Some Weighted Clique Minimization Problems and their Approximability Properties

Geoffrey Solano[1], Jhoirene Clemente[2], Richelle Ann Juayong[2], and Ivy Ordanel[2]

[1] Department Physical Sciences and Mathematics
College of Arts and Sciences
University of the Philippines Manila, Philippines
gasolano@up.edu.ph
[2] Department of Computer Science
College of Engineering
University of the Philippines Diliman, Philippines
jbclemente@up.edu.ph, rbjuayong@up.edu.ph, idordanel@up.edu.ph

Abstract. Two minimum clique problems are considered in this study along with some of their important subclasses. The first is the Minimum Edge-Weighted Clique Problem (MIN-EWCP), which is the problem of finding the clique of order m with the least total weight in a complete edge-weighted graph G. The other one is the Minimum Weighted t-partite Clique Problem (MWtCP), which is the problem of finding the t-clique with the least total weight in a complete edge-weighted t-partite graph G. The results show that the $(1,2)$-*metric* MWtCP, *the* σ-*metric* MWtCP, and the *metric* MWtCP *are NP-hard*. The general MWtCP is inapproximable. The ultrametric variants of MWtCP and MIN-EWCP, on the other hand, are approximable with a performance guarantee of $\frac{3}{2}$. These problems were explored mainly because of their potential as a computational models of the biological problem of finding approximate gene clusters as is shown in some related studies, particularly the Approximate Gene Cluster Discovery Problem ($AGCDP$). Among the results shown in this paper is that $AGCDP$, when $w^+ = w^-$, is in *APX*.

Keywords: cliques, graph theory, edge-weighted clique, approximation

1 Introduction

Combinatorial optimization is a subfield of mathematical optimization that consists of finding an optimal object from a finite set of alternative objects with respect to some criterion. An example of such problems is that of searching for a discrete structure of extremal property in a given graph. Paths, cycles, spanning trees, matchings, and cliques are just some examples of such structures. While some of these combinatorial problems are polynomially solvable, such as the minimum spanning tree problem [10], unfortunately, many are intractable. Among such problems is the Weighted Clique Problem (WCP), which is the problem of finding a clique of a given order in a complete weighted graph, that is of extremal weight, measured as the sum of weights of the edges included in the said clique. Up to the most recent years, much is still being explored on the WCP

© The Author(s) 2024
J. Caro et al. (eds.), *Proceedings of the Workshop on Computation: Theory and Practice (WCTP 2023)*, Atlantis Highlights in Computer Sciences 20,
https://doi.org/10.2991/978-94-6463-388-7_3

and its variants [13, 11, 8, 6, 2]. These problems are still extensively studied because of their important applications in different domains.

In this paper we explore the approximability properties of two clique-finding minimization problems. One of them is the Minimum Edge-Weighted Clique Problem (MIN-EWCP), which is the problem of finding the clique of order m with the least total weight in a complete edge-weighted graph G. The other one is the Minimum Weighted t-partite Clique Problem (MWtCP), which is the problem of finding the t-clique with the least total weight in a complete edge-weighted t-partite graph G.

These problems were explored because of their potential as a computational models of the biological problem of finding approximate gene clusters[5, 7, 14–17]. Gene clusters are sets of closely related genes which are arranged in close proximity with each other, even after genome sequences have evolved in multiple events such as gene duplication and gene loss[9]. This problem has been formalized as the Approximate Gene Cluster Discovery Problem (AGCDP) [12].

This paper is organized as follow. The edge-weighted clique problems are defined in Section 2. Section 3 presents the results on the approximability properties of the problems and of the AGCDP. Finally, Section 4 presents our conclusion.

2 Some Edge-Weighted Clique Problems

This section discusses the two clique-finding problems highlighted in this study, namely, the Minimum Weighted t-partite Clique Problem (MWtCP) and the Minimum Edge-Weighted Clique Problem (MIN-EWCP). Both problems involve graphs where the edges have weights. The terms weight and cost are interchangeably used in this study.

2.1 Minimum Weighted t-partite Clique Problem

Given a positive integer t, where $1 \leq t \leq n$, a t-partite graph is a graph $G^t = (V^t, E^t)$ whose vertex set V^t can be partitioned into t different independent sets $\{V_1, V_2, \ldots V_t\}$. Thus, $V^t = V_1 \cup V_2 \cup \ldots \cup V_t$ and all the vertices in a partition V_i are non-adjacent for all $i \in I^+$ where $1 \leq i \leq t$. A **complete edge-weighted t-partite** graph is a t-partite graph $G = (V, E, c)$, where $c_{ij} > 0$, in which there is an edge between every pair of vertices from each pair of distinct independent sets. In this study, each partition or independent set has m vertices. Thus, $\mid V \mid = mt$, making the graph G $m(t-1)$-regular. Using the Handshaking Lemma, the total number of edges in the graph G, is $\mid E \mid = m\binom{t}{2}$[15].

