An improved upper bound and algorithm for clique covers

Ryan McIntyre and Michael Soltys

California State University Channel Islands
Dept. of Computer Science
One University Drive
Camarillo, CA 93012, USA
{ryan.mcintyre466@myci.csuci.edu,michael.soltys@csuci.edu}

Abstract. Indeterminate strings have received considerable attention in the recent past; see for example [1] and [3]. This attention is due to their applicability in bioinformatics, and to the natural correspondence with undirected graphs. One aspect of this correspondence is the fact that the minimum alphabet size of indeterminates representing any given undirected graph equals the size of the minimal clique cover of this graph. This paper first considers a related problem proposed in [3]: characterize $\Theta_n(m)$, which is the size of the largest possible minimal clique cover (i.e., an exact upper bound), and hence alphabet size of the corresponding indeterminate, of any graph on n vertices and m edges. We provide improvements to the known upper bound for $\Theta_n(m)$ in section 3.3. [3] also presents an algorithm which finds clique covers in polynomial time. We build on this result with a heuristic for vertex sorting which significantly improves their algorithm's results, particularly in dense graphs.

1 Background

Given an undirected graph $G = (V, E)$, we say that $c \subseteq V$ is a *clique* if every pair of distinct vertices $(u, v) \in c \times c$ comprises an edge—that is, $(u, v) \in E$. A vertex u is *covered* by c if $u \in c$. Similarly, edge (u, v) is covered by c if $\{u, v\} \subseteq c$; we will often write $(u, v) \in c$ instead, a convenient abuse of notation. Similarly, instead of saying "the edges incident on v", we will say "v's edges".

$C = \{c_1, c_2, \cdots, c_k\}$ is a *clique cover* of G if size k if each c_i is a clique, and furthermore every edge and vertex in G is covered by at least one such c_i. Note that there are several variants of this definition. In some contexts, it is only necessary to cover the edges; in others, only the vertices. We consider the case in which both edges and vertices must be covered, and we will call these three variations the *edge cover*, *vertex cover*, and *complete cover* respectively. Whenever we say "clique cover" or "cover" without specifying the type, it should be assumed that we are talking about a complete cover.

The *neighborhood* of a vertex v, denoted $\mathcal{N}_v$ is the set of all vertices adjacent to v; that is, $u \in \mathcal{N}_v$ if $(u, v) \in E$. Every $u \in \mathcal{N}_v$ is a *neighbor* of v. The *degree* of v, denoted d_v, is the cardinality of $\mathcal{N}_v$; $d_v = |\mathcal{N}_v|$. We denote by $\mathcal{R}_v$ the set of vertices which are neither v nor in $\mathcal{N}_v$. We say that v is *isolated*, or that v is a *singleton*, if $d_v = 0$.

The clique cover problem is the problem of algorithmically finding a minimal clique cover, and is $\mathcal{NP}$-hard. The decision version, finding a clique cover whose cardinality is below a given value (or determining that no such cover exists) is $\mathcal{NP}$-complete.

Remark 1 *If a graph has no singletons, then any edge clique cover is also a complete clique cover. Otherwise, any complete cover consists of an edge cover with the addition of a clique for each singleton.*

Given two integers n and m such that $n > 0$ and $0 \leq m \leq \binom{n}{2}$, we let $\mathcal{G}_{n,m}$ denote the set of all simple, undirected graphs on n vertices and m edges. Given any graph G, we denote by $\theta(G)$ the size of a smallest cover of G ([6]). Finally, we denote by $\Theta_n(m)$ the largest $\theta(G)$ of all graphs $G \in \mathcal{G}_{n,m}$. For example, figure 1 shows $\Theta_8(m)$ and $\Theta_7(m)$ plotted together. The plot suggests that $\Theta_n(m)$ is a very uniform function (parametrized by n).

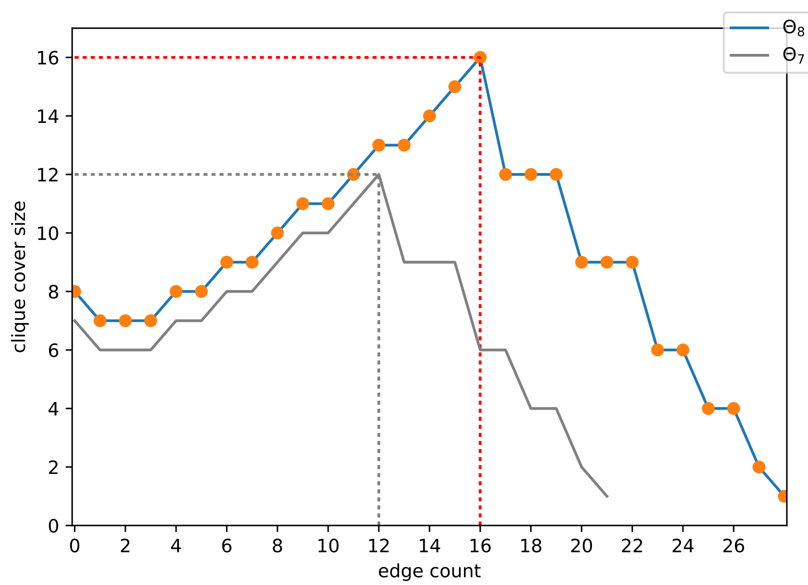


Fig. 1. $\Theta_8(m)$ and $\Theta_7(m)$

We denote by i_G the number of singletons in G, and with c_G the number of non-isolated vertices. Clearly, if $G \in \mathcal{G}_{n,m}$ then $i_G + c_G = n$. We let I_G denote the subgraph of G consisting of the all singletons, and C_G the subgraph consisting of all non-singletons and edges—$|I_G| = i_G$ and $|C_G| = c_G$. Finally, we let S_G (with cardinality s_G) denote the set of vertices which are adjacent to all other vertices (we call them *stars*). That is, $v \in S_G$ if $\mathcal{N}_v = V - \{v\}$.

We define D_G to be the degree sum of G, and A_G the average degree in G. That is, $D_G = \sum_V d_v$ and $A_G = D_G/|G|$. These will usually be denoted simply with D and A if G is implied by the context.

Given a vertex or set of vertices v in graph G, we denote by $G - v$ the graph which results from removing v (or every vertex in v), along with all edges incident to v, from G.

