Universal-Clique Amplification and Hardness of Assignment-Optimal Edge Clique Covers*

Matěj Kripner Václav Rozhoň Robert Šámal
Martin Šindelář Adrián Zámečník

hexagon:2611.00053v1

Abstract

An edge clique cover has assignment cost equal to the sum of its clique sizes. Unless $P = NP$, this cost admits no polynomial-time $O(N^{1/2-\eta})$-approximation for any fixed $0 < \eta < 1/2$, even on connected N-vertex graphs. A universal-clique join transfers coloring hardness without the additive obstruction in earlier reductions. Its exact residual formula also exhibits a change of optimal cover when a universal twin is added. Explicit counterexamples refute two proposed twin-extension rules, while weighted compression of true-twin classes preserves the optimum.

1 Introduction

1.1 Background and the source problem

An edge clique cover compresses a graph by storing complete subgraphs rather than individual edges. Counting the stored cliques is only one measure of its size: a large clique requires more vertex memberships than a small one. For a clique family $\mathcal{C}$, the corresponding assignment cost and optimum are

$$\operatorname{size}(\mathcal{C}) = \sum_{C \in \mathcal{C}} |C|, \qquad A(G) = \min_{\mathcal{C} \text{ covering } E(G)} \operatorname{size}(\mathcal{C}).$$

Singleton cliques are allowed; an edgeless graph has optimum zero and is solved by the empty cover. Approximation ratios are stated for graphs with at least one edge.

Ullah and Pothen call a cover minimizing this quantity *assignment-optimal* [15, Def. 3.6]. Their dual clique-cover representation stores the vertex–clique incidence relation in both directions [15, Def. 3.15]. Double-counting these incidences gives representation size $2\operatorname{size}(\mathcal{C})$, so assignment cost controls the storage objective directly. This optimization problem also appears as *assignment-minimum clique cover* [6] and *Sigma Clique Cover* [7, Def. 3.1].

The distinction between the two objectives matters for complexity. Kou, Stockmeyer, and Wong related cardinality-minimum edge clique cover to coloring [10]; the coloring hardness of Lund and Yannakakis [12] consequently yields hardness for that objective, as reviewed in [15, Appendix A.2]. Exact NP-hardness for assignment cost is established by Ullah [14, Thm. 1.3 and Sec. 3.4] and by Firbas et al. [7, Lem. 3.4 and Thm. 3.5]. Those reductions add mutually nonadjacent vertices, each adjacent to the base graph. Their cost includes a repeated contribution from every base vertex. Ullah and Pothen calculate explicitly that a ρ-approximation then gives at most $(\rho - 1)n + \rho\chi(X) + O(\rho)$ colors for an n-vertex input X [15, Lem. A.2 and the following discussion]. The term $(\rho - 1)n$ prevents the required multiplicative transfer.

Ullah and Pothen therefore leave constant-factor approximability open in Section 3.2, repeat this status in Appendix D.1, and include the objective among the directions in Section 7 [15]. The explicit question is the following.

*This work was produced within the Bolzano project (https://bolzano.app), using AI systems for mathematical research and manuscript preparation. The listed authors contributed to the project and the preparation of this release. The manuscript received only limited human checking.

Question 1.1 (Ullah–Pothen)**.** Does the assignment-optimal edge clique cover problem admit a polynomial-time constant-factor approximation? [15, Sec. 3.2]

1.2 Contribution

The reduction replaces the independent added vertices by a clique U of size L. If S is a clique in the base graph H, a single clique $U \cup S$ can now serve all vertices of U while paying for the vertices of S only once. In the opposite direction, the cover cliques containing any fixed $u \in U$, intersected with H, cover every vertex of H. Thus each universal vertex still pays for a vertex clique cover, although the base-vertex assignments can be shared. Writing $\theta(H) = \chi(\overline{H})$, this gives

$$L\theta(H) \le A(K_L \vee H) \le L\theta(H) + n^2, \qquad n = |V(H)|.$$

Taking $L = n^2$ controls the remaining internal-edge cost. The resulting graph has $N = n + n^2$ vertices, which explains the exponent $1/2$ in the hardness bound.

The results are organized as follows.

(i) Theorem 3.1 converts a $\rho(N)$-approximation for assignment cost into a $2\rho(n + n^2)$-approximation for vertex clique cover. Together with Zuckerman's chromatic-number hardness [16, Thm. 1.2], Corollary 3.2 excludes $O(N^{1/2-\eta})$ approximation for every fixed $0 < \eta < 1/2$ unless $P = NP$, even for connected graphs without isolated vertices. This answers Question 1.1 negatively.

(ii) Theorem 4.2 determines the exact assignment cost of the join as $\min_{q \ge \theta(H)}(Lq + R_q(H))$, where the marked residual $R_q(H)$ measures the base-graph assignments needed when at most q cliques must also cover all base vertices. Corollary 4.3 and the triangle-free examples describe this residual explicitly.