Provided a weight for each edge (*e.g.*, defined as a function $w : E \rightarrow \mathbb{N}$), a main problem in graph theory lies in finding a clique with maximum or minimum weight. The MINIMUM WEIGHTED t-PARTITE CLIQUE PROBLEM (MWtCP) [15, 17] was defined as follows:

Definition 1 (The Minimum Weighted t-partite Clique Problem (MWtCP) [17]). .
Input: *a complete edge-weighted t-partite graph* $G = (V, E, w)$, *where* $w_{ij} \geq 0$ *and* $\mid V \mid = n$.
Question: *What is the minimum weighted complete subgraph (clique) of order t in G?*

A polynomial-time approximation algorithm for MWtCP, Algorithm 1, was also presented by Solano et. al in [17].

Input: A complete weighted t-partite graph $G = ((V, E), w)$
Output: a *minimum* weighted t-Clique
1. For every node x in G find the minimum weighted $t-1$-star with x as the hub or center vertex, by selecting the least-weighted edge from x to each of the remaining $t-1$ partitions
2. Find the minimum weighted t-star from the n candidate $t-1$-stars in #1
3. Return the t-clique formed from the vertex set of the $t-1$-star in #2

Algorithm 1: Minimum Weighted t-1 Star Algorithm for MWtCP

2.2 Minimum Edge-Weighted Clique Problem

A **complete edge-weighted graph** $G = (V, E, c)$, where $c_{ij} > 0$, is a graph where all the n vertices are adjacent to each other. For every vertex v_i of a complete graph, $deg(v_i) = n - 1$ and the size of G is $\binom{n}{2}$.

Eremin et. al., in [6], presented a general combinatorial problem called the Weighted Clique Problem (WCP).

Definition 2 (Weighted Clique Problem (WCP)[6]). .
Input: *a complete weighted undirected graph $G = (V, E, a, c)$, where weight functions $a : V \to \mathbb{Q}$ and $c : E \to \mathbb{Q}$, define the vertex weights and the edge weights respectively, and a positive integer m.*
Question: *What is the minimum(maximum) weighted complete subgraph (clique) of the graph G of order m?*

Here, the sum $\Sigma_{v \in V} a_v + \Sigma_{e \in E} c_e$ is called the weight of the graph G. A variant of WCP was further defined in [6] as the minimum edge-weighted clique problem.

Definition 3 (Minimum Edge-Weighted Clique Problem (MIN-EWCP)[6]). .
Input: *a complete weighted undirected graph $G = (V, E, a, c)$, where $a_i \geq 0$ and $c_{ij} \geq 0$, and a positive integer m*
Question: *What is the minimum weighted complete subgraph (clique) of the graph G of order m?*

It was shown that this problem is not in APX, or the class of NP optimization problems that allow polynomial-time approximation algorithms with approximation ratio bounded by a constant.

A graph-theoretic version of Eremin's solution to MIN-EWCP [6] using the Rows Subset of Symmetric Matrix Problem(RSSM) Problem is Algorithm 2.

Input: A complete graph $G = ((V, E), w)$
Output: a *minimum* weighted m-Clique
1. For every node x find the minimum weighted $m-1$-star with x as the hub or center vertex
2. Find the minimum weighted $m-1$-star from the n candidate $m-1$-stars in #1
3. Return the m-clique formed from the vertex set of the $m-1$-star in #2

Algorithm 2: Minimum Weighted $m-1$ Star Algorithm for MIN-EWCP

3 Some Hardness and Approximability Results

3.1 Results on Edge-Weighted Clique Problems

Some variants of edge-weighted clique problems are introduced in this section along with some results on the computational complexity of the variants of the problems. These variants are based on the instance of the edge weights of the input graph G. Recall that a complete edge-weighted t-partite graph $G = (V, E, w)$, where $w_{ij} \geq 0$ and $\mid V \mid = n$. A case of MWtCP is called *metric* if all edge weights in G follow the triangle inequality, $i.e.$, for any distinct vertices $a, b, c \in V$, $w(a, b) \leq w(b, c) + w(a, c)$, where $w(a, b)$ is the weight of the edge (a, b). Similarly, a case of MWtCP is called *ultrametric* if all edge weights in G follow the rule: for any distinct vertices $a, b, c, \in V$, $w(a, b) = max(w(b, c), w(a, c))$. Moreover, we refer to a case MWtCP as σ*-metric* if for any distinct vertices $a, b, c \in V$ and $\sigma \in \mathbf{Q}$, $w(a, b) \leq \sigma(w(b, c) + w(a, c))$. Finally, we refer to a case MWtCP as $(1, 2)$-metric MWtCP if for every edge $e \in E, w(e) \in (1, 2)$.