2 Summary of Results

In this paper, we explore two topics. First, we aim to characterize $\Theta_n(m)$ in section 3. We synthesize theorems from Lovász (Theorem 3), Mantel and Erdős (Theorem 2) to establish an upper bound for $\Theta_n(m)$ which is exact for some values of m but not for others. We establish that $\Theta_n(m)$ has recursive properties, which we use to characterize it for some values of m and bound it in others. We improve Lovász's bound in Theorems 12 and 17. These improvements are likely extendible to the complete characterization of $\Theta_n(m)$ (see conjecture 14). A succinct summary of these results can be found in section 3.3.

Next, in section 4, we establish a heuristic to order vertices and edges. The motivation is an algorithm developed in [3] which outputs a clique cover in polynomial time with respect to the number of vertices; this algorithm does not necessarily output a minimal or small cover, but it works quickly. Moreover, it outputs covers of different sizes when presented with vertices in a different order. We develop and explore a heuristic reminiscent of the `PageRank` algorithm (we call it `CliqueRank`) and apply it in combination with some naïve heuristics. The resulting covers are significantly smaller than those from the original algorithm, particularly in dense graphs.

3 Characterizing $\Theta_n(m)$

In [3, Problem 11] the authors pose the following problem: describe the function $\Theta_n(m)$ for every n. They provide as an example a (slightly flawed) graph for $\Theta_7(m)$, where $m \in [21] = [\binom{7}{2}]$ (see [3, Fig. 3]). For $n > 7$, the number of graphs quickly becomes unwieldy, so it is desirable to compute $\Theta_n(m)$ analytically. Our results do not necessarily apply to very small graphs; we assume throughout that any graph worth discussing has at least 4 vertices, as we can characterize $\Theta_n(m)$ for $n < 4$ easily by brute force. In fact, we have found Θ_n by brute force for all $n \leq 8$.

We know from [3] and from the results of Mantel and Erdős [5, 2] that the global maximum of $\Theta_n(m)$ is reached at $m = \lfloor n^2/4 \rfloor$. The reason is that this is the largest number of edges which can fit on n vertices without forcing triangles. This maximum is realized in complete bipartite graphs—such graphs have no triangles or singletons, so covers consist of all edges. The expression

'$\lfloor n^2/4 \rfloor$' will be used frequently, so we abbreviate it: for any expression exp, we let $\overline{\text{exp}} = \lfloor \text{exp}^2/4 \rfloor$.

Figure 2 displays the largest complete bipartite graphs on five and six vertices respectively: $\mathcal{K}_{3,2}$ and $\mathcal{K}_{3,3}$. Note that $\theta(\mathcal{K}_{3,2}) = 6 = \overline{5}$ and $\theta(\mathcal{K}_{3,3}) = 9 = \overline{6}$. For any natural n, $\theta(\mathcal{K}_{\lceil n/2 \rceil, \lfloor n/2 \rfloor}) = \overline{n}$.

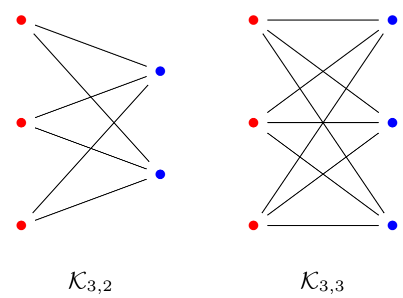


Fig. 2. Complete bipartite graphs

Theorem 2 (Mantel, Erdős) *If a graph on n vertices contains no triangle, then it contains at most $\overline{n}$ edges.*

Theorem 3 (Lovász) *Given $G \in \mathcal{G}_{n,m}$, let k be the number of missing edges (i.e. $k = \binom{n}{2} - m$), and let t be the largest natural number such that $t^2 - t \leq k$. Then $\theta(G) \leq k + t$. Moreover, this bound is exact if $k = t^2$ or $k = t^2 - t$.*

For $m \leq \overline{n}$, we rely primarily on the theorems above, provided by Mantel and Erdős [5, 2] (Theorem 2) and Lovász [4] (Theorem 3); we use them to prove our first contribution, namely that $\Theta_n(m)$ has some recursive properties. These properties provide an exact upper bound when $m \leq \overline{n}$. Lovász provides an inexact upper bound when $m \geq \overline{n}$. We propose two improvements to Lovász's bound in Theorems 12 and 17, for which proofs can be found in section 3.2; these improvements comprise our most notable theoretical results in this paper. We also give conjecture 14; if proven true, this conjecture finishes the complete exact upper bound of for $m \geq \overline{n}$.

Theorem 12 *If $m > \overline{n}$ then $\Theta_n(m) \leq \overline{n-1}$.*

Theorem 17 *If $m > \binom{n}{2} - \overline{n-2}$ then $\Theta_n(m) \leq \overline{n-2}$.*

Conjecture 14 *If $k < \overline{p}$, then $\Theta_n(\binom{n}{2} - k) \leq \overline{p}$.*

3.1 Pre-maximum: $\Theta_n(m)$ for $m \leq \overline{n}$

We begin by introducing our results informally. We then prove a sequence of auxiliary results which will help us characterize $\Theta_n(m)$. The forthcoming material is rather technical, but the reader will find it easier to follow by keeping the graph in figure 3 in mind.

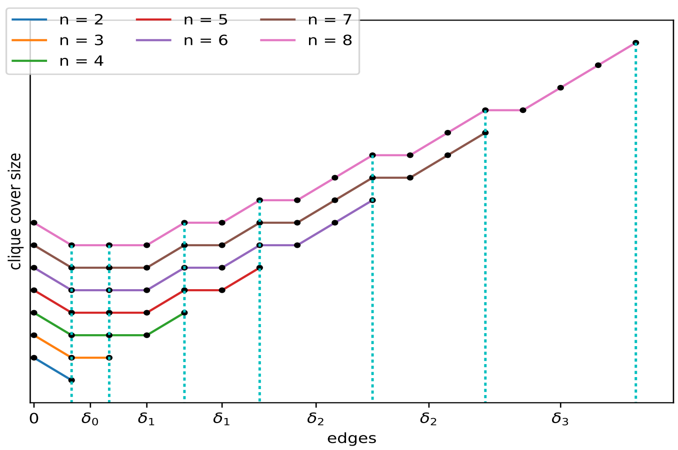


Fig. 3. Left sides of $\Theta_n(m)$ for $n \in [2, 8]$

We will refer to the portion of $\Theta_n(m)$ where $m \leq \overline{n}$ as the *left side* of the function. As figure 3 displays, we can obtain the left side of $\Theta_n(m)$ for $n \geq 3$ by translating that of $\Theta_{n-1}(m)$ upward by one, and then extending it by a new segment $\delta_{\lfloor (n-1)/2 \rfloor}$. Here, δ_k represents a series of changes $(\Delta x, \Delta y)$, consisting first of $(+1, +0)$ followed by k iterations of $(+1, +1)$. For example, $\delta_3 = \{(+1, +0), (+1, +1), (+1, +1), (+1, +1)\}$, as shown in figure 4.

