On the minimum edge cover and vertex partition by quasi-cliques problems

Oguz Kaya, Enver Kayaaslan, Bora Uçar

On the minimum edge cover and vertex partition by quasi-cliques problems

Oguz Kaya, Enver Kayaaslan, Bora Uçar

RESEARCH REPORT
N° 8255
February 2013
Project-Team ROMA

ISSN 0249-6399 ISRN INRIA/RR--8255--FR+ENG

On the minimum edge cover and vertex partition by quasi-cliques problems

Oguz Kaya*, Enver Kayaaslan†, Bora Uçar‡

Project-Team ROMA

Research Report n° 8255 — February 2013 — 11 pages

Abstract: A γ-quasi-clique in a simple undirected graph is a set of vertices which induces a subgraph with the edge density of at least γ for $0 < \gamma < 1$. A cover of a graph by γ-quasi-cliques is a set of γ-quasi-cliques where each edge of the graph is contained in at least one quasi-clique. The minimum cover by γ-quasi-cliques problem asks for a γ-quasi-clique cover with the minimum number of quasi-cliques. A partition of a graph by γ-quasi-cliques is a set of γ-quasi-cliques where each vertex of the graph belongs to exactly one quasi-clique. The minimum partition by γ-quasi-cliques problem asks for a vertex partition by γ-quasi-cliques with the minimum number of quasi-cliques. We show that the decision versions of the minimum cover and partition by γ-quasi-cliques problems are NP-complete for any fixed γ satisfying $0 < \gamma < 1$.

Key-words: clique, quasi-clique, edge cover, vertex partition, undirected graph

* College of Computing (GATECH) Georgia Institute of Technology (Georgia Tech), Atlanta, GA

† Department of Computer Engineering, Bilkent University, Ankara, Turkey

‡ CNRS and LIP, ENS Lyon, France

RESEARCH CENTRE GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Sur les problèmes de la couverture minimale des arêtes et de la partition minimale des sommets d'un graphe par des quasi-cliques

Résumé : Un γ-quasi-clique, pour $0 < \gamma < 1$, dans un graphe simple non-orienté est un sous-ensemble de sommets dont le sous-graphe induit a une densité d'arêtes supérieure ou égale à γ. Un ensemble de γ-quasi-cliques couvrant toutes les arêtes d'un graphe est appelé une couverture par des γ-quasi-cliques. Le problème de couverture minimale par des γ-quasi-cliques consiste à trouver une couverture par des γ-quasi-cliques ayant le plus petit nombre de quasi-cliques. Une partition des sommets d'un graphe par des quasi-cliques est un ensemble de γ-quasi-cliques telle que chaque sommet du graphe appartient à un seul quasi-clique de l'ensemble. Le problème de partition minimale par des γ-quasi-cliques consiste à trouver une partition par des γ-quasi-cliques ayant le plus petit nombre de quasi-cliques. Nous démontrons que les problèmes de décision associés aux problèmes de couverture minimale et partition minimale par des γ-quasi-cliques sont NP-complets pour tout γ satisfaisant $0 < \gamma < 1$ fixé.

Mots-clés : clique, quasi-clique, couverture des arêtes, partition des sommets, graphe nonorienté

1 Introduction and definitions

Let $G = (V, E)$ be an undirected simple graph and γ be a positive real less than 1. A γ-quasi-clique or a γ-clique in G is a set S of vertices such that $|E[S]| \geq \gamma\binom{|S|}{2}$ where $E[S] = \{\{u, v\} \in E : u, v \in S\}$. A cover by γ-cliques is a set $\mathcal{C}$ of γ-cliques that cover all the edges in G. In other words, for each edge $(u, v) \in E$, there is at least one γ-clique $C \in \mathcal{C}$ such that $u, v \in C$. The minimum cover by γ-cliques problem asks for a γ-clique cover of the minimum cardinality. A partition by γ-cliques is a set $\mathcal{C}$ of γ-cliques such that each vertex $v \in V$ belongs to exactly one γ-clique $C \in \mathcal{C}$. The minimum partition by γ-cliques problem asks for a γ-clique partition of the minimum cardinality.

It has been shown that the decision version of the minimum cover by $\frac{1}{2}$-cliques problem is NP-complete [6]. In this work, we show that the decision versions of the minimum cover and partition by γ-cliques problems are NP-complete for any fixed γ satisfying $0 < \gamma < 1$.