(iii) Proposition 4.5 gives a six-vertex base whose optimal residual changes with the number of marked cliques. Applied to twin extensions, this yields a counterexample to [14, Rule 3.26]; a separate six-vertex example refutes Rule 3.27. Proposition 5.4 gives an exact weighted compression for true-twin classes.

The issue with the twin rules is that changing the number of copies of a vertex changes the relative costs of competing covers. Synchronizing twins inside a fixed graph is valid, but extending an arbitrary optimum of the smaller graph need not preserve optimality. The marked-residual formula makes this distinction explicit. These counterexamples concern the two stated rules in the 2022 preprint, not its separate NP-hardness reduction.

The quantitative approximation threshold remains open. In particular, the hardness statement does not exclude $\sqrt{N}/\log N$ or other factors of the form $N^{1/2-o(1)}$. The general edge-by-edge upper bound is linear; Theorem 6.2 and Corollary 6.4 give a logarithmic bound on graph classes with polynomially many listable maximal cliques. These restricted-class algorithms are compatible with hardness on arbitrary graphs.

1.3 Related work and the scope of the auxiliary results

The same sum-of-clique-sizes objective has a substantial extremal literature under the notation $\mathrm{scc}(G)$. Davoodi, Javadi, and Omoomi [3, Thm. 2(i)] prove $2|E(G)|/(\omega(G)-1) \le \mathrm{scc}(G) \le 2|E(G)|$. Consequently the entire $(\omega(G) - 1)$ edge-by-edge guarantee in Proposition 6.1, including its linear specialization, is already contained in their result. The degeneracy bound follows by $\omega(G) \le d + 1$. Davoodi, Gerbner, Methuku, and Vizer [4, Thms. 6–7 and Cor. 8] study the same objective on complete multipartite graphs, obtaining asymptotically matching upper and lower bounds of order $t \log t$ when the number t of parts grows and each part has fixed size. Those extremal results concern the value of the optimum on specified families, rather than a polynomial-time approximation ratio for arbitrary graphs.

Recent work also treats a different valency objective. The local clique cover number minimizes the *maximum* number of cover cliques containing one vertex, whereas assignment cost minimizes the *sum* of these numbers. Davoodi's September 2026 preprint [5, Thms. 1.1–1.2] bounds the local parameter in terms of chromatic number. Its Lemma 4.1 extends a cover through a universal clique by marking occurrences selected from a maximum matching. This is the matching extension used in the triangle-free example below and already implies that example's upper bound. The local-valency results do not state approximation hardness for the sum objective or the exact marked-residual minimization formula.

Practical clique-cover algorithms address a complementary question. Ennis, Fayle, and Ennis [6] study the assignment objective; Abdullah and Hossain [1, Sec. 5] discuss post-processing for assignment minimization after computing covers aimed at small clique cardinality. The worst-case reduction here neither evaluates nor precludes successful compression on such practical instances.

True-twin classes are also called critical cliques [7, Def. 5.2]. Firbas et al. give a valid shrinking rule when the neighborhood of the class in the quotient graph is a clique [7, Lem. 5.5]. The universal class over CP_3 does not satisfy that hypothesis. The counterexamples below concern the unconditional extensions in [14, Rules 3.26–3.27]; the weighted quotient identity follows from an elementary normalization inside the original graph.

The greedy analysis is the classical weighted set-cover argument: Johnson and Lovász established the unweighted guarantees [9, 11], and Chvátal treated positive set costs [2]. The specific implementation in Lemma 6.3 lists maximal cliques and optimizes over their subsets using maximum-density subgraph, which is polynomial-time solvable by flow [8]. It is essential to optimize over subsets: a maximal clique may contain many vertices but few currently uncovered edges. For chordal graphs, perfect elimination supplies the required short maximal-clique list [13]. Thus the greedy guarantee and its flow primitive are standard. The join amplification, its exact marked residual, and the consequences for twin extensions form the main argument.

2 Definitions and the join gadget

All graphs are finite, simple, and undirected. Join statements below assume that the base graph H has at least one vertex; this harmless convention is necessary because an empty base would leave the edges of K_L outside the marked-family description. For a graph H, let

$$\theta(H)$$

denote the minimum number of cliques whose union of vertices is $V(H)$; this is the vertex clique cover number. Equivalently, $\theta(H) = \chi(\overline{H})$.

Definition 2.1 (Universal-clique join)**.** For an integer $L \geq 1$, define

$$J_L(H) = K_L \vee H.$$

The vertex set is the disjoint union $U \cup W$, where U induces a clique of size L, $W = V(H)$, all edges between U and W are present, and the graph induced by W is H.

The key point is that each universal vertex sees a complete copy of the vertex clique cover problem for H.

Lemma 2.2 (Extraction from a universal vertex)**.** *Let H be nonempty, let $G = J_L(H)$, and let $\mathcal{C}$ be any edge clique cover of G. One can extract from $\mathcal{C}$ a vertex clique cover of H with at most* $\mathrm{size}(\mathcal{C})/L$ *cliques. Consequently,*

$$A(J_L(H)) \geq L\theta(H).$$

Proof. Index the cliques of $\mathcal{C}$ as $C_1, \ldots, C_t$. For a universal vertex $u \in U$, let

$$I_u = \{i : u \in C_i \text{ and } C_i \cap W \neq \emptyset\}.$$

For each $i \in I_u$, the intersection $C_i \cap W$ is a clique of H, because $G[W] = H$. These intersections cover every vertex of H: for $w \in W$, the edge uw must be covered by some clique containing both u and w, and that clique contributes an intersection containing w.