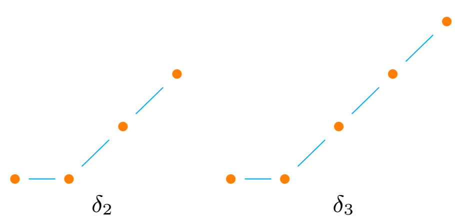


Fig. 4. δ_2 and δ_3

We can easily determine the first seven points in $\Theta_n(m)$ via brute force. Clearly, $\Theta_n(0) = n$; each vertex must be covered individually by a single clique, as there are no edges. The addition of a single edge allows two vertices to be covered with this edge, so $\Theta_n(1) = n - 1$. Figure 5 provides visual justification for the first seven points of $\Theta_n(m)$ for $n \geq 5$.

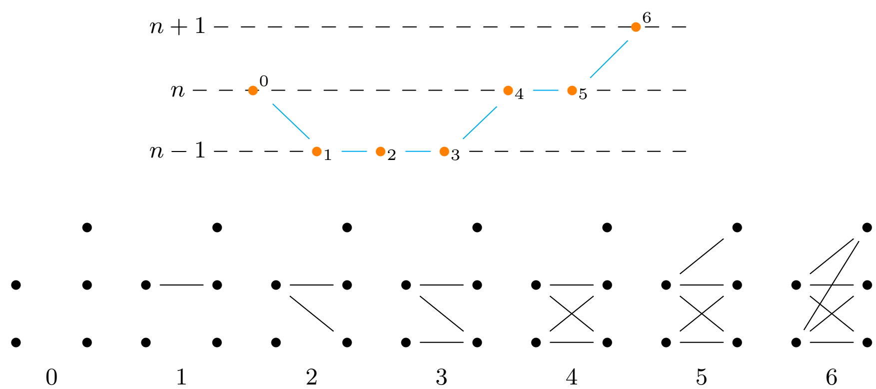


Fig. 5. $\Theta_n(m)$ for $n \geq 5$ and $m \leq 6$ with corresponding graphs

Claim 4 *If $n \geq 4$, then $\Theta_n(0) = \Theta_n(4) = n$ and $\Theta_n(1) = \Theta_n(2) = \Theta_n(3) = n - 1$.*

Claim 5 $\Theta_n(m + 1) \leq \Theta_n(m) + 1$

Claim 4 can be verified quickly by checking every possible configuration of 0-4 edges. Claim 5 is true because any edge added to a graph can simply be covered by a single additional clique consisting of that edge's vertices.

Lemma 6 *For any graph G, $\theta(G) \leq \overline{c_G} + i_G$, where i_G is the number of singletons and c_G the non-isolated vertices.*

Proof. Theorem 2 guarantees that C_G can be covered by $\overline{c_G}$ cliques. I_G can be covered by i_G cliques, each consisting of a singleton vertex. Every edge is in C_G, and every vertex is either in I_G or C_G, so the union of the covers of C_G and I_G covers G and contains at most $\overline{c_G} + i_G$ cliques. □

Lemma 7 *If $G \in \mathcal{G}_{n,m}$ for some $m \leq \overline{n}$, then*
$\big(G$ *contains triangles* $\implies \theta(G) < \Theta_n(m)\big)$

Proof. Assume that $G \in \mathcal{G}_{n,m}$ for some $m \leq \overline{n}$, and furthermore that G has at least one triangle. Three edges can be covered with this triangle, so $\theta(G) \leq m - 2 + i_G$.

Case 1: $m \leq \overline{c_G}$. We can construct a triangle-free graph $C \in \mathcal{G}_{c_G,m}$ and singleton graph $I \in \mathcal{G}_{i_G,0}$. Let $G' = C \cup I$. Then $G' \in \mathcal{G}_{n,m}$. Moreover, it has no triangles, so every edge must be covered individually, as must the singletons. Thus, $\theta(G') = m + i_G > m + i_G - 2 \geq \theta(G)$, so $\theta(G) < \Theta_n(m)$.
Case 2: $m > \overline{c_G}$. We must first note that $i_G \neq 0$, as this would imply that $m > \overline{n}$, which directly contradicts the hypothesis.

Lemma 6 guarantees that $\theta(G) \leq \overline{c_G} + i_G$; call this upper bound β_0. Consider a graph $G_1 \in \mathcal{G}_{n,m}$ such that $c_{G_1} = c_G + 1$ and $i_{G_1} = i_G - 1$. Such a graph can be constructed easily—G contains a triangle, so simply remove an edge from this triangle and use it to connect a vertex in I_G to one in C_G. Again, lemma 6 grants $\theta(G_1) \leq \overline{c_{G_1}} + i_{G_1}$; call this bound β_1. Let's compare these two bounds.

If c_G is even, then $\overline{(c_G+1)} - \overline{c_G} = c_G/2$. Since C_G contains a triangle and has an even vertex count, $c_G \geq 4$. Thus, $\overline{c_{G_1}} - \overline{c_G} \geq 2$. Otherwise, c_G is odd, so $\overline{(c_G+1)} - \overline{c_G} = (c_G+1)/2$. Again, there are at least three vertices in C_G, so $\overline{c_{G_1}} - \overline{c_G} \geq 2$.

Whether c_G is even or odd, $\beta_1 > \beta_0$. Of course, this does not prove that $\theta(G_1) > \theta(G)$. The process can be repeated on G_1 to gain G_2 with bound $\beta_2 > \beta_1$, and so on, until a G_α is reached such that $c_{G_\alpha} \geq m$. Since $m \leq \overline{n}$, this will necessarily happen before or when we run out of singletons.

If $\alpha = 1$, then $\overline{(c_G+1)} \geq m$, so we can construct a triangle-free graph $C \in \mathcal{G}_{(c_G+1),m}$ and a graph $I \in \mathcal{G}_{(i_G-1),0}$ consisting of $(i_G - 1)$ singletons. Let $G' = C \cup I$. Clearly, $G' \in \mathcal{G}_{n,m}$. Moreover, $\theta(G') = m + i_{G_1} = m + i_G - 1$. Recall that $\theta(G) \leq m + i_G - 2$. So we have found a $G' \in \mathcal{G}_{n,m}$ such that $\theta(G') > \theta(G)$. Therefore, $\theta(G) < \Theta_n(m)$.