Pattillo et al. [7] show that the problem of finding the maximum cardinality γ-clique in a given undirected simple graph is NP-hard for any fixed γ satisfying $0 < \gamma < 1$. In doing so, they first show that the associated decision problem remains NP-complete, if γ is replaced by a rational number $\frac{p}{q}$ for integers p and q such that $1 \leq p < q$. Next they show [7, Corollary 1] how to find a p and a q for a given γ so that any γ-clique is a $\frac{p}{q}$-clique and vice versa, thus establishing the NP-completeness of the decision version of the maximum γ-clique problem. This last result help us demonstrate the NP-completeness of the decision versions of the minimum cover and partition by γ-cliques problems in two steps. We show the NP-completeness results with $\frac{p}{q}$-cliques as the first step. The NP-completeness for γ-cliques where $0 < \gamma < 1$ follows by the cited result of Pattillo et al. as the second step.

We define the notation used in what follows, though most are standard or very intuitive. For a graph $G = (V, E)$, we use V to refer to the set of its vertices; when the graph is not clear from the context, we use $V(G)$ to refer to the vertex set. Similarly, we use E and $E(G)$ to refer to the set of edges of G. For a vertex u, we use $adj(u)$ to denote the set of vertices it is connected to, that is $adj(u)$ is the set of neighbors of u. The number of neighbors of u is denoted as $degree(u)$. The maximum degree of a vertex in a graph G is denoted by $\Delta(G)$. For a set S, we use $|S|$ to denote its cardinality. For a nonnegative number x, we use $\lfloor x \rfloor$ to denote the largest integer smaller than or equal to x.

Our NP-completeness proofs use a reduction from the classical problem CLIQUE [4, GT19]: Given a graph $G = (V, E)$ and a positive integer $k \leq |V|$, does G contain a clique of size k or more? We assume $|E| \geq \binom{k}{2}$, as the other case is not interesting.

2 NP-completeness

2.1 Minimum cover by $\frac{p}{q}$-cliques problem

We consider the following decision problem associated with the minimum cover by $\frac{p}{q}$-cliques problem.

COVERBY-$\frac{p}{q}$-CLIQUES: Given a simple undirected graph $G = (V, E)$, fixed integers p and q satisfying $1 \leq p < q$, and an integer c, is there a cover of G

with c many $\frac{p}{q}$-cliques.

Theorem 1. *The* COVERBY-$\frac{p}{q}$-CLIQUES *problem is NP-complete for any positive integer constants p and q satisfying $1 \leq p < q$.*

The proof, which is deferred to the end of this section, uses a reduction from CLIQUE problem. For any fixed $1 \leq p < q$, we will therefore construct a graph $G' = (V', E')$ and an integer c for a given $G = (V, E)$ and $k \geq 2$ such that G has a clique of size k, assuming $|E| \geq \binom{k}{2}$, if and only if G' has a cover by $\frac{p}{q}$-cliques with $c = |E| - \binom{k}{2} + 1$ many $\frac{p}{q}$-cliques. We will discuss a few important properties of the constructed graph by a series of propositions before proving Theorem 1.

For a given positive integer m satisfying $1 \leq m \leq |V|$, we construct an auxiliary graph $Q_m = (V_m, E_m)$ with a technique similar to the one used by Pattillo et al. [7]. Let $z = (4qk+2)|V|^2$. We start with a set of $|V_m| = qz - m = q(4qk+2)|V|^2 - m$ vertices and enumerate them with indices $0, \ldots, |V_m|-1$. Then we add $|E_m| = \frac{p}{q}\binom{qz}{2} - \binom{m}{2} = \frac{p}{q}\binom{|V_m|+m}{2} - \binom{m}{2}$ edges in $w = \left\lfloor \frac{|E_m|}{|V_m|} \right\rfloor + 1$ passes. In each pass $r < w$, using $|V_m|$ edges we connect all vertex pairs $v_i, v_j \in V_m$ such that $j = (i + r) \mod |V_m|$, that is v_j comes as the rth vertex after v_i in the cyclic order. Hence, after pass $w - 1$, we obtain $2(w - 1) = 2 \left\lfloor \frac{|E_m|}{|V_m|} \right\rfloor$ regular graph Q_m. In the final pass, if need be, we add $(|E_m| \mod |V_m|)$ different edges formed by arbitrary pairs of vertices $v_i, v_j \in V_m$ satisfying the constraint $j = (i + w) \mod |V_m|$. This construction guarantees that $\Delta(Q_m) = \max_{v_i \in V_m} degree(v_i) \leq 2 \left\lfloor \frac{|E_m|}{|V_m|} \right\rfloor + 2$; because the difference in the degrees of vertices can only occur in the last pass, in which each vertex can be connected to at most two other vertices. Moreover, we have the following upper bound on $\Delta(Q_m)$ for $2 \leq m \leq |V|$:

$$\begin{aligned}
\Delta(Q_m) &\leq 2 \left\lfloor \frac{|E_m|}{|V_m|} \right\rfloor + 2 \\
&\leq 2 \left\lfloor \frac{p}{q}\binom{qz}{2} \frac{1}{qz - m} \right\rfloor + 2 \\
&\text{since } 2\lfloor a \rfloor \leq \lfloor 2a \rfloor \text{ for } a \geq 0 \\
&\leq \left\lfloor \frac{pz(qz-1)}{qz - m} \right\rfloor + 2 \\
&= \left\lfloor \frac{pz(qz-m)}{qz - m} + (m-1)\frac{qz - (q-p)z}{qz - m} \right\rfloor + 2 \\
&\text{since the first term is an integer, } q > p \text{ and } z > m \\
\Delta(Q_m) &\leq pz + (m-2) + 2 = pz + m\,. \qquad (1)
\end{aligned}$$

In addition, the following inequalities also hold for any $1 \leq m \leq |V|$:

$$\begin{aligned}
\binom{|V_m|}{2} &= \frac{1}{2}(qz - m)(qz - m - 1) \\
&= \frac{1}{2}\left(q^2 z^2 - (2m+1)qz + m^2 + m\right) \\
&> \frac{1}{2}\left(qz(qz - 2m - 1)\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\big(qz(pz + (q-p)z - 2m - 1)\big) \\
&> \frac{1}{2}(qpz^2) > \frac{p}{q}\frac{qz(qz-1)}{2} - \binom{m}{2}
\end{aligned}
$$

$$
\binom{|V_m|}{2} > \frac{p}{q}\binom{|V_m|+m}{2} - \binom{m}{2}\ . \tag{2}
$$

We note that $\frac{p}{q}\binom{|V_m|+m}{2} - \binom{m}{2}$ is integer, as $|V_m| + m = q(4qk+2)|V|^2$ is a multiple of $2q$. Starting with this, we continue to show the next inequality:

$$
\begin{aligned}
\frac{p}{q}\binom{|V_m|+m}{2} - \binom{m}{2} &= \frac{1}{2}\Big(\frac{p}{q}qz(|V_m| + m - 1) - m(m-1)\Big) \\
&= \frac{1}{2}\big(pz|V_m| + pz(m-1) - m(m-1)\big) \\
&\geq \frac{1}{2}\big(z|V_m| + (m-1)(z-m)\big) \\
&\geq \frac{1}{2}z|V_m| = (2qk+1)|V|^2|V_m|\ . \qquad (3)
\end{aligned}
$$

With the inequality (2) we guarantee that Q_m can accommodate the number of edges we desire to place. In inequality (3) we ensure that the number of edges in E_m is sufficient to construct a $z = (4qk+2)|V|^2$ regular graph among all vertices in V_m (before we add the last $|E_m| \mod |V_m|$ edges during the construction of Q_m). Therefore, each vertex $v_i \in V_m$ is connected to the preceding $\frac{z}{2}$ and the next $\frac{z}{2}$ vertices in cyclic order.

Now we construct $G' = G \cup Q_k \cup \bigcup_{i=1}^{|E|-\binom{k}{2}} Q_2^i$ where each Q_2^i is a separate instance of Q_2. Observe that the graphs G, Q_k and Q_2^i are not connected to each other and that $|V_k| < |V_2^i|$ when $k > 2$. Also note that the inequality (1) implies $\Delta(G') = \max(\Delta(G), \Delta(Q_2), \Delta(Q_k)) \leq pz + k$.

Proposition 1. *In G' there exists no $\frac{p}{q}$-clique with $qz+2qk+1$ or more vertices.*

Proof. Consider the number of edges in a $\frac{p}{q}$-clique T with $|V(T)| = t \geq qz + 2qk + 1$ vertices:

$$
\begin{aligned}
|E(T)| \geq \frac{p}{q}\binom{t}{2} &= \frac{1}{2}\Big(\frac{p}{q}t(t-1)\Big) \\
&\geq \frac{1}{2}\Big(\frac{p}{q}t(qz + 2qk)\Big) \\
&= \frac{1}{2}t(pz + 2pk) \\
&> \frac{1}{2}t(pz + k)
\end{aligned}
$$

using the bound (1) on the maximum degree, we obtain

$$
|E(T)| > \frac{1}{2}t\Delta(G')\ . \tag{4}
$$

Hence, no subgraph of G' with $t \geq qz + 2qk + 1$ vertices can achieve the edge density required to form a $\frac{p}{q}$-clique. □