In [17] it was shown that the general case of MWtCP is *NP*-hard. This was done by showing that there is a reduction from the Clique Problem, to MWtCP. The Clique Problem is defined as follows :

Definition 4 (Clique Problem). .
Input: *a graph* $G = (V, E)$*, where* $\mid V \mid = n$ *and an integer* t*.*
Question: *Does there exist a clique* Q *of order* t *in* G*,* $i.e.$*,* $V(Q) = V_c$ *where* $V_c \subseteq V$*,* $\mid V_c \mid = t$ *and* $\forall u, v \in V_c, (u, v) \in E$ *?*

Part of that proof is showing that given any instance $G = ((V, E), t)$ of the CLIQUE PROBLEM where $V = \{v_1, v_2, v_3, \ldots, v_n\}$, a corresponding instance $G' = (V', E', w)$ of MWtCP can be constructed in polynomial time. We now show that given any instance $G = ((V, E), t)$ of the CLIQUE PROBLEM as defined above, may also be used to construct a corresponding instance in $(1, 2)$-metric MWtCP.

Lemma 1. *An instance* $G' = (V', E', w)$ *of the* $(1, 2)$*-metric* MWtCP *may be constructed from an instance* $G = ((V, E), t)$ *of* CLIQUE PROBLEM *in polynomial time.*

Proof. We define a graph $G' = (V', E')$. We let the set of vertices $V' = V'_1 \cup V'_2 \cup V'_3 \cup \ldots V'_t$, where for $1 \leq i \leq t$, and there is a copy of each vertex of V in V'_i where for every $i, V'_i = \{v^i_1, \ldots, v^i_n\}$. As for the edge set, we let $E' = (v^x_i, v^y_j)$ where

$(v_i, v_j) \in E$ such that $\forall 1 \leq x, y \leq t$ and $x \neq y$. Essentially, for each edge $(v_i, v_j) \in E$, a complete bipartite graph $K_{t,t}$ is built using the following two disjoint sets of vertices $\{v_i^1, v_i^2, \dots v_i^t\}$ and $\{v_j^1, v_j^2, \dots v_j^t\}$ and then remove all the edges $\{(v_i^x, v_j^x) \mid 1 \leq x \leq t\}$.

We then set assign edge weights $w(e) = 1$ for all such edges e currently in E'. We then define another edge set E^* where $E^* = \{(v_i^x, v_j^y), \mid \forall 1 \leq x < y \leq t; 1 \leq i \leq j \leq n; (v_i^x, v_j^y) \notin E)\}$, and assign each edge $e \in E^*$ a weight $w(e) = 2$. We then add edges to E' to come up with a complete t-partite graph, $i.e.$, $E' = E' \cup E^*$. It is clear that the constructed graph is an instance of $(1, 2)$-metric MWtCP and can be done in polynomial time. This is illustrated in Figure 1.

□

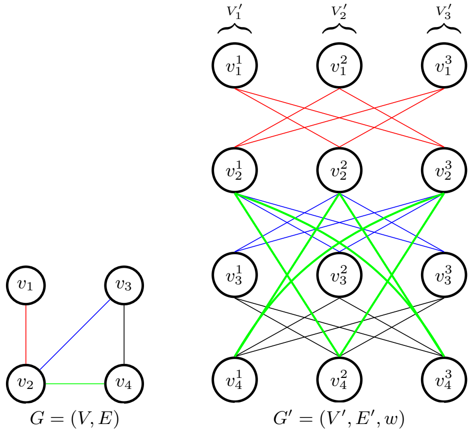

Fig. 1. The graph G of CLIQUE PROBLEM and the corresponding graph G' of $(1, 2)$-metric MWtCP, for $t = 3$, where only arcs of unit weight have been drawn; edges in E^*, which are not shown for readability, are of weight 2.