If $\alpha > 1$, then we can construct a triangle-free graph $C \in \mathcal{G}_{c_{G_\alpha},m}$ (Theorem 2 guarantees that such a graph exists) and singleton graph $I \in \mathcal{G}_{i_{G_\alpha},0}$. Let $G' = C \cup I$. Then $\theta(G') = m + i_{G_\alpha}$. Moreover, $m \geq \overline{c_{G_{\alpha-1}}} + 1$ or we would have stopped before G_α; $i_{G_\alpha} = i_{G_{\alpha-1}} - 1$ by construction, so $\theta(G') \geq \beta_{\alpha-1}$. Thus, $\theta(G') > \theta(G)$, so $\theta(G) < \Theta_n(m)$.

Regardless of α's value, we have shown that $\theta(G) < \Theta_n(m)$. □

An immediate consequence of lemma 7 is: if $G \in \mathcal{G}_{n,m}$ for some $m \leq \overline{n}$ and $\theta(G) = \Theta_n(m)$ then G contains no triangles. With this, we can fully characterize $\Theta_n(m)$ for $m \leq \overline{n}$ in Theorem 8.

Theorem 8 *If $m \leq \overline{n}$, let p be the smallest natural number such that $\overline{p} \geq m$. Then $\Theta_n(m) = m + n - p$.*

Proof. Let $m \leq \overline{n}$, and let G be a graph in $\mathcal{G}_{n,m}$ such that $\theta(G) = \Theta_n(m)$. G has no triangles, so its minimal cover consists of a clique for each edge, and one for each singleton vertex. That is, $\theta(G) = m + i_G$. m is constant, so $\theta(G)$ is entirely dependent on i_G. As such, G is any triangle-free graph on n vertices and m edges which maximizes i_G, or equivalently minimizes c_G. Theorem 2 grants that m edges can be placed without triangles on c_G vertices if and only if $\overline{c_G} \geq m$, so c_G must be the smallest number meeting this condition; $c_G = p$, where p is the smallest natural number such that $\overline{p} \geq m$. As such $i_G = n - c_G = n - p$, so $\theta(G) = m + n - p$. □

The following conclusions can quickly be drawn from Theorem 8:

Lemma 9 *If $p < n$, then $\Theta_n(\overline{p}) = \Theta_n(\overline{p}+1)$.*

Proof. Theorem 8 implies that $\Theta_n(\overline{p}) = \overline{p} + n - p$ and that $\Theta_n(\overline{p}+1) = (\overline{p}+1) + n - (p+1) = \overline{p} + n - p$. □

Lemma 10 *If $m \leq \overline{n}$, then $\Theta_{(n+1)}(m) = \Theta_n(m) + 1$.*

Proof. Since $m \leq \overline{n} < \overline{n+1}$, Theorem 8 proves that $\Theta_n(m) = m + n - p$ and $\Theta_{n+1}(m) = m + (n+1) - p = \Theta_n(m) + 1$, where p is the smallest natural number such that $\overline{p} \geq m$. □

While lemma 9 is not necessary for the characterization, it does explain the distribution of short plateaus throughout the left side of Θ_n.

Lemma 10 shows that $\Theta_n(m)$ behaves recursively on the left side; while this fact is not needed to prove our results, it displays their structural causes. Note that lemma 10 is actually a direct result of lemma 7, and could be used to prove Theorem 8—in fact, this was the approach we used in early versions of the proofs above. As such, lemma 10 should be considered the recursive version of Theorem 8. The δs described in figures 3 and 4 are necessary to complete the recursion; after moving the left side of $\Theta_n(m)$ upward by 1, we must extend it by $\delta_{\lfloor n/2 \rfloor}$ to complete the left side of $\Theta_{n+1}(m)$. The shape of these extensions can be proven accurate with lemma 7 or 9 in conjunction with claim 5.

3.2 Post-maximum: $\Theta_n(m)$ for $m \geq \overline{n}$

Again, we begin by informally discussing our results before delving into proofs. We will refer to the part of $\Theta_n(m)$ where $m \geq \overline{n}$ as the *right side* of the function. The left side was shown to behave recursively with respect to n. The right side appears to do the same for small n, and we conjecture that it does for all n.

Lovász's Theorem (Theorem 3) provides an upper bound for $\Theta_n(m)$ based on the number of missing edges. Here, we restate it:

Theorem 3 (Lovász) *Given $G \in \mathcal{G}_{n,m}$, let k be the number of missing edges (i.e. $k = \binom{n}{2} - m$), and let t be the largest natural number such that $t^2 - t \leq k$. Then $\theta(G) \leq k + t$. Moreover, this bound is exact if $k = t^2$ or $k = t^2 - t$.*

First, note that Theorem 3 relies solely on the number of missing edges. It is exact at the specified values of k, but only if $k \leq \overline{n-1}$. If $k > \overline{n-1}$, then $m < \overline{n}$ and a better bound can be found using our characterization of the left side of Θ_n.

Of course, Lovász's bound is not exact for all $m \geq \overline{n}$. As shown in figure 6 and stated in Theorem 3, it is only necessarily exact if $k = \overline{t}$. Between these exact values, Lovász's bound appears to be a smoother version of Θ; where the right side of Θ is a jagged series of plateaus, Lovász's bound is nearly linear.

Lovász bound is difficult to apply as presented. We rephrase it here. Clearly, $\overline{2t} = t^2$ and $\overline{2t \pm 1} = t^2 \pm t$. Moreover, any natural number can be written as $2t$ or $2t - 1$ for some value of t. As such, we can adopt Theorem 3 to our notation:

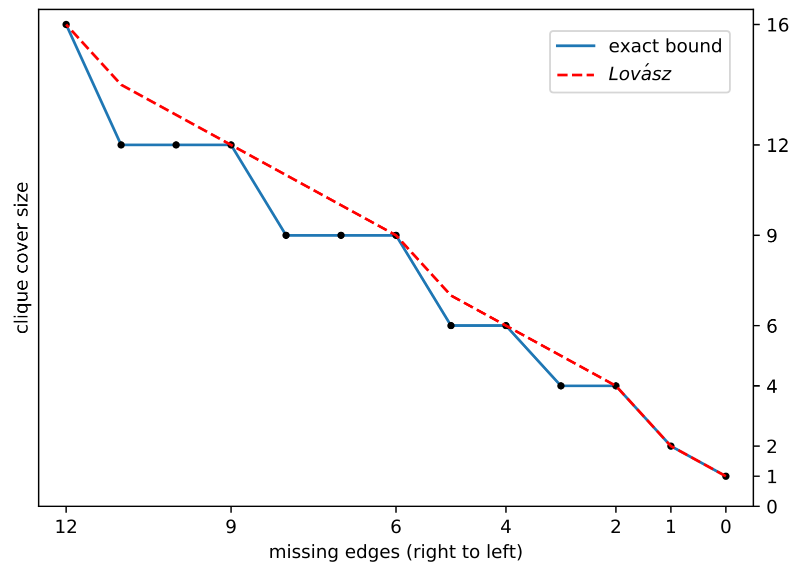


Fig. 6. Lovász's bound vs the right side of Θ_8

Theorem 3 (Lovász restated) *Given* $m \geq \overline{n}$ *and* $k = \binom{n}{2} - m$*:*