Proposition 2. *Let $H = Q_m \cup R_t$ where R_t is an arbitrary graph with t vertices whose vertex set $V(R_t)$ is disjoint from $V(Q_m)$. Then, the following three statements hold:*

1. *If $0 \leq t < m$, then H is a $\frac{p}{q}$-clique.*
2. *If $t = m$, then H is a $\frac{p}{q}$-clique if and only if R_t is a clique with t vertices.*
3. *If $m < t \leq |V|$, then H is not a $\frac{p}{q}$-clique.*

Proof. Let $\delta = m - t$, in which case we obtain $|V(H)| = qz - \delta$. Now we consider the cardinality of $E(H)$:

$$\begin{aligned}
|E(H)| &= |E(Q_m)| + |E(R_t)| \\
&= \frac{p}{q}\binom{qz}{2} - \binom{m}{2} + |E(R_t)| \\
&= \frac{p}{2q}(qz(qz-1)) - \binom{m}{2} + |E(R_t)| \\
&= \frac{p}{2q}((qz-\delta)(qz-\delta-1) + 2\delta qz - \delta(\delta+1)) - \binom{m}{2} + |E(R_t)| \\
&= \frac{p}{q}\binom{qz-\delta}{2} + 2\delta pz - \frac{p}{2q}(\delta(\delta+1)) - \binom{m}{2} + |E(R_t)|
\end{aligned}$$

If $t < m$ as in the first case of the proposition, that is $\delta > 0$, we obtain $|E(H)| > \frac{p}{q}\binom{qz-\delta}{2}$; because $2\delta pz > z = (4qk+2)|V|^2 > \frac{p}{2q}(\delta(\delta+1)) + \binom{m}{2} - |E(R_t)|$ trivially holds; hence H is a $\frac{p}{q}$-clique.

If $m < t \leq |V|$ as in the third case, that is $-|V| \leq \delta < 0$, we have $2\delta pz < \frac{p}{2q}(\delta(\delta+1)) + \binom{m}{2} - |E(R_t)|$; therefore, $|E(H)| < \frac{p}{q}\binom{qz-\delta}{2}$ and H is not a $\frac{p}{q}$-clique.

When $t = m$ as in the second case, we have $\delta = 0$. It holds that $|E(H)| = \frac{p}{q}\binom{qz}{2} - \binom{m}{2} + |E(R_t)|$; which implies that H is a $\frac{p}{q}$-clique if and only if $|E(R_t)| \geq \binom{m}{2}$. Since R_t has $t = m$ vertices, the constraint is satisfied only when R_t is a complete graph. □

Proposition 3. *For any $m \geq 1$, there exists no vertex separator S of Q_m with $|S| < \frac{z}{2}$, which partitions V_m into disjoint vertex sets A, B, S with $|A|, |B| > 0$ so that there is no edge $(v_a, v_b) \in E_m$ where $v_a \in A$ and $v_b \in B$.*

Proof. We assume for the sake of contradiction that there exists such a vertex partition of V_m where $|S| < \frac{z}{2}$. Then, we pick any $v_{p_1} \in A$ as the first *pivot* vertex. Next, we consider the set $N_{p_1} = \{v_{(p_1+1) \bmod (|V_m|)}, \ldots, v_{(p_1+z/2) \bmod (|V_m|)}\}$ of next $\frac{z}{2}$ vertices in cyclic order connected to v_{p_1}. Since $|S| < \frac{z}{2} = |N_{p_1}|$, we have $R_{p_1} = N_{p_1} - S \neq \emptyset$ which must only consist of vertices from A; since the existence of any vertex in $R_{p_1} \cap B$ contradicts S being a vertex separator. Next, we pick another pivot vertex v_{p_2} from $R_{p_1} \subset A$, then recursively apply the same argument on v_{p_2}, N_{p_2} and R_{p_2} to scan the neighbors of v_{p_2} in N_{p_2}. We continue this process until we eventually obtain a pivot v_{p_t} for which we get $v_{p_1} \in N_{p_t}$; which implies the completion of a cyclic scan of all vertices of V_m. In other words, we obtain $N_{p_0} \cup N_{p_1} \cup \ldots \cup N_{p_t} = V_m$. This can be seen by observing that $|N_{p_i}| = \frac{z}{2}$ while $|p_{i+1} - p_i| \leq \frac{z}{2}$ for all $1 \leq i < t$. Therefore, in this case we have $(N_{p_1} \cap B) \cup \ldots \cup (N_{p_t} \cap B) = V_m \cap B = B = \emptyset$ which is a contradiction. Hence, the assumption is false and $|S| \geq \frac{z}{2}$ must hold.