Thus $(C_i \cap W)_{i \in I_u}$ is a vertex clique cover of H, so $|I_u| \geq \theta(H)$. Summing over all $u \in U$ gives

$$\sum_{u \in U} |I_u| = \sum_{i : C_i \cap W \neq \emptyset} |C_i \cap U| \leq \sum_{i=1}^{t} |C_i| = \text{size}(\mathcal{C}).$$

Therefore some $u^* \in U$ has $|I_{u^*}| \leq \text{size}(\mathcal{C})/L$, and its intersections give the desired extracted cover. Since every $|I_u| \geq \theta(H)$, the same displayed inequality also gives $\text{size}(\mathcal{C}) \geq L\theta(H)$ for every $\mathcal{C}$. □

Lemma 2.3 (Upper bound for the join)**.** *If H has $n \geq 1$ vertices and m edges, then*

$$A(J_L(H)) \leq L\theta(H) + n + 2m \leq L\theta(H) + n^2.$$

Proof. Let $T_1, \ldots, T_\theta$ be a minimum vertex clique cover of H, where $\theta = \theta(H)$. Assign each vertex $w \in V(H)$ to one clique T_i that contains it, and let S_i be the set assigned to T_i. Each nonempty S_i is a clique, and no S_i is empty, since otherwise the nonempty assigned sets would form a vertex clique cover using fewer than θ cliques. Hence $S_1, \ldots, S_\theta$ partition $V(H)$ into θ cliques.

For each i, include the clique $U \cup S_i$ in the edge clique cover of $J_L(H)$. These cliques cover all edges inside U and all edges between U and W. Cover every remaining internal edge of H by its two endpoints. The total assignment cost is at most

$$\sum_{i=1}^{\theta} (L + |S_i|) + 2m = L\theta + n + 2m \leq L\theta + n^2.$$

□

3 Approximation hardness

The reduction is now immediate. The additive residual in Lemma 2.3 is made negligible by taking $L = n^2$.

Theorem 3.1 (Approximation transfer)**.** *Suppose there is a polynomial-time algorithm which, on every connected N-vertex graph G, outputs an edge clique cover of assignment cost at most $\rho(N)A(G)$. Then there is a polynomial-time $2\rho(n + n^2)$-approximation algorithm for vertex clique cover on n-vertex graphs.*

Proof. Given an n-vertex graph H, set $L = n^2$ and $G = J_L(H)$. Then G has $N = n + n^2$ vertices. Run the assumed approximation algorithm on G, obtaining a cover $\mathcal{C}$ with

$$\text{size}(\mathcal{C}) \leq \rho(N)A(G).$$

The graph G is connected and has no isolated vertices, so it belongs to the promised input class. By Lemma 2.2, extract from $\mathcal{C}$ a vertex clique cover of H with at most $\text{size}(\mathcal{C})/L$ cliques. Lemma 2.3 gives

$$A(G) \leq L\theta(H) + n^2 = L(\theta(H) + 1).$$

Therefore the extracted vertex clique cover has size at most

$$\frac{\rho(N)A(G)}{L} \leq \rho(N)(\theta(H) + 1) \leq 2\rho(N)\theta(H),$$

because $\theta(H) \geq 1$ for nonempty H. This is the claimed approximation ratio. □

Corollary 3.2 (Nearly square-root hardness)**.** *For every fixed $0 < \eta < 1/2$ and every constant $C > 0$, there is no polynomial-time $CN^{1/2-\eta}$-approximation for assignment-optimal edge clique cover on connected N-vertex graphs, unless $P = NP$. Equivalently, no polynomial-time $O(N^{\delta})$-approximation exists for any fixed $0 \le \delta < 1/2$, unless $P = NP$. In particular, no polynomial-time constant-factor approximation exists unless $P = NP$.*

Proof. If $H = \overline{X}$, then $\theta(H) = \chi(X)$. A vertex clique cover of H is exactly a coloring of X: assign each vertex to one clique containing it, and the assigned clique index is a color because clique vertices in H are independent in X.