Lemma 2. *Let $G = ((V, E), t)$ be an instance of the* CLIQUE PROBLEM *problem and the corresponding graph $G' = (V', E', w)$ of the $(1, 2)$-metric* MWtCP *obtained by the above construction. G' contains a t-clique of minimum cost if G contains a t-clique.*

Proof. Let V_c, where $V_c \subseteq V$, be the vertex set of the t-clique in G. A corresponding solution, a minimum-weighted t-partite clique, may be constructed in $(1, 2)$-metric MWtCP, having the vertex set V_c', where $V_c' \subseteq V'$. If edge $(v_i, v_j) \in E$ is part of the t-clique in G, then $(v_i^x, v_j^y) \in E', \forall 1 \leq x, y \leq t$ and $x \neq y$. Thus, if edge (v_i, v_j) is part of the t-clique in G then $(v_i^p, v_j^q) \in E'$ where $1 \leq p \leq t$ and $q = p + 1$ if $p < t$ and $q = 1$ otherwise. We note that $\forall 1 \leq p \leq t, w(v_i^p, v_j^q) = 1$. Furthermore, a t-clique in G will result in a t-clique in G' whose total weight is $\binom{t}{2}$ and it is minimum total weight any t-clique could have.

□

Lemma 3. *Let $G = ((V, E), t)$ be an instance of the* Clique Problem *problem and the corresponding graph $G' = (V', E', w)$ of the $(1, 2)$-metric* MWtCP *obtained by the above construction. G contains a t-clique if G' contains a t-clique of minimum cost.*

Proof. Suppose we have a t-clique in G' whose total weight is $\binom{t}{2}$. We note that each edge is of unit weight. For every vertex $v_i^x \in V_c'$, we add the vertex v_i as part of V_c.

By Lemma 1, for all vertices $v_i^x \in V_c'$, i and x are unique. Thus, for all vertices $v_i \in V_c$, i is also unique. Furthermore, since the weight of each edge (v_i^p, v_j^q) in the t-clique of minimum weight is 1, then each pair of vertices $(v_i, v_j) \in V_c$ are adjacent in G. Thus, V_c forms a clique in G.

□

Theorem 1. *The $(1, 2)$-metric variant of* MWtCP *is NP-hard.*

Proof. Lemma 1 shows the mapping of the instances of the Clique Problem to the $(1, 2)$-metric MWtCP. It was also shown in Lemma 2 that a solution the Clique Problem can be used to obtain a solution to the $(1, 2)$-metric MWtCP, while the converse is shown in Lemma 3. Thus, the $(1, 2)$-metric MWtCP is also *NP*-hard.

□

Theorem 2. *The σ-metric and the metric variants of* MWtCP *are NP-hard.*

Proof. It is shown in Theorem 1 that the $(1, 2)$- metric MWtCP is NP-Hard. However, since the instances in the $(1, 2)$-metric MWtCP are also instances of the σ-metric as well as the metric variants of MWtCP, then the reduction of the Clique Problem to these variants of MWtCP to prove their hardness may use the same mapping of instances from the Clique Problem to instances of the $(1, 2)$-metric MWtCP. This is illustrated in Figure 2. Therefore, the σ-metric and the metric variants of MWtCP are *NP*-hard as well. □

We now show that there is no efficient approximation algorithm with a constant performance guarantee for the general case of MWtCP.

Theorem 3. *General* MWtCP *is inapproximable.*

Proof. We first note that the Clique Problem is a decision problem which cannot be decided in polynomial time, unless $P = NP$.

Suppose there is an efficient approximation algorithm M, with an approximation ratio ρ for any instance I of MWtCP, for some $\rho > 1$. Thus M can obtain a t-clique in polynomial time with respect to $|I|$ such that the weight of the solution is upper bounded by $\rho \cdot OPT(I)$. Let $K = \frac{t(t-1)}{2}$, following from a proof direction in [6]. From an instance of the Clique Problem ($G = (V, E), t$), we construct a corresponding instance I of MWtCP, using the same construction in Lemma 1, defined by the complete t-partite graph $G' = (V', E', w)$, this time, with the following edge weight assignment

$$w_{ij} = \begin{cases} 1, & \text{if } e \in E, \\ \rho K, & \text{otherwise} \end{cases}$$

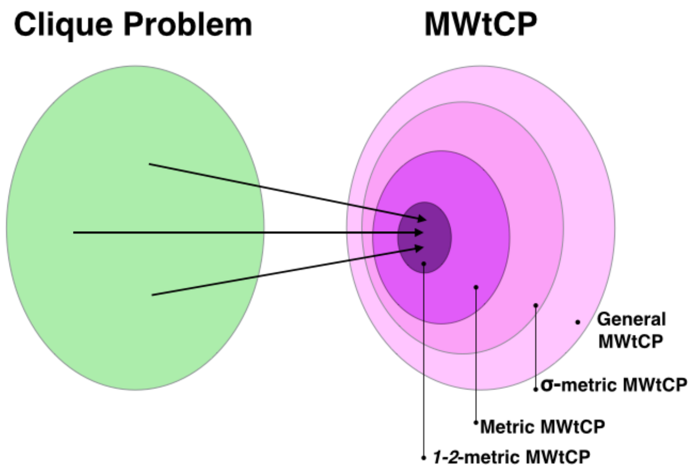

Fig. 2. Mapping of instances of the Clique Problem to the *1-2* metric-MWtCP

.