□

Proposition 4. *In a cover by $\frac{p}{q}$-cliques of G' with $c = |E| - \binom{k}{2} + 1$ many $\frac{p}{q}$-cliques, the vertices of each constituting subgraph Q_k and Q_2^i for $i = 1, \ldots, c-1$ must reside in a separate $\frac{p}{q}$-clique.*

Proof. Let $C_1, \ldots, C_c$ be the set of $\frac{p}{q}$-cliques in the cover so that $C_1 \cup \ldots \cup C_c = G'$. Let $Q \in \{Q_2^1, \ldots, Q_2^{c-1}, Q_k\}$ be a subgraph of G' whose vertices are not contained in a single $\frac{p}{q}$-clique. Let $S_i = C_i \cap Q$ be the subgraph of C_i partially covering Q for $1 \le i \le c$. Suppose that $|V(S_t)| = \max_{1 \le i \le c} |V(S_i)|$, i.e., C_t is the quasi-clique which possesses the largest subgraph of Q. Then, we assume contrary that there exists a quasi-clique cover of G' in which $|V(S_t)| < |V(Q)|$. Let $R_t = Q - S_t$ be the edge-induced subgraph of Q containing the edges of Q that are to be covered by the quasi-cliques except C_t. Clearly, we have $V(R_t), E(R_t)$ and $V(R_t) - V(S_t)$ non-empty by our assumption.

We first analyze the case when $|V(S_t)| \ge \frac{z}{2}$. Consider the vertex set $I = V(S_t) \cap V(R_t)$. If $I = V(S_t)$, we have $|I| = |V(S_t)| \ge \frac{z}{2}$ holds. Otherwise we have $I \subset V(S_t)$, which implies that $V(S_t) - I$ and $V(R_t) - I$ are disjoint and non-empty sets. Moreover, there exists no edge $(v_s, v_r) \in E(Q)$ such that $v_s \in V(S_t) - I$ and $v_r \in V(R_t) - I$; because, such an edge is contained neither in S_t nor in R_t, which contradicts $S_t \cup R_t = Q$. Hence we obtain a vertex partition $\{V(S_t) - I, V(R_t) - I, I\}$ of $V(Q)$, where each set is non-empty and there is no edge in $E(Q)$ between the vertices in the first two sets. Therefore I is a vertex separator and by Proposition 3 it has cardinality $|I| \ge \frac{z}{2}$. We thus have $|I| \ge \frac{z}{2}$ satisfies whenever $|V(S_t)| < |V(Q)|$ and $|V(S_t)| \ge \frac{z}{2}$. Now, we first note that in a $\frac{p}{q}$-clique cover of G', vertices of each subgraph in $\{Q_2^1, \ldots, Q_2^{c-1}, Q_k\}$ must appear at least once in any of the quasi-cliques; since they have nonzero vertex degrees. Moreover, we have $|V(S_t)|$ vertices of Q in C_t to cover S_t. In the rest of the quasi-cliques, we similarly need at least $|V(R_t)| = |V(Q) - V(S_t) + I| = |V(Q)| - |V(S_t)| + |I|$ vertices to cover $R_t = Q - S_t$. Thus, covering Q requires at least $|V(S_t)| + |V(R_t)| = |V(Q)| + |I|$ vertices; which implies at least $|I| \ge \frac{z}{2}$ vertex repetitions. Hence, we have the total number of vertices in all quasi-cliques, V_{tot}, lower-bounded by:

$$\begin{aligned} V_{tot} &\ge \sum_{i=1}^{c-1} |V(Q_2^i)| + |V(Q_k)| + |I| \\ &\ge \sum_{i=1}^{c-1} |V(Q_2^i)| + |V(Q_k)| + \frac{z}{2} \qquad (5) \\ &= \left(|E| - \binom{k}{2}\right)(qz-2) + (qz-k) + (2qk+1)|V|^2 \\ &> \left(|E| - \binom{k}{2} + 1\right) qz - 2\left(|E| - \binom{k}{2}\right) - k + (2qk+1)2|E| \\ &> \left(|E| - \binom{k}{2} + 1\right) qz + (2qk+1)|E| \\ &\ge \left(|E| - \binom{k}{2} + 1\right)(qz + 2qk + 1) = c(qz + 2qk + 1) \end{aligned}$$

This requires the existence of a quasi-clique C_i with $|V(C_i)| \ge (qz+2qk+1)$, which contradicts Proposition 1.