Fix $\delta < 1/2$ and suppose an assignment-cover approximation with ratio at most cN^{δ} existed for some fixed c. By Theorem 3.1, it would give a coloring approximation ratio at most

$$2c(n + n^2)^{\delta} \le 2^{1+\delta} c\, n^{2\delta}.$$

Choose

$$\varepsilon = \frac{1 - 2\delta}{2} > 0.$$

Then $2\delta < 1 - \varepsilon$, and the constant $2^{1+\delta}c$ is absorbed by the positive exponent gap for all sufficiently large n. Thus the displayed ratio is at most $n^{1-\varepsilon}$ for large inputs. Zuckerman proved that, for every $\varepsilon > 0$, approximating chromatic number within $n^{1-\varepsilon}$ is NP-hard [16, Thm. 1.2]. The finitely many smaller values of n can be solved exactly without affecting polynomial running time, yielding a contradiction. Writing $\delta = 1/2 - \eta$ gives the first formulation. A constant factor is $O(N^{\delta})$ for every fixed $\delta > 0$, so constants are also ruled out. □

Remark 3.3. Corollary 3.2 answers Question 1.1 in the negative and strengthens the answer to every fixed approximation exponent below $1/2$. It also explains why the earlier non-approximation-preserving exact reduction can be bypassed: the universal clique forces L separate mixed assignments, while the internal residual is only $O(n^2)$ and can be amortized by choosing $L = n^2$.

4 The exact residual in the join

The simple inequality

$$L\theta(H) \le A(J_L(H)) \le L\theta(H) + n^2$$

is enough for the hardness transfer, but it hides the exact residual. The residual matters for improving the exponent: if one could guarantee a much smaller yes-case residual for coloring-hard instances, the join size L could be smaller and the hardness exponent would improve.

Definition 4.1 (Marked residual)**.** For an integer $q \ge 0$, a q-admissible marked clique family in H is an indexed family of nonempty cliques

$$\mathcal{F} = (F_1, \dots, F_t)$$

together with a marked set $M \subseteq \{1, \dots, t\}$ such that $|M| \le q$, the full family $\mathcal{F}$ covers every edge of H, and the marked subfamily $(F_i)_{i \in M}$ covers every vertex of H. Let $R_q(H)$ be the minimum of $\sum_i |F_i|$ over all such families, and set $R_q(H) = +\infty$ if none exists.

Plainly $R_q(H) < +\infty$ if and only if $q \ge \theta(H)$.

Theorem 4.2 (Exact residual formula)**.** *For every nonempty graph H and every $L \ge 1$,*

$$A(J_L(H)) = \min_{q \ge \theta(H)} \big(Lq + R_q(H)\big).$$

Proof. Let $G = J_L(H)$ with universal clique U and nonempty base set $W = V(H)$.

First take an arbitrary edge clique cover $\mathcal{C} = (C_1, \dots, C_t)$ of G. Put $W_i = C_i \cap W$ and $U_i = C_i \cap U$. Ignore indices with $W_i = \emptyset$ for the moment; they can only increase the assignment cost and are irrelevant to covering edges incident with W.

For each $u \in U$, define $I_u = \{i : u \in U_i,\ W_i \neq \emptyset\}$. Choose u^* minimizing $|I_u|$ and set $q = |I_{u^*}|$. As in Lemma 2.2, the marked subfamily $(W_i)_{i \in I_{u^*}}$ covers every vertex of H, so $q \geq \theta(H)$. The full nonempty family (W_i) covers every edge of H, because any edge internal to H must be covered by some clique of $\mathcal{C}$ whose intersection with W contains both endpoints. Thus the W_i form a q-admissible marked family, and

$$\sum_{i : W_i \neq \emptyset} |W_i| \geq R_q(H).$$

The mixed universal assignments satisfy

$$\sum_{i : W_i \neq \emptyset} |U_i| = \sum_{u \in U} |I_u| \geq Lq.$$

Therefore

$$\operatorname{size}(\mathcal{C}) \geq Lq + R_q(H) \geq \min_{r \geq \theta(H)} (Lr + R_r(H)).$$

Conversely, let $(\mathcal{F}, M)$ be a q-admissible marked family with total cost $R_q(H)$. For each marked clique F_i, include $U \cup F_i$ in a cover of $J_L(H)$; for each unmarked clique F_i, include F_i itself. The marked cliques cover every vertex of H, so these joined cliques cover all edges between U and W. Since W is nonempty, the marked subfamily is nonempty; hence the joined cliques also cover all edges inside U. The full family covers all edges of H. The assignment cost is at most

$$L|M| + \sum_i |F_i| \leq Lq + R_q(H).$$

Minimizing over q gives the reverse inequality. □

Corollary 4.3 (Large-amplification residual)**.** *Let H be nonempty, and let $\iota(H)$ be its number of isolated vertices. For all sufficiently large L, in particular for $L \geq |V(H)| - \iota(H)$,*

$$A(J_L(H)) = L\theta(H) + R_{\theta(H)}(H).$$

Moreover, for every integer $q \geq \theta(H)$,

$$A(H) + \iota(H) \leq R_q(H), \qquad R_{\theta(H)}(H) \leq A(H) + |V(H)|.$$

Proof. The inequalities follow from the definition. Every q-admissible family has to cover all edges of H, at cost at least $A(H)$, and each isolated vertex must appear in an additional singleton marked clique. This gives $R_q(H) \geq A(H) + \iota(H)$ for every q. For the upper bound at $q = \theta(H)$, take an assignment-optimal edge clique cover of H and add a minimum vertex clique cover of H as marked cliques; assigning each vertex once costs at most $|V(H)|$.