Therefore, if G has a t-clique, then $OPT(I) = K$, else $OPT(I) > \rho K$. The assignment of edge weights w_{ij}, the value K and the performance guarantee ρ imply that M will always return the optimal clique for MWtCP. It also consequently implies the polynomial time solvability of the Clique Problem, and that $P = NP$.

Therefore, the general case of MWtCP is inapproximable. Thus, if there exists a polynomial time approximation algorithm with an arbitrary fixed approximation performance guarantee of $\rho > 1$ for the general case of MWtCP, then $P = NP$.

□

Though the general case of MWtCP was shown to be inapproximable, it has also been shown [17] that certain variants of the problem may be approximable by narrowing the case of inputs. Algorithm 1 has been presented as a 2-approximation algorithm for the metric case of MWtCP.

We now show the performance guarantee of Algorithm 1 in the ultrametric case of inputs. The more precise characterization of the problem instances presented here compared to that in a previous study [17] promises a more accurate performance guarantee.

Theorem 4. *Algorithm 1 is a $\frac{3}{2}$-approximation algorithm for the ultrametric case of* MWtCP.

Proof. Given a complete graph G with positive edge weights, we define as Q as the optimal t-clique or the t-clique of minimum weight as seen at the left part of Figure 3. We let e be the number of edges of the t-clique, thus $e = \binom{t}{2}$, having the weights $w_1, w_2, w_3, \ldots w_e$. We let Q' be the solution or the t-clique returned by Algorithm 1 whose edges have the weights $w'_1, w'_2, w'_3, \ldots, w'_e$, seen at the right part of Figure 3. We also let Q^* be the minimum-weighted $(t-1)$-star in G, shown as the subgraph of Q' with dashed edges.

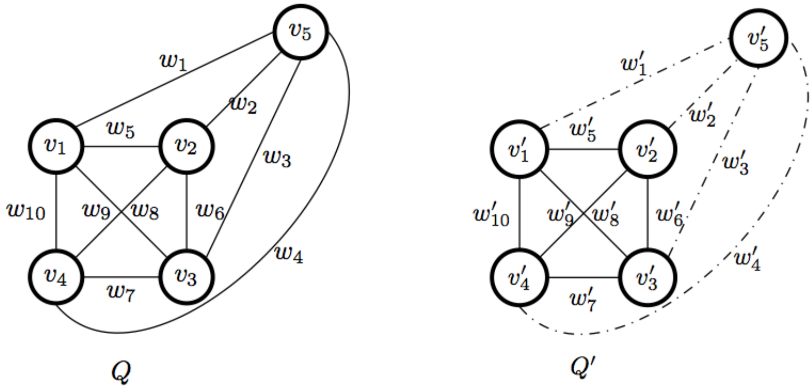

Fig. 3. The optimal t-clique Q and the solution Q' in graph G by Algorithm 1 for $t = 5$

We also note that

$$\begin{aligned} cost(Q') &= \sum_{i=0}^{e} w'_i \\ &= \sum_{i=0}^{t-1} w'_i + \sum_{i=t}^{e} w'_i \end{aligned}$$

thus,

$$cost(Q') = cost(Q^*) + \sum_{i=t}^{e} w'_i \tag{1}$$

and

$$cost(Q^*) \leq \frac{2}{t} \cdot cost(Q). \tag{2}$$

We also note that the average cost of an edge in Q^* is

$$\frac{cost(Q^*)}{t-1} \tag{3}$$

In obtaining the performance guarantee of Algorithm 1 for the ultrametric case of inputs, we wish to obtain an upper bound for $cost(Q')$. In other words, what is being asked here is a charaterization of the instance when solution returned by Algorithm 1 is the largest for the ultrametric case of inputs. Given that for all edge weights in G, for any distinct vertices $a, b, c, \in V$, $w(a,b) = max(w(b,c), w(a,c))$, this instance happens when there are only two values for the edge weights in Q^*, i.e. the edges that have less weight will have $w'_i = \frac{cost(Q^*)}{t-1} - \delta$, while the edges that have larger weight will have $w'_i = \frac{cost(Q^*)}{t-1} + \delta$. Since all the edges in G are non-negative, then $0 \leq \frac{cost(Q^*)}{t-1} - \delta$. This gives us an upper bound on δ, that is:

$$\delta \leq \frac{cost(Q^*)}{t-1}. \tag{4}$$

Given the instance mentioned, of the $\binom{t-1}{2}$ edges in $Q' \setminus Q^*$, because of the ultrametric property, half of the edges in Q^*, which have the less cost, cause a clique of order $\frac{t-1}{2}$ to be formed in $Q' \setminus Q^*$ having the same total cost. Thus, the number of edges with cost $\frac{cost(Q^*)}{t-1} - \delta$ is $\frac{(t-1)(t-3)}{8}$, while the number of edges with cost $\frac{cost(Q^*)}{t-1} + \delta$ is $\binom{t-1}{2} - \frac{(t-1)(t-3)}{8}$.