Next, we consider the case when $|V(S_t)| < \frac{z}{2}$. Now we have $\max_{1 \leq i \leq c} \{\max_{v \in V(S_i)} degree(v)\} < \frac{z}{2}$; since the cardinality of $V(S_i)$ is upper-bounded by $|V(S_t)| < \frac{z}{2}$. Note also that since Q is either Q_k or an instance of Q_2, we have $|E(Q)| \geq \min(|E_2|, |E_k|) = \frac{p}{q}\binom{qz}{2} - \binom{k}{2}$. As a result, we obtain T, the total number of vertices used to cover all edges in $E(Q)$ in all quasi-cliques, to be lower-bounded by:

$$\begin{aligned} T &> \frac{2|E(Q)|}{z/2} \\ &\geq \frac{2\left(\frac{p}{q}\binom{qz}{2} - \binom{k}{2}\right)}{z/2} \\ &\geq \frac{2\big(pz(qz-1) - k(k-1)\big)}{z} \\ &\geq \frac{2\big(pz(qz-2) + pz - k(k-1)\big)}{z} \\ &> 2(qz-2) > (qz-2) + \frac{z}{2} \geq |V(Q)| + \frac{z}{2} \, . \end{aligned}$$

That is, we need at least $\frac{z}{2}$ vertex repetitions to cover the edges of Q, which yields to the same contradiction as in (5). □

We are now ready to prove the main theorem of this section.

Proof of Theorem 1. We can easily show that the CoverBy-$\frac{p}{q}$-Cliques $\in$ NP. Let us consider an instance of Clique problem, that is a graph G and an integer k. We build an instance of CoverBy-$\frac{p}{q}$-Cliques with G' and c as defined at the beginning of this section. We now show that G has a clique of size k if and only if G' has a cover of size c by $\frac{p}{q}$-cliques. Suppose that G has a clique of size k. We can then construct a cover by $\frac{p}{q}$-cliques of G' with c quasi-cliques $\mathcal{C} = \{C_1, \ldots, C_c\}$ by putting the k-clique of G together with Q_k in C_c, and each of the remaining $|E| - \binom{k}{2}$ edges (or cliques of two vertices) together with an arbitrary Q_2^i in C_i, for which each C_i and C_c becomes a $\frac{p}{q}$-clique by Proposition 2. Suppose now that G' has a cover of size c by $\frac{p}{q}$-cliques. We have shown in Proposition 4 that in such a cover of G', each subgraph Q_2^i as well as Q_k must reside in a separate $\frac{p}{q}$-clique. Proposition 2 implies that each quasi-clique containing one of Q_2^i can cover at most one edge (or 2-clique) from G, leaving at least $\binom{k}{2}$ edges for the remaining $\frac{p}{q}$-clique containing Q_k. Since Proposition 2 also implies that any $\frac{p}{q}$-clique containing Q_k cannot contain another subgraph with vertex count $k < t \leq |V|$, the remaining $\binom{k}{2}$ edges of G must belong to a subgraph with k vertices; hence form a clique of size k. This finalizes the proof of NP-completeness for the CoverBy-$\frac{p}{q}$-Cliques problem. □

Pattillo et al. [7, Corollary 1] discuss how to find a p and a q satisfying $1 \leq p < q$ for a given γ so that any γ-clique is a $\frac{p}{q}$-clique and vice versa. Using this result, the NP-completeness of the decision version of the minimum cover by γ-cliques problem follows as a corollary to Theorem 1, which we give below for completeness.

Corollary 1. *The decision version of the minimum cover by γ-cliques problem is NP-complete for any fixed γ satisfying $0 < \gamma < 1$.*

2.2 Minimum partition by $\frac{p}{q}$-cliques problem

We consider the following decision problem associated with the minimum partition by $\frac{p}{q}$-cliques problem.
PartitionBy-$\frac{p}{q}$-Cliques: Given a simple undirected graph $G = (V, E)$, fixed integers p and q satisfying $1 \le p < q$, and a positive integer c, is there a partition of V into a set $\mathcal{C}$ of c many $\frac{p}{q}$-cliques such that for each $v \in V$ there exists a unique $C \in \mathcal{C}$ such that $v \in V(C)$?

Theorem 2. *The* PartitionBy-$\frac{p}{q}$-Cliques *problem is NP-complete for any positive integer constants p and q satisfying $1 \le p < q$.*

We use a similar technique to the one in the proof of Theorem 1 by constructing a graph $G'' = (V'', E'')$ and an integer c for a given graph $G = (V, E)$ and $2 \le k \le |V|$ in a way that G has a clique of size k if and only if G'' has a partition of its vertices into $c = (|V| - k + 1)$ many $\frac{p}{q}$-cliques.