If $q > \theta(H)$, then the formula in Theorem 4.2 pays at least $L(q - \theta(H))$ more in the leading term. The possible residual saving relative to $R_{\theta(H)}(H)$ is at most $|V(H)| - \iota(H)$ by the displayed bounds. Therefore, when $L \geq |V(H)| - \iota(H)$, no $q > \theta(H)$ can improve the minimum. □

Remark 4.4 (Triangle-free bases). The matching extension below also appears in [5, Lem. 4.1]. If H is triangle-free with n vertices, m edges, ι isolated vertices, and maximum-matching size $\nu(H)$, then every internal edge of H must be covered by its own two-vertex clique. Moreover,

$$\theta(H) = n - \nu(H).$$

Indeed, a maximum matching together with singleton cliques for its unmatched vertices is a vertex clique cover with $n - \nu(H)$ members. Conversely, assign each vertex to one member of an arbitrary

vertex clique cover. The members assigned two vertices are pairwise vertex-disjoint edges and hence form a matching. If the cover has q members, at least $n-q$ of them must receive two vertices, so $\nu(H) \geq n-q$ and $q \geq n-\nu(H)$. To obtain a marked subfamily without paying for additional nonisolated singletons, start with a maximum matching, choose one incident edge for every unmatched nonisolated vertex, and mark these edge cliques together with the isolated singleton cliques. After duplicate choices are identified, at most $n-\nu(H) = \theta(H)$ members are marked and they cover every vertex. Thus, for every $q \geq \theta(H)$, the forced edge cliques together with the isolated singletons form a q-admissible family of cost $2m+\iota$, and no admissible family can cost less. Hence the residual formula gives, for every $L \geq 1$,

$$A(J_L(H)) = L\theta(H) + 2m + \iota.$$

For the explicit base $H = K_{r,r}$, a maximum matching has size r, so its vertex clique cover number is $\theta(H) = 2r - r = r$; moreover $m = r^2$ and $\iota = 0$. Hence

$$A(J_L(K_{r,r})) = Lr + 2r^2.$$

Thus the join gadget can have an intrinsically quadratic residual even when the leading term is controlled by $\theta(H)$.

4.1 A residual that changes with the number of marks

The triangle-free residual is constant once enough marks are available. The following small example shows why this need not hold in general. Let CP_3 be the graph obtained from K_6 by deleting the matching $\{x_iy_i : 1 \leq i \leq 3\}$. Write

$$X = \{x_1, x_2, x_3\}, \qquad Y = \{y_1, y_2, y_3\},$$

and, for $1 \leq i \leq 3$, put

$$C_i = \{x_i\} \cup (Y \setminus \{y_i\}), \qquad D_i = \{y_i\} \cup (X \setminus \{x_i\}).$$

Proposition 4.5. *For $H = CP_3$, one has*

$$\theta(H) = 2, \quad A(H) = 12, \quad R_2(H) = 15, \quad R_q(H) = 12 \quad (q \geq 3).$$

Consequently, for every integer $L \geq 1$,

$$A(J_L(CP_3)) = \min\{2L + 15,\ 3L + 12\}. \tag{1}$$

Proof. The cliques X, Y partition the six vertices, and H is not complete, so $\theta(H) = 2$. There are twelve edges and no clique larger than three. Every clique F therefore covers at most $\binom{|F|}{2} \leq |F|$ edges, giving $A(H) \geq 12$. The four triangles X, C_1, C_2, C_3 cover all edges: X covers the edges within X, each pair within Y occurs in the C_i with the missing index, and x_iy_j belongs to C_i whenever $i \neq j$. Their total cost is twelve. Marking C_1, C_2, C_3 covers all vertices, so $R_q(H) = 12$ for every $q \geq 3$.

For $R_2(H)$, mark X, Y and add D_1, D_2, D_3. The latter three triangles cover every cross edge, and the total cost is fifteen. Conversely, two marked cliques covering all six vertices must be disjoint triangles P, Q. Each deleted pair has one endpoint in each triangle, so exactly six edges run between P and Q. These edges require the unmarked cliques. If such a clique F contains a vertices of P and b of Q, then $a + b \leq 3$ and it covers ab cross edges. For nonnegative integers with $a + b \leq 3$, one has $ab \leq \frac{2}{3}(a+b)$: the only case with both positive and total three has product two, and the case of total two has product one. Thus

$$6 \leq \sum_{F \text{ unmarked}} |F \cap P|\,|F \cap Q| \leq \frac{2}{3} \sum_{F \text{ unmarked}} |F|.$$

The unmarked cost is at least nine, in addition to the six marked assignments. Hence $R_2(H) = 15$. Theorem 4.2 now gives (1); for $q \geq 4$, the term $Lq + 12$ exceeds $3L + 12$. □

5 Twin extensions and weighted compression

For a vertex x, write $N(x)$ for its open neighborhood and $N[x] = N(x) \cup \{x\}$. Distinct vertices x, y are *true twins* if $N[x] = N[y]$, and *false twins* if $N(x) = N(y)$ and xy is not an edge. The preceding residual example describes what happens when the size of a true-twin class changes.