Because the set of edges with less cost will even out the part of the set of edges with larger cost, the number of edges with the average cost $\frac{cost(Q^*)}{t-1}$ is $\frac{(t-1)(t-3)}{4}$, while the number of edges with cost $\frac{cost(Q^*)}{t-1} + \delta$ is $(\binom{t-1}{2} - \frac{(t-1)(t-3)}{4})$.

Thus,

$$\begin{aligned} \sum_{i=t}^{e} w'_i &= \left[\frac{(t-1)(t-3)}{4} \cdot \frac{cost(Q^*)}{t-1}\right] + \left[\binom{t-1}{2} - \frac{(t-1)(t-3)}{4}\right] \cdot \left[\frac{cost(Q^*)}{t-1} + \delta\right] \\ &= \tfrac{t-3}{4} \cdot cost(Q^*) + \left[\tfrac{(t^2-2t+1)}{4}\right] \cdot \left[\tfrac{cost(Q^*)}{t-1} + \delta\right] \end{aligned} \tag{5}$$

Also,

$$cost(Q') = cost(Q^*) + \frac{t-3}{4} \cdot cost(Q^*) + \left[\frac{(t-1)^2}{4}\right] \cdot \left[\frac{cost(Q^*)}{t-1} + \delta\right] \tag{6}$$

Furthermore, given Equation 2 and the upper bound of δ

$$cost(Q') \leq \tfrac{2}{t} \cdot cost(Q) + \tfrac{t-3}{4} \cdot \tfrac{2}{t} \cdot cost(Q) + \tfrac{t-1}{2} \cdot \tfrac{2}{t} \cdot cost(Q)$$

$$\leq \left[\frac{4+t-3+2t-2}{2t}\right] \cdot cost(Q)$$

$$\leq \left[\frac{3t-1}{2t}\right] \cdot cost(Q)$$

$$\leq \left[\frac{3}{2} - O(\frac{1}{t})\right] \cdot cost(Q)$$

Algorithm 1 is therefore a $\frac{3}{2}$-approximation algorithm for the ultrametric case of MWtCP

□

Corollary 1. *The ultrametric case of* MIN-EWCP *is also* $\frac{3}{2}$*-approximable.*

Proof. Both Algorithms 1 and 2 approximate the minimum weighted clique by finding the minimum weighted star. The only difference is that Algorithm 1 considers partitions and thus tries to get a candidate star by obtaining the minimum-weighted edges incident from a node in a given partition to each of the other $t-1$ partitions, whereas Algorithm 2 simply obtains the minimum-weighted $m-1$ edges. Thus, if the order m of the clique of interest in MIN-EWCP is equal to t, then Q^* is the minimum weighted $t-1$, and Q' is the solution t-clique returned by both algorithms derived from Q^*. Thus, the exact same proof shown in Theorem 4 also shows that Algorithm 2 is a $\frac{3}{2}$-approximation algorithm for MIN-EWCP.

. □

We just showed that the weight of the clique returned by Algorithm 1 is at the worst case $\frac{3}{2}$ that of the optimal solution for the ultrametric case of inputs. We note that this case happens when the cost of each of the $\binom{t-1}{2}$ edges that make up $Q' \setminus Q^*$ is equal to the sum of cost of the two edges in Q^* it forms a triangle with, and when half of the edges Q^* common edge weight and the other half also have a common edge weight.

It was already mentioned that Algorithm 1 was presented is a 2-approximation algorithm [17] for the metric case of MWtCP. Clearly, the metric MWtCP is in APX. We now use this property of MWtCP to show that a related problem in the area of gene cluster discover is also in *APX*. This is the Approximate Gene Cluster Discovery Problem (AGCDP).