We now similarly construct a graph $G'' = G \cup Q_k \cup \bigcup_{i=1}^{(|V|-k)} Q_1^i$. where each Q_1^i is a separate instance of Q_1.

Proposition 5. *In a vertex partition of G'' into $\frac{p}{q}$-cliques $\mathcal{C} = \{C_1, \ldots, C_c\}$, the total number of edges in all $\frac{p}{q}$-cliques, $E_{tot} = \sum_{i=1}^{c} |E(C_i)|$, is lower-bounded by $c\frac{p}{q}\binom{qz}{2}$.*

Proof. First note the fact that $|V''| = |V| + |V_k| + \sum_{i=1}^{c-1} |V_1| = cqz$. Let $s_i = |V(C_i)|$ denote the number of vertices in C_i. Since each C_i is a $\frac{p}{q}$-clique, the total number of edges in all quasi-cliques is lower-bounded by:

$$E_{tot} \ge L = \sum_{i=1}^{c} \frac{p}{q}\binom{s_i}{2} .$$

Since $V(C_1)+\cdots+V(C_c)$ is a partition of V'', we also have $\sum_{i=1}^{c} s_i = |V''| = cqz$. In this case, L is minimum when $s_i = qz$ for all $1 \le i \le c$ by Cauchy-Schwarz inequality, which yields

$$E_{tot} \ge L = \sum_{i=1}^{c} \frac{p}{q}\binom{qz}{2} = c\frac{p}{q}\binom{qz}{2}.$$

hence completes the proof. □

Proposition 6. *Let v_a, v_b be two vertices in Q_m for any $m \ge 2$ such that $\{v_a, v_b\} \in E_m$. Then, there are at least $\frac{z}{2} - 1$ vertices in $adj(v_a) \cap adj(v_b)$.*

Proof. Assume without loss of generality that v_a precedes its neighbor v_b in cyclic order, and that the vertices are enumerated in such a way that $a < b$. Since $\{v_a, v_b\} \in E_m$, we have $b - a \le \frac{z}{2} + 1$ as we add the edges $\{v_a, v_{a+1}\}$, $\{v_a, v_{a+2}\}, \ldots, \{v_a, v_{a+z/2}\}$ in the first $\left\lfloor \frac{|E_m|}{|V_m|} \right\rfloor$ passes of the construction, and possibly add one more edge in the last pass. If $\{v_a, v_b\}$ is added to E_m in the last pass (where we add $|E_m| \mod |V_m| \neq 0$ edges), all vertices between v_a and $v_b = v_{a+z/2+1}$ are connected to both v_a and v_b. The number of such vertices is $\frac{z}{2}$. If we did not add any edges containing v_a in the last pass, then v_b is among the vertices $\{v_{a+1}, \ldots, v_{a+z/2}\}$, and therefore other vertices $v_u \neq v_b$ for $a < u \le a + \frac{z}{2}$ are connected to both v_a and v_b. There are $\frac{z}{2} - 1$ such vertices. □

We note that $\frac{z}{2}-1$ in the proof above is tight and holds for a v_a and $v_{a+z/2}$ when we did not add the edge $\{v_a, v_{a+z/2+1}\}$ in the last pass.

Proposition 7. *In a $\frac{p}{q}$-clique vertex partition of G'' with c many $\frac{p}{q}$-cliques, the vertices of each constituting subgraph Q_1^i for $i = 1, \ldots, (|V| - k)$ and Q_k must reside in a separate $\frac{p}{q}$-clique.*

Proof. Proposition 6 implies that in case a Q_m is split, there are at least $\frac{z}{2}-1$ edges in the cut between split sets of vertices. By definition, no vertex can belong to more than one quasi-clique in a $\frac{p}{q}$-clique vertex partition of G''. This implies not only that the cut-edges cannot belong to any $C_i \in \mathcal{C}$, but also for each of the remaining $|E''| - \frac{z}{2} + 1$ edges there can be at most one quasi-clique C_i containing it; otherwise the corresponding vertices need to appear more than once in quasi-cliques. Therefore, in this case the total number of edges in all quasi-cliques, E_{tot} is upper-bounded by:

$$E_{tot} \leq |E| + |E_k| + \sum_{i=1}^{c-1} |E_1| - \frac{z}{2} + 1$$
$$= |E| + \frac{p}{q}\binom{qz}{2} - \binom{k}{2} + \sum_{i=1}^{c-1} \frac{p}{q}\binom{qz}{2} - \frac{z}{2} + 1$$

since the negative terms are strictly larger in absolute value than $|E|+1$

$$E_{tot} < c\frac{p}{q}\binom{qz}{2}$$

which contradicts Proposition 5. Thus, no Q_m may be split into different quasi-cliques in a $\frac{p}{q}$-clique vertex partition of G'' with c many quasi-cliques. □

Now we continue with the proof of Theorem 2.