Definition 5.1 (Twin extensions, [14, Rules 3.26–3.27])**.** Given an edge clique cover $\mathcal{C}$ of $G - x$ and a twin y of x, the true-twin extension replaces each clique C containing y by $C \cup \{x\}$. The false-twin extension retains every clique of $\mathcal{C}$ and adds $(C \setminus \{y\}) \cup \{x\}$ for each C containing y.

When y has a neighbor in $G - x$, both operations produce covers of G. The two cited rules assert that these operations preserve assignment optimality. They fail even on graphs with no isolated vertices.

Proposition 5.2 (Failure of the true-twin extension)**.** *There is a ten-vertex graph G with true twins z, u and an assignment-optimal cover of $G - z$ whose true-twin extension has cost* 24*, whereas* $A(G) = 23$*.*

Proof. Take $G = J_4(CP_3)$, let U_4 be its universal clique, and choose $z \in U_4$ and $u \in U_3 := U_4 \setminus \{z\}$. Use the notation of Proposition 4.5. The cover

$$X, \quad U_3 \cup C_1, \quad U_3 \cup C_2, \quad U_3 \cup C_3$$

of $G - z$ costs $3 + 3 \cdot 6 = 21$, which equals its optimum by (1). The vertex u occurs in the three joined cliques, so the stated extension adds z to all three and costs $3 + 3 \cdot 7 = 24$. However,

$$U_4 \cup X, \quad U_4 \cup Y, \quad D_1, \quad D_2, \quad D_3$$

is a cover of G of cost $7 + 7 + 3 + 3 + 3 = 23$. Equation (1) gives $A(G) = \min\{23, 24\} = 23$. □

Proposition 5.3 (Failure of the false-twin extension)**.** *There is a six-vertex graph G with false twins x, y and an assignment-optimal cover of $G - x$ whose false-twin extension has cost* 16*, whereas* $A(G) = 12$*.*

Proof. Let $1, 2, 3, 4$ induce a four-cycle in that cyclic order, and join each of the nonadjacent vertices x, y to all four cycle vertices. The graph $G - x$ has the cover

$$\{y, 1, 2\}, \quad \{y, 3, 4\}, \quad \{2, 3\}, \quad \{4, 1\}$$

of cost ten. Every clique intersects the cycle in at most two vertices, so covering its four edges requires at least eight cycle-vertex assignments. Covering the four edges incident to y requires at least two occurrences of y. Thus the displayed cover is optimal.

The false-twin extension adds $\{x, 1, 2\}$ and $\{x, 3, 4\}$ while retaining the original four cliques, and therefore costs sixteen. In contrast,

$$\{x, 1, 2\}, \quad \{x, 3, 4\}, \quad \{y, 2, 3\}, \quad \{y, 4, 1\}$$

covers G at cost twelve. The graph has twelve edges and clique number three, so the edge-count argument in Proposition 4.5 gives the matching lower bound. □

These examples distinguish extension from normalization. Within a fixed graph it is possible to synchronize true twins without increasing cost. After compression, their multiplicity must remain in the objective.

Proposition 5.4 (Weighted true-twin quotient)**.** *Partition $V(G)$ into its maximal true-twin classes. Let Q have these classes as vertices, with two classes adjacent when their vertices are adjacent in G. Give a quotient vertex v the weight w_v equal to the size of its class. Let $A_w(Q)$ be the minimum weighted edge clique cover cost $\sum_C \sum_{v \in C} w_v$, where the minimum is over clique covers of $E(Q)$. Then*

$$A(G) = A_w(Q) + \sum_{\substack{v \text{ isolated in } Q \\ w_v \geq 2}} w_v. \tag{2}$$

Moreover, an optimum of G exists in which each clique contains either all or none of each true-twin class.

Proof. Fix a cover and a true-twin class T with at least two vertices. Choose $t \in T$ occurring in the fewest cliques. Remove $T \setminus \{t\}$ from all cliques and insert all of T in each clique containing t. These remain cliques because the vertices in T are pairwise adjacent and have the same neighbors outside T. Every edge from T to an outside vertex v is covered: some original clique covered tv, and its replacement contains $T \cup \{v\}$. The vertex t occurs at least once, since its edges to $T \setminus \{t\}$ were covered, so the internal edges of T are covered as well. Edges disjoint from T retain their original covers.

If the old occurrence counts were r_s for $s \in T$, then, keeping the clique indices, the new cost on T is $|T| r_t \leq \sum_{s \in T} r_s$; outside assignments are unchanged. Empty cliques may be omitted. Apply this operation successively to all classes. An already synchronized class stays synchronized because later operations alter only vertices of other classes. Starting from an optimum proves the normalization assertion.

Every nonisolated quotient vertex appears in a cover of $E(Q)$, so lifting that cover to whole classes covers both edges between classes and internal edges of those classes. An isolated quotient vertex of weight at least two represents a complete connected component, whose optimal cost is exactly its order. Such components contribute the additional sum in (2); isolated classes of size one contribute zero. This proves the upper bound. Conversely, project a normalized optimum to Q. Its cliques involving nonisolated quotient vertices have weighted cost at least $A_w(Q)$, while each isolated class of size at least two costs at least its order. This gives the reverse inequality. □

In (1), the coefficient L is precisely the weight of the universal class. Increasing it from three to four changes which cover minimizes the total. The weighted quotient retains this change; the unweighted extension rule does not.