– *If* $k = \bar{t}$ *for some natural number* t*, then* $\Theta_n(m) = \overline{t+1}$*.*
– *Otherwise, if* t *is the largest natural number such that* $\overline{2t-1} \leq k$*, then* $\theta(G) \leq k + t$*.*

The plateaus on the right side of Θ are identical between different n for $n \leq 8$, and we conjecture this is true for larger n. It appears that if $m \geq \overline{n}$, then $\Theta_n(m)$ is a function of the number of missing edges, independent of the vertex count. This is displayed in figure 7.

The differences between the left and right sides raise an immediate, fundamental question: why is the right side of Θ characterized by large value changes where the left is smooth (i.e. never changing by more than one clique per edge)? What are the structural causes behind this difference? It seems that the answer to this question can be reduced to the behavior of complete bipartite graphs; if such a graph is missing an edge, then its cover size is simply one less. If it has an extra edge, however, this edge completes several triangles, resulting in a larger drop in cover size coupled with the capability of adding some additional edges without affecting cover size. As an example of this phenomenon, we provide figure 8, which shows the graphs corresponding to maximum cover size on the right side of Θ_7.

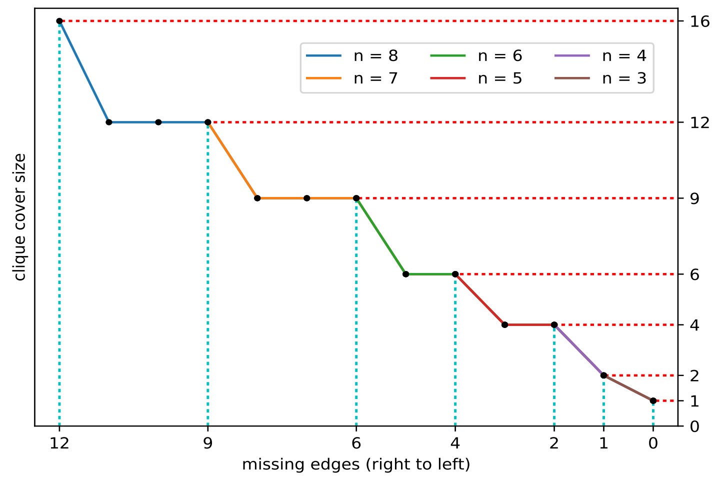


Fig. 7. The right sides of Θ_n for $3 \leq n \leq 8$

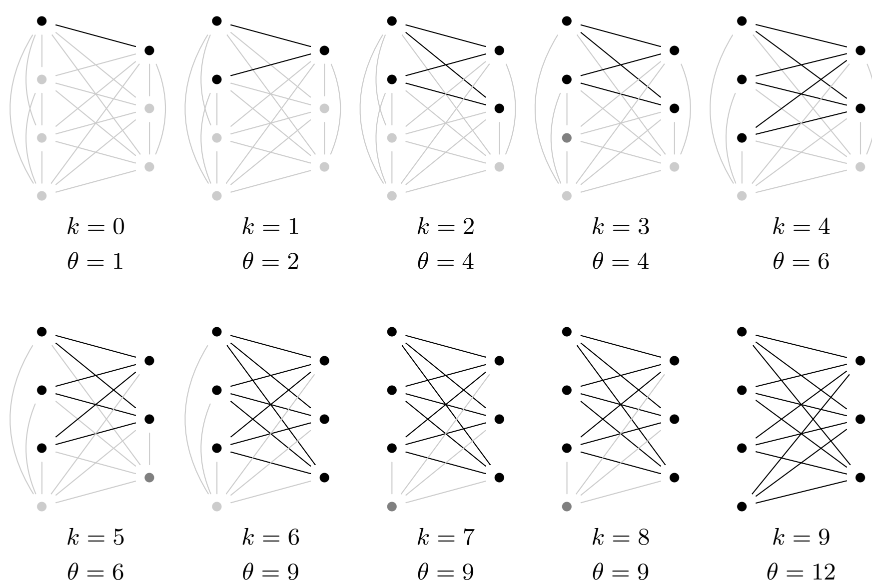


Fig. 8. The largest possible bipartite graph (black) for k missing edges corresponds to $\Theta_n((\binom{n}{2} - k))$ for $n \leq 8$. We conjecture that this is true for all n (see conjecture 14).

We will now begin a series of proofs to show that some of the plateaus in the right side (specifically, the first two after the global maximum) necessarily exist for all n. These are a direct improvement to Lovász bound.

Lemma 11 *Given a graph G with triangle Δ, remove all edges from Δ to obtain G'. Then $\theta(G') \geq \theta(G) - 1$.*

Proof. Let C' be a minimal cover of G'. Let $C = C' \cup \{\Delta\}$; that is, C is C' with a single additional clique, containing only the three vertices in Δ. Clearly, C covers G, so $\theta(G) \leq \theta(G') + 1$. □

We can now prove Theorems 12 and 17. Recall:

Theorem 12 *If $m > \overline{n}$ then $\Theta_n(m) \leq \overline{n-1}$.*

Proof. This can be shown quickly through enumeration of all arrangements of edges for three or four vertices. We present a proof by strong induction for larger graphs. That is, we assume that it is true for all $n_0 \leq n$, and prove that it is true for $n+1$.

If $m \geq \binom{n}{2} - \overline{n-2}$, then Theorem 3 provides proof. As such, we assume that

$$\overline{n} < m < \binom{n}{2} - \overline{n-2} \tag{1}$$

Case 1: n is even and at least 4. Consider $G \in \mathcal{G}_{(n+1),m}$ where $\overline{n+1} < m < \binom{n+1}{2} - \overline{n-1}$. The degree sum D of G is exactly twice the number of edges, so (1) grants that $2\overline{n+1} < D < 2\binom{n+1}{2} - 2\overline{n-1} = \frac{n^2+4n}{2}$. Clearly, the average degree A must be $D/(n+1)$. Thus

$$A < \frac{n^2+4n}{2(n+1)} < \frac{n+3}{2}$$

Let v be a minimum degree vertex. The minimum degree is at most the average degree, so $d_v \leq \lfloor A \rfloor \leq \lfloor (n+3)/2 \rfloor$. Since n is even, this means $d_v \leq n/2+1$.