3.2 The Approximate Gene Cluster Discovery Problem (AGCDP)

A formalisation for the problem of gene cluster discovery was provided by Rahmann and Klau as the Approximate Gene Cluster Discovery Problem (AGCDP) [12]. A set of genomes is represented as $\mathcal{G} = \{g^1, g^2, \ldots, g^t\}$ where each g^i has length $n_i \in \mathbb{Z}^+$. A linear interval J^i in a genome g^i is an index set which can either be empty or $J^i = \{j, j+1, ..., k\}$, which can also be denoted as $J^i_{j_i,k_i}$ where $1 \leq j_i \leq k_i \leq n_i$. The gene content $G(J^i_{j,k})$ of a linear interval $J^i_{j,k}$ in genome g^i is the set of unique genes contained in that interval.

In their study, they defined the AGCDP the following double minimization problem.

Problem 1 (The Approximate Gene Cluster Discovery Problem (AGCDP) [12]). Given the gene universe $\mathcal{U} = \{0, 1, \ldots, N\}$, a set of genomes $\mathcal{G} = \{g^1, g^2, \ldots, g^t\}$, a positive constant D or size range $[D^-, D^+]$, and the cost of missing or additional genes in an interval represented by integer weights w^- and w^+, identify the set of genes X where $X \subset \mathcal{U}$, $0 \notin X$, $D^- \leq |X| \leq D^+$ or $|X| = D$ and a set of linear intervals $J = \{J^i_{j_i,k_i}\}$, $\forall\, i$ such that the cost function

$$cost(X, J) = \sum_{i=1}^{t} [(w^- \cdot |X \setminus G(J^i_{j_i,k_i})|) + (w^+ \cdot |G(J^i_{j_i,k_i}) \setminus X|)] \tag{7}$$

is **minimum**.

Rahmann and Klau also provided an integer linear programming formulation of this problem in the said paper. In another study, this time by Cabunducan et. al., it was shown that this problem is *NP*-hard [3]. A proposed transformation of AGCDP as a graph minimization problem has been presented by Aborot et. al. in [1].

We now show that $AGCDP$, when $w^+ = w^-$, is in *APX*

Theorem 5. *$AGCDP$, when $w^+ = w^-$, is in APX.*

Proof. It was shown in a previous study [15] how instances in AGCDP can be modeled as a complete edge-weighted t-partite graph G^* in MWtCP and it was shown as well how the corresponding solution in MWtCP provides an approximate gene cluster solution for AGCDP.

We note, however, that in the cost function between vertices defined as

$$cost(u, v) = [(w^- \cdot |G(J^x_{j_x,k_x}) \setminus G(J^y_{j_y,k_y})|) + (w^+ \cdot |G(J^y_{j_y,k_y}) \setminus G(J^x_{j_x,k_x})|)] \tag{8}$$

when $w^+ = w^-$, the cost between vertices in G^* will be the symmetric difference of the respective set of gene contents they represent, making the edge weights have the metric property.

In the work of Crescenzi [4], it was shown that given an NP Optimization Problem Γ, an instance x of Γ, a feasible solution y of x and the objective function m, the performance ratio of y with respect to x was defined as

$$R(x, y) = max \left\{ \frac{m(x, y)}{opt(x)}, \frac{opt(x)}{m(x, y)} \right\} \tag{9}$$

Furthermore, it was shown that in the reduction of NP Optimization problems $\mathit{\Gamma}$ into Z, there should be a polynomial-time computable function f mapping instances of $\mathit{\Gamma}$ into instances of Z, and also a polynomial-time computable function g mapping back solutions of Z into solutions of $\mathit{\Gamma}$, i.e. $x \to f(x)$ and $g(x,y) \in sol(x) \leftarrow y \in sol(f(x))$.

A reduction (f,g) was defined to be an A-reduction if a computable function $c : Q \cap (1,\infty) \to Q \cap (1,\infty)$ exists such that for any instance x of $\mathit{\Gamma}$, for any $y \in sol_Z(f(x))$, and for any $r > 1$, the following holds:

$$R_Z(f(x),y) \le r \Rightarrow R_{\mathit{\Gamma}}(x,g(x,y)) \le c(r) \tag{10}$$

Since the metric MWtCP and this case of AGCDP are both NP optimization problems, by A-reduction [4], if we set the computable function c, where $c : Q \cap (1,\infty) \to Q \cap (1,\infty)$, as the identity function, and since there is a 2-approximation algorithm for the metric MWtCP, such 2-approximation algorithm may also be used to obtain a solution for this case of AGCDP.

Thus, AGCDP, when $w^+ = w^-$, is in *APX*, as well.

□

4 Conclusion and Future Work

In this paper we explored some hardness as well as approximability properties of the Minimum Weighted t-partite Clique Problem and of the Minimum Edge-Weighted Clique Problem, along with some of their important subclasses, namely the general, the metric, σ-metric and the ultrametric case of inputs. An approximability result was also presented on the Approximate Gene Cluster Discovery problem when $w^+ = w^-$.