6 Upper bounds and structured classes

The hardness above does not eliminate all approximation guarantees. The elementary edge-by-edge algorithm gives a linear bound, and the same counting proof becomes sharper when large cliques are absent.

Proposition 6.1 (Edge-by-edge approximation)**.** *Let G have N vertices and $m > 0$ edges. Covering every edge by its two endpoints is an $(N-1)$-approximation. More generally, it is an $(\omega(G)-1)$-approximation, where $\omega(G)$ is the clique number. In particular, it is a d-approximation on graphs of degeneracy d.*

Proof. The edge-by-edge cover has cost $2m$. In any assignment cover $\mathcal{C}$, each clique C covers at most

$$\binom{|C|}{2} \leq \frac{|C|(N-1)}{2}$$

edges. Hence

$$m \leq \sum_{C \in \mathcal{C}} \binom{|C|}{2} \leq \frac{N-1}{2} \operatorname{size}(\mathcal{C}),$$

so $A(G) \geq 2m/(N-1)$. The ratio of the edge-by-edge cover is at most $N-1$.

If every clique has size at most $\omega(G)$, the same argument gives

$$m \leq \frac{\omega(G)-1}{2}\,\mathrm{size}(\mathcal{C}),$$

and the ratio is at most $\omega(G)-1$. A d-degenerate graph has clique number at most $d+1$. □

The full clique-number estimate is due to [3, Thm. 2(i)]; the proof is included for completeness. For a set $R \subseteq E(G)$ of uncovered edges and a clique C, put

$$g_R(C) = |E(C) \cap R|, \qquad d_R(C) = \frac{g_R(C)}{|C|}.$$

Theorem 6.2 (Weighted set-cover greedy specialization)**.** *Suppose that, for every nonempty current set R of uncovered edges, one can find a positive-gain clique maximizing $d_R(C)$. Repeatedly selecting such a clique gives an H_m-approximation, where $m = |E(G)|$ and*

$$H_m = \sum_{j=1}^{m} \frac{1}{j} \leq 1 + \ln m.$$

More generally, an oracle returning a clique of density at least $1/\alpha$ times the maximum gives an αH_m-approximation.

Proof. This is weighted greedy set cover with universe $E(G)$ and a set of cost $|C|$ for every clique C [9, 11, 2]. The standard charging proof is recalled for completeness. Let $\mathcal{O}$ be an optimal assignment cover. At a step with R containing r edges,

$$\sum_{O \in \mathcal{O}} g_R(O) \geq r, \qquad \sum_{O \in \mathcal{O}} |O| = A(G).$$

Consequently some $O \in \mathcal{O}$ has density at least $r/A(G)$. If the oracle is an α-approximation and returns C with $h = g_R(C)$ and $c = |C|$, then

$$\frac{c}{h} \leq \frac{\alpha A(G)}{r}.$$

Charge the cost c equally to the h newly covered edges. If this step changes the number of uncovered edges from r to $r-h$, then

$$c \leq \alpha A(G)\frac{h}{r} \leq \alpha A(G) \sum_{j=r-h+1}^{r} \frac{1}{j}.$$

The index intervals from successive steps partition $\{1, \ldots, m\}$. Summing the charges therefore gives total cost at most $\alpha H_m A(G)$. □

Lemma 6.3 (Pricing from maximal cliques)**.** *If every graph in a class has polynomially many maximal cliques and they can all be listed in polynomial time, then the exact density oracle in Theorem 6.2 is polynomial-time implementable on that class.*

Proof. Every clique is contained in a maximal clique, so list the maximal cliques and optimize separately inside each one. Fix a maximal clique M. Every subset $S \subseteq M$ is a clique of G, and its density is

$$\frac{|R \cap \binom{S}{2}|}{|S|}.$$

Thus the subproblem inside M is the maximum-density-subgraph problem in the auxiliary graph with vertex set M and edge set $R \cap \binom{M}{2}$. This problem is polynomial-time solvable [8]. For

completeness, a threshold λ can be tested by a maximum-closure instance: create a profit-1 node for each auxiliary edge, a profit-$-\lambda$ node for each vertex, and implication arcs from an edge node to its endpoints of capacity $|R| + 1$. This finite capacity exceeds the total positive profit, so no minimum cut violates an implication. The maximum closure has positive value exactly when some $S \subseteq M$ satisfies

$$|R \cap \tbinom{S}{2}| - \lambda|S| > 0.$$

Maximum closure reduces to one minimum s–t cut. Every candidate density is a/b with $0 \le a \le |R|$ and $1 \le b \le |M|$; enumerating these polynomially many rational thresholds (and scaling capacities by b) recovers an exact optimum. If the auxiliary graph has an edge, let λ_- be the largest candidate threshold at which the optimum closure value is positive. The next distinct candidate is its maximum density d^*: the test is positive precisely below d^*. A positive closure at λ_- supplies a vertex set of density greater than λ_-, which must therefore have density d^*. If the auxiliary graph has no edges, any nonempty subset has optimum density zero. Taking the best solution over all listed maximal cliques proves the lemma. □