Subcase 1.1: $d_v = n/2+1$.

Subcase 1.1.1: $m = \overline{n+1}+1$. Then

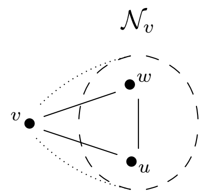
$$D = 2\overline{n+1}+2 = \frac{(n+1)^2+3}{2}, \text{ so } A = \frac{D}{n+1} = \frac{n+1}{2} + \frac{3}{2(n+1)}$$

and since $n \geq 4\ldots$

$$A < \frac{n}{2} + 1$$

$A < n/2+1$, so $d_v \leq n/2$, which contradicts the conditions of this subcase. This set of conditions cannot occur, so it need not be considered any further.

Subcase 1.1.2: $m \geq \overline{n+1}+2$. Let w be any vertex in $\mathcal{N}_v$; clearly $d_w \geq n/2+1$ as well. Other than v and w, there are $(n-1)$ vertices in G. Moreover, v and w are each connected to at least $n/2$ of these $(n-1)$ vertices; they have at least one neighbor u in common. (u, v, w) is a triangle.



Remove v and all $n/2+1$ of its edges from G to obtain G'. G' has n vertices and at least $\overline{n+1}+2-(n/2+1) > \overline{n}$ edges. The hypothesis grants that $\theta(G') \leq \overline{n-1}$. Moreover, v and the two edges (v,u) and (v,w) can be covered with the triangle (u,v,w). The remaining $n/2-1$ edges adjacent to v can each be covered by their own clique. Thus, $\theta(G) \leq \theta(G') + n/2 \leq \overline{n-1} + n/2 = \overline{n}$. That is, $\theta(G) \leq \overline{n}$.

Subcase 1.2: $d_v \leq n/2$. Our job is much easier in this case; v and its edges can be covered by at most $n/2$ cliques. The rest of G consists of n vertices and more than $\overline{n+1} - n/2 = \overline{n}$ edges; the hypothesis grants that it can be covered by at most $\overline{n-1}$ cliques. Thus, $\theta(G) \leq \overline{n-1} + n/2 = \overline{n}$.

Case 2: n is odd and at least 3. Consider $G \in \mathcal{G}_{(n+1),m}$ such that $\overline{n+1} < m < \binom{n+1}{2} - \overline{n-1}$.

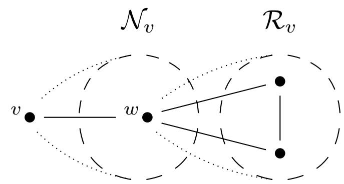
$$2\overline{n+1} < D < 2\binom{n+1}{2} - 2\overline{n-1}$$
$$\frac{(n+1)^2}{2} < D < \frac{(n+1)(n+3)}{2} - 2$$

so, since $A = D/(n+1)$

$$\frac{n+1}{2} < A < \frac{n+3}{2} - \frac{2}{n+1}$$

Let v be a minimum degree vertex. $d_v \leq \lfloor A \rfloor$, so $d_v \leq \frac{n+1}{2}$.

Subcase 2.1: $d_v = (n+1)/2$. We first show that there is a vertex w such that $d_w = (n+1)/2$ and w is in a triangle. If v is in a triangle, then w is v. Otherwise, $\mathcal{N}_v$ contains $(n+1)/2$ vertices and no edges. The vertices in $\mathcal{N}_v$ each have at least $(n+1)/2$ neighbors themselves—but there are only $(n+1)/2$ vertices (including v) which are not in $\mathcal{N}_v$. Thus, every vertex in $\mathcal{N}_v$ has a degree of exactly $(n+1)/2$, and these edges connect every vertex in $\mathcal{N}_v$ to every vertex in $\mathcal{R}_v$. Let's count edges: there are $(n+1)/2$ edges connecting v to $\mathcal{N}_v$, none within $\mathcal{N}_v$, and another $(n+1)(n-1)/4$ connecting $\mathcal{N}_v$ to $\mathcal{R}_v$. This totals $\overline{n+1}$ edges; at least one edge is unaccounted for, and the only remaining space is between vertices which are neither v nor in $\mathcal{N}_v$. This edge is in triangles with every element of $\mathcal{N}_v$, each of which have degree $(n+1)/2$; let w be any one of $\mathcal{N}_v$'s vertices.



We have a vertex w such that $d_w = (n+1)/2$ and w is in a triangle. As such, w and all of its adjacent edges can be covered with $(n-1)/2$ cliques—two of the edges are covered by this triangle. The rest of the graph consists of n vertices and more than $\overline{n+1} - (n+1)/2 = \overline{n}$ edges. By the hypothesis, it can be covered by at most $\overline{n-1}$ cliques. As such, G can be covered by at most $\overline{n-1} + (n-1)/2 = \overline{n}$ cliques.

Subcase 2.2: $d_v \leq (n-1)/2$. Let v be a vertex with $d_v \leq (n-1)/2$. Then v and all of its incident edges can be covered by at most $(n-1)/2$ cliques. The rest of G consists of n vertices and more than $\overline{n}$ edges, so the hypothesis grants that it can be covered with $\overline{n-1}$ cliques. Thus, $\theta(G) \leq \overline{n}$. □

Theorem 12 provides the bound shown in the first "plateau" of $\Theta_n(m)$ after m passes $\overline{n}$; it is easy to construct a graph with this exact cover size; simply create the largest complete bipartite subgraph possible with the number of missing edges. Thus, this upper bound is exactly $\Theta_n(m)$ for $\overline{n} < m \leq \binom{n}{2} - \overline{n-2}$

Remark 13 *$\overline{n} = \binom{n}{2} - \overline{n-1}$, so Theorem 12 identically reads: If $m > \binom{n}{2} - \overline{n-1}$ then $\Theta_n(m) \leq \overline{n-1}$.*

The largest complete bipartite graph that can be constructed on n vertices has $m = \overline{n}$ edges and $k = \overline{n-1}$ missing edges; in fact, even for larger numbers of vertices this is the largest such graph with less than $\overline{n}$ edges missing. Moreover, for $n \leq 8$ we have determined via brute force that, when $m > \overline{n}$, the largest possible complete bipartite subgraph matches the maximum cover size. We suspect that this is true for larger n:

Conjecture 14 *If $k < \overline{p}$, then $\Theta_n(\binom{n}{2} - k) \leq \overline{p}$.*

We prove in Theorems 12 and 17 that conjecture 14 holds when p is $n-1$ or $n-2$, respectively. Remarks 15 and 16, while not necessary to prove our results, may be useful in proving conjecture 14.

Remark 15 *If $G \in \mathcal{G}_{n,m}$ for some $m > \overline{n}$ and $i_G > 0$, then $\theta(G) < \Theta_n(m)$.*

Proof. Assume g contains singleton vertex z. Because $m > \overline{n}$, G necessarily contains a triangle $\Delta = (u, v, w)$. Let G' be G, without the edges in Δ. Lemma 11 grants that $\theta(G') \geq \theta(G) - 1$. Let G'' be G', with three additional edges: (u, z), (v, z), and (w, z). z was a singleton prior, and there are no edges between u, v, and w in G'', so none of these three new edges is in a triangle; they must be covered individually, but they also cover z (which required its own clique in G). Thus, $\theta(G'') \geq \theta(G') + 2 \geq \theta(G) + 1$. Clearly $G'' \in \mathcal{G}_{n,m}$, so $\theta(G) < \Theta_n(m)$. □

The proof of lemma 15 could be easily improved to apply whenever $m > \overline{n-1}$. To see this, note that if $\overline{n-1} < m \leq \overline{n}$, then a singleton guarantees (via Theorem 2) that the remaining $(n-1)$ vertices contain triangles. Lemma 7 finishes the proof.

Remark 16 *If S_G is nonempty for some graph G with at least two vertices, then let $s \in S_G$ and let G' be the result of removing s and all of its edges from G. Then $\theta(G') = \theta(G)$.*

Proof. Let C' be a cover for G'. Define C with:

$$C = \bigcup_{c \in C'} \{c \cup \{s\}\}$$

$|C| = |C'|$ and C covers G, so $\theta(G) \leq \theta(G')$.

Similarly, let C be a cover for G; we can assume without loss of generality that s is in every clique in C, because s can be part of any clique in G due to its adjacency with every vertex in G. Construct C':

$$C' = \bigcup_{c \in C} \{c - \{s\}\}$$

C' covers G' and has the same cardinality as C. Thus, $\theta(G') \leq \theta(G)$. □

Finally, we extend the bound in conjecture 14 to a second plateau. The proof of Theorem 17 is lengthy and technical with many subcases.

Theorem 17 *If $m > \binom{n}{2} - \overline{n-2}$ then $\Theta_n(m) \leq \overline{n-2}$.*

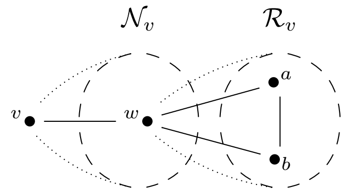
Proof. We have determined through exhaustive search that this lemma is true for all $n \leq 8$. We present an inductive proof for $n > 8$. Note that $\binom{n}{2} - \overline{n-2} = \overline{n+1} - 1$, so $m \geq \overline{n+1}$ provides an identical lower bound for m; this is version of the bound we'll use in this proof. Much like in Theorem 12, we can rely on Lovász's (Theorem 3) for $m \geq \binom{n}{2} - \overline{n-3}$. As such, we assume throughout that $m < \binom{n}{2} - \overline{n-3}$.

Case 1: n is even and at least 10. Let $G \in \mathcal{G}_{n,m}$. We assume $m \geq \overline{n+1}$, so the degree sum D of G is at least $2\overline{n+1}$. That is, $D \geq (n^2 + 2n)/2$. As such, the average degree A of G is at least $n/2 + 1$. Similarly, $m < \binom{n}{2} - \overline{n-3}$, so $D < (n^2)/2 + 2n - 4$. Thus, $A < n/2 + 2$.

Let v be a minimum degree vertex in G; $d_v \leq \lfloor A \rfloor \leq n/2 + 1$.

Subcase 1.1: $d_v \leq n/2 - 1$. We can cover v and all of its edges with at most $n/2 - 1$ cliques. Let G' be the rest of G; it consists of $n-1$ vertices and at least $\overline{n+1} - (n/2 - 1) > \overline{n}$ edges, so by the hypothesis $\theta(G') \leq \overline{n-3}$. Therefore, $\theta(G) \leq \overline{n-3} + n/2 - 1 = \overline{n-2}$.

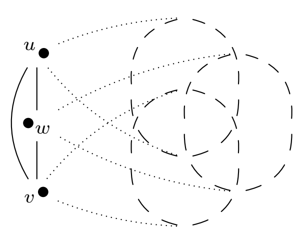
Subcase 1.2: $d_v = n/2$. We first prove that there is a vertex w of degree $n/2$ which is in a triangle. If v is in a triangle, we're done. Otherwise, there are no edges within $\mathcal{N}_v$. There are $n/2$ vertices in $\mathcal{N}_v$, $n/2 - 1$ in $\mathcal{R}_v$, and of course v itself. Notice that, if there are no edges within $\mathcal{N}_v$, then each vertex in $\mathcal{N}_v$ has at most $n/2$ edges (those leading to v or $\mathcal{R}_v$). Since $n/2$ is the minimum degree, every vertex in $\mathcal{N}_v$ must be connected to v and all of $\mathcal{R}_v$. So, there are $n/2$ edges connecting v to $\mathcal{N}_v$, no edges within $\mathcal{N}_v$, and another $\frac{n}{2}(\frac{n}{2}-1)$ between $\mathcal{N}_v$ and $\mathcal{R}_v$. We have counted $\overline{n}$ edges; there are at least $n/2$ edges unaccounted for, and these edges must be within $\mathcal{R}_v$. So, choose any edge (a, b) in $\mathcal{R}_v$ and any vertex $w \in \mathcal{N}_v$; (w, a, b) is a triangle and $d_w = n/2$.



So there is some vertex w in a triangle, such that $d_w = n/2$. Because w is in a triangle, it and its $n/2$ edges can be covered with at most $n/2 - 1$ cliques. Let G' be the rest of G; G' has $n-1$ vertices and at least $\overline{n}$ edges, so $\theta(G') \leq \overline{n-3}$ by the hypothesis. Thus, $\theta(G) \leq \overline{n-3} + n/2 - 1 = \overline{n-2}$.

Subcase 1.3: $d_v = n/2 + 1$.

Subcase 1.3.1: $m = \overline{n+1}$. All n vertices have degree of at least $n/2+1$; this alone accounts for all $\overline{n+1}$ edges, so every vertex has this degree exactly. Since $m > \overline{n}$, there is some triangle (u, v, w) in G. u, v and w each have $n/2+1$ edges, two of which are within this triangle. As such, they each have $n/2 - 1$ edges connecting them to the other $n-3$ vertices. In other words, there are a total of $3n/2 - 3$ edges connecting (u, v, w) to the rest of G. Let G' be G without u, v, w or their edges. There are exactly $n-3$ vertices in G'; given a vertex a in G', all edges (if any exist) from (u, v, w) to a can be covered by a single clique. As such, the edges from (u, v, w) to G' can be covered by $n-3$ cliques. Moreover, these cliques necessarily cover the edges in (u, v, w) because $2(n/2 - 1) = n-2$, so each pair in (u, v, w) has at least one neighbor in G' in common.



G' consists of $n-3$ vertices and $\overline{n+1} - 3n/2 = \overline{n-2} - 1$ edges. $n > 4$, so $\overline{n-2} - 1 > \overline{n-3}$. Theorem 12 grants that $\theta(G') \leq \overline{n-4}$. Thus, $\theta(G) \leq \overline{n-4} + n - 3 = \overline{n-2}$.

Subcase 1.3.2: $m > \overline{n+1}$. So $m = \overline{n+1} + d$ for some $d > 0$. v is in a triangle (u, v, w), just like the previous subcase, but there are up to d additional edges connecting (u, v, w) to the rest of G; any of the d extra edges not between

(u, v, w) and G' are within G'. As such, we can use the exact bound described in the previous case, both for the edges connecting (u, v, w) to the rest of G, and for the rest of G. Additional edges in G' do not invalidate our upper bound for $\theta(G')$, nor can the extra edges between (u, v, w) increase the number of cliques necessary to cover these edges with the method described in the previous subcase. Thus, we have the same bound: $\theta(G) \leq \overline{n-2}$.

Case 2: n is odd and at least 9. Just as in the previous case, the average degree A of graph $G \in \mathcal{G}_{n,m}$ is less than $n/2 + 2$. Therefore the minimum degree d_v is at most $(n + 3)/2$.

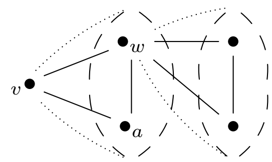
Subcase 2.1: $d_v = (n+3)/2$. In this case, the degree sum is at least $(n^2+3n)/2$, so $(n^2 + 3n)/4 \leq m < \binom{n}{2} - \overline{n-3}$. In other words, $m = (n^2 + 3n)/4 + d$ where $0 \leq d < (n - 9)/4$. By Theorem 2, any graph $G \in \mathcal{G}_{n,m}$ has a triangle (u, v, w). Let G' be G without (u, v, w). Given a vertex a in G', every edge between (u, v, w) and a can be covered in a single clique, so the edges from this triangle to G' can be covered by at most $n - 3$ cliques. Moreover, the minimum degree $(n+3)/2$ guarantees that any vertex in (u, v, w) is adjacent to at least $(n-1)/2$ vertices outside of this triangle, so any two vertices in (u, v, w) have a common neighbor not in (u, v, w). Thus, the edges (u, v), (u, w) and (v, w) are necessarily covered in triangles with the $n - 3$ cliques covering the edges from (u, v, w) to G'. The minimum degree accounts for $(n^2 + 3n)/4$ of the edges, so there are at most d additional edges (other than those implied by the minimum degree sum) between (u, v, w) and G'. There are 3 edges in (u, v, w), and at most $3(n-1)/2+d$ edges from (u, v, w) to G', so there are $m' \geq m - 3 - 3(n - 1)/2 - d$ edges in G'. $m = (n^2 + 3n)/4 + d$, so $m' \geq (n^2 - 3n - 6)/4$. Moreover, since $n \geq 9$, this implies that $m' \geq (n^2-4n+3)/4$; that is, $m' \geq \overline{n-2}$. G' only has $n-3$ vertices, so by the hypothesis $\theta(G') \leq \overline{n-5}$. As such, $\theta(G) \leq \overline{n-5} + n - 3 < \overline{n-2}$. In fact, in this case our upper bound for $\theta(G)$ is $\overline{n-3} + 1$.

Subcase 2.2: $d_v = (n + 1)/2$. For every vertex $w \in \mathcal{N}_v$, the minimum degree guarantees that $\mathcal{N}_v \cap \mathcal{N}_w \neq \emptyset$; that is, v and w have at least one neighbor in common. This common neighbor corresponds to an edge in $\mathcal{N}_v$.

Subcase 2.2.1: All of these edges within $\mathcal{N}_v$ share a common vertex, a. Then every $w \in \mathcal{N}_v - \{a\}$ has no neighbors other than a in $\mathcal{N}_v$. So, w is connected to v and a, along with at least $(n - 3)/2$ other vertices, none of which can be in $\mathcal{N}_v$. There are only $(n - 3)/2$ vertices in $\mathcal{R}_v$, so every such w must be adjacent to them all. That is, $\mathcal{N}_w = \{v, a\} \cup \mathcal{R}_v$, and $d_w = (n + 1)/2$.

If a is connected to every element of $\mathcal{R}_v$, then $a \in S_G$ and can be removed without reducing the cover size (remark 16), leaving a graph on $n - 1$ vertices and $\overline{n+1} - (n - 1)$ edges. $\overline{n+1} - (n - 1) = \overline{n-1} + 1$, so Theorem 12 shows that $\theta(G) \leq \overline{n-2}$.

As such, it is safe to assume that a is adjacent to at most $(n - 5)/2$ of the $(n - 3)/2$ vertices in $\mathcal{R}_v$. There are at most $(n^2 - 4n + 3)/4$ additional edges between the other $(n - 1)/2$ vertices in $\mathcal{N}_v$ and those in $\mathcal{R}_v$. Finally, there are the $(n + 1)/2$ edges from v to $\mathcal{N}_v$ and the $(n - 1)/2$ within $\mathcal{N}_v$; we have a total of at most $(n^2 + 2n - 7)/4$ edges accounted for. There are at least 2 more edges, which must be in $\mathcal{R}_v$.



Consider any $w \in \mathcal{N}_v - \{a\}$. $d_w = (n+1)/2$, and $\mathcal{N}_w = \{v, a\} \cup \mathcal{R