Proof of Theorem 2. We can easily show that PARTITIONBY-$\frac{p}{q}$-CLIQUES $\in$ NP. Now we consider an instance of CLIQUE problem with a graph G and an integer k, and construct an instance of PARTITIONBY-$\frac{p}{q}$-CLIQUES with G'' and c as described. We claim that G has a clique of size k if and only if G'' has a $\frac{p}{q}$-clique vertex partition $\mathcal{C} = \{C_1, \ldots, C_c\}$ of size c. Suppose that G has a clique of size k. In this case, we can construct c many $\frac{p}{q}$-cliques which cover V'' by putting the vertices in k-clique of G together with Q_k in a quasi-clique C_c, and each of the remaining $|V| - k$ vertices together with an arbitrary Q_1^i in C_i. Proposition 2 implies that each C_i constructed this way is a $\frac{p}{q}$-clique, and clearly they partition V''. Now suppose that G'' has a vertex partition by c many $\frac{p}{q}$-cliques. Proposition 5 implies that each Q_1^i must be contained in a separate quasi-clique, say C_i, and Proposition 2 implies that each such C_i can contain at most one vertex from V, thereby leaving at least $t \geq k$ vertices for the last quasi-clique, say C_c, which contains Q_k. Since Proposition 2 also implies that C_c cannot contain a graph with $k < t \leq |V|$ vertices, not only t must equal to k but also these t vertices must form a complete graph by Proposition 2, which finalizes the proof of Theorem 2. □

Using again the result of Pattillo et al., the NP-completeness of the decision version of the minimum partition by γ-cliques problem follows as a corollary to Theorem 2, which we give below for completeness.

Corollary 2. *The decision version of the minimum partition by γ-cliques problem is NP-complete for any fixed γ satisfying $0 < \gamma < 1$.*

3 Conclusion

We have shown that the decision versions of the minimum cover and partition by γ-quasi-cliques problems are NP-complete for any fixed γ satisfying $0 < \gamma < 1$. Kaya et al. [5] discuss an application of the minimum cover by quasi-cliques problem in sparse matrix ordering methods. Matsuda et al. [6] discuss and Blanchette et al. [3] mention the use of quasi-clique covers in bioinformatics applications. Cliques and quasi-cliques also arise in various problems of network analysis (see, for example [1, 2]), where the parameter γ is used to identify similar entities with a guaranteed coherence. The associated covering and partitioning problems treated in this work can therefore find applications in such settings. We leave the investigation of the use of covers and partitions by quasi-cliques in network analysis as a future work.

Bibliography

[1] J. Abello, M. G. C. Resende, and S. Sudarsky. Massive quasi-clique detection. In S. Rajsbaum, editor, *LATIN 2002: Theoretical Informatics*, volume 2286 of *Lecture Notes in Computer Science*, pages 598–612. Springer Berlin Heidelberg, 2002.

[2] M. Behrisch and A. Taraz. Efficiently covering complex networks with cliques of similar vertices. *Theoretical Computer Science*, 355(1):37–47, 2006.

[3] M. Blanchette, E. Kim, and A. Vetta. Clique cover on sparse networks. In D. A. Bader and P. Mutzel, editors, *2012 Meeting on Algorithm Engineering & Experiments, ALENEX 2012*, Kyoto, Japan, January 2012. SIAM / Omnipress.

[4] M. R. Garey and D. S. Johnson. *Computers and Intractabilityd: A Guide to the Theory of NP-Completeness.* W. H. Freeman & Co., New York, NY, USA, 1979.

[5] O. Kaya, E. Kayaaslan, B. Uçar, and I. S. Duff. On hypergraph-partitioning based fill reducing methods, 2013. Working paper.

[6] H. Matsuda, T. Ishihara, and A. Hashimoto. Classifying molecular sequences using a linkage graph with their pairwise similarities. *Theoretical Computer Science*, 210(2):305–325, 1999.

[7] J. Pattillo, A. Veremyev, S. Butenko, and V. Boginski. On the maximum quasi-clique problem. *Discrete Applied Mathematics*, 161(1–2):244–257, 2013.

RESEARCH CENTRE
GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399