Corollary 6.4 (Polynomial-maximal-clique classes)**.** *Every class satisfying Lemma 6.3 admits a polynomial-time H_m-approximation. In particular, chordal graphs admit a polynomial-time ratio*

$$H_m \le 1 + \ln m \le 1 + 2 \ln n.$$

Proof. Combine Lemma 6.3 with Theorem 6.2. A chordal graph has a polynomial-time computable perfect elimination ordering [13]. For each vertex, take it together with its later neighbors; these form a clique. Every maximal clique equals one of these n sets, by considering its earliest vertex. Deleting repeated sets and sets properly contained in another gives all maximal cliques in polynomial time. Finally, $m < n^2$ for $n \ge 2$. □

Remark 6.5. The pricing hypothesis is restrictive on arbitrary graphs. With $R = E(G)$, a clique of order s has density $(s-1)/2$, so exact density pricing already contains maximum clique. The logarithmic result is therefore a structured-class complement to the worst-case hardness, not a general upper-bound improvement.

7 Concluding remarks

The residual formula suggests a precise route for sharpening the lower bound. The quadratic blowup arises because the reduction must dominate the internal residual of the base graph. Any coloring-hard promise family whose complement has small marked residual $R_{\theta(H)}(H)$ would permit a smaller universal clique and hence a stronger hardness exponent. Conversely, proving that no such residual control is available for hard instances would explain why the simple amplification is close to its natural limit. On the algorithmic side, improving the linear worst-case approximation without solving the maximum-density clique pricing problem appears to require ideas beyond ordinary set-cover greedy.

References

[1] W. M. Abdullah and S. Hossain. A sparse matrix approach for covering large complex networks by cliques.

In *Computational Science—ICCS 2022*, LNCS 13352, pp. 505–517, 2022.

doi:10.1007/978-3-031-08757-8_43. Author manuscript.

[2] V. Chvátal. A greedy heuristic for the set-covering problem. *Mathematics of Operations Research*, 4(3):233–235, 1979. doi:10.1287/moor.4.3.233.

[3] A. Davoodi, R. Javadi, and B. Omoomi. Edge clique covering sum of graphs. *Acta Mathematica Hungarica*, 149(1):82–91, 2016. doi:10.1007/s10474-016-0586-1.

[4] A. Davoodi, D. Gerbner, A. Methuku, and M. Vizer. On clique coverings of complete multipartite graphs. *Discrete Applied Mathematics*, 276:19–23, 2020. doi:10.1016/j.dam.2019.09.014; arXiv:1809.01443.

[5] A. Davoodi. Local clique covers and chromatic number. arXiv:2609.07988v1, 7 September 2026.

[6] J. M. Ennis, C. M. Fayle, and D. M. Ennis. Assignment-minimum clique coverings. *ACM Journal of Experimental Algorithmics*, 17(1):1.5:1–1.5:17, 2012. doi:10.1145/2133803.2275596.

[7] A. Firbas, A. Dobler, F. Holzer, J. Schafellner, M. Sorge, A. Villedieu, and M. Wißmann. The complexity of cluster vertex splitting and company. *Discrete Applied Mathematics*, 365:190–207, 2025. doi:10.1016/j.dam.2025.01.012.

[8] A. V. Goldberg. Finding a maximum density subgraph. Technical Report UCB/CSD-84-171, EECS Department, University of California, Berkeley, 1984. Berkeley EECS technical-report record.

[9] D. S. Johnson. Approximation algorithms for combinatorial problems. *Journal of Computer and System Sciences*, 9(3):256–278, 1974. doi:10.1016/S0022-0000(74)80044-9.

[10] L. T. Kou, L. J. Stockmeyer, and C.-K. Wong. Covering edges by cliques with regard to keyword conflicts and intersection graphs. *Communications of the ACM*, 21(2):135–139, 1978. doi:10.1145/359340.359346.

[11] L. Lovász. On the ratio of optimal integral and fractional covers. *Discrete Mathematics*, 13(4):383–390, 1975. doi:10.1016/0012-365X(75)90058-8.

[12] C. Lund and M. Yannakakis. On the hardness of approximating minimization problems. *Journal of the ACM*, 41(5):960–981, 1994. doi:10.1145/185675.306789.

[13] D. J. Rose, R. E. Tarjan, and G. S. Lueker. Algorithmic aspects of vertex elimination on graphs. *SIAM Journal on Computing*, 5(2):266–283, 1976. doi:10.1137/0205021.

[14] A. Ullah. Computing clique cover with structural parameterization. arXiv:2208.12438v1, 2022.

[15] A. Ullah and A. Pothen. Succinct graph representations and algorithmic applications. arXiv:2604.28096v1, 2026.

[16] D. Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory of Computing*, 3(6):103–128, 2007. doi:10.4086/toc.2007.v003a006.