., vol.22, pp.207–216, June 1993.