A problem area open for exploration is hardness of the of the ultrametric variants of these problems. Another open problem is the approximability of the general AGCDP.

Acknowledgements

We would like to thank our mentors, Dr. Henry Adorna, Dr. Guillaume Blin, Dr. Mathieu Raffinot and Dr. Jaime Caro, for their guidance.

Data Availability Statement

No datasets were analysed nor generated in this study as this work proceeded within a theoretical and mathematical approach.

References

1. Aborot, J.A., Adorna, H., Clemente, J.B. and de Jesus, B.K. and Solano, G.A.: Search for a Star: Approximate Gene Cluster Discovery Problem (AGCDP) as Minimization Problem on Graph. Philippine Computing Journal (2012)
2. Alidaee, B., Glover, F.W., Kochenberger, G.A., Wang, H.: Solving the maximum edge weight clique problem via unconstrained quadratic programming, Eur. J. Oper. Res., 181,592-597,(2007)
3. Cabunducan, G., Clemente, J.B., Adorna, H.N., Relator, R.T. :Probing the Hardness of the Approximate Gene Cluster Discovery Problem (AGCDP), Theory And Practice Of Computation - Proceedings Of Workshop On Computation: Theory and Practice, (2014)
4. Crescenzi: A short guide to approximation preserving reductions,Proceedings of Computational Complexity. Twelfth Annual IEEE Conference,262-27(1997)
5. Dela Rosa, J.L., Magpantay, A.E., Gonzaga, A.C., Solano, G.A.: Cluster center genes as candidate biomarkers for the classification of Leukemia,IISA 2014, The 5th International Conference on Information, Intelligence, Systems and Applications, IEEE, 124-129, (2014)
6. Eremin, I. I., Gimadi, E. Kh, Kelmanov, A. V., Pyatkin, A. V., Khachai, M.Y.: 2-Approximation Algorithm for Finding a Clique with Minimum Weight of Vertices and Edges, Proceedings of the Steklov Institute of Mathematics,284, SUPPL.1, 87-95, (2014)
7. Florendo, P.D.V., Solano, G.A.: InteGene: An Integer Linear Programming Tool for Discovering Approximate Gene Clusters, 10th International Conference on Information, Intelligence, Systems and Applications, IISA 2019, Patras, Greece, July 15-17, IEEE, 1-8, (2019)
8. Gimadi, E. Kh and Kelmanov, A. V. and Pyatkin, A. V. and Khachai, M. Y.: Efficient algorithms with performance guarantees for some problems of finding several cliques in a complete undirected weighted graph, Proceedings of the Steklov Institute of Mathematics, 289, (2015)
9. Lawrence, J. G.: Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes, Current Opinion in Genetics & Development, 9,6,642-648, (1999)
10. Prim, R.C.: Shortest Connection Networks And Some Generalizations, Bell System Technical Journal, vol. 36, issue 6, 1389-1401, (1957)
11. Pullan, W.J.: Approximating the maximum vertex/edge weighted clique using local search, Journal of Heuristics, 14, 117-134, (2008)
12. Rahmann, S., Klau, G.W.: Integer Linear Programming Techniques for Discovering Approximate Gene Clusters, Bioinformatics Algorithms: Techniques and Applications, 9, 203-222, (2008)
13. Shimizu, S., Yamaguchi, K., Masuda, S.: A Branch-and-Bound Based Exact Algorithm for the Maximum Edge-Weight Clique Problem, Computational Science/Intelligence & Applied Informatics, Springer International Publishing, 27-47 (2019)
14. Silmaro, B.C., Solano, G.A. : Clique-finding Tool for Detecting Approximate Gene Clusters, 10th International Conference on Information, Intelligence, Systems and Applications, IISA 2019, Patras, Greece, July 15-17, 2019, IEEE, 1-8 (2019)
15. Solano, G.A., Blin, G., Raffinot, M., Caro, J.: A Clique Finding Algorithm for the Approximate Gene Cluster Discovery Problem, Theory and Practice of Computation, World Scientific, 72-88 (2018)
16. Solano, G.A., Silmaro, B.C., Ojeda, S.A., Caro, J: Discovering Approximate Gene Clusters as a Minimum Weighted Clique Problem, Philippine Computing Journal, Computing Society of the Philippines, 13, 2 (2018)

17. Solano, G.A., Blin, G., Raffinot, M., Clemente, J.B., Caro, J.: On the Approximability of the Minimum Weight t-partite Clique Problem, Journal of Graph Algorithms and Applications, 24, 3, 171-190 (2020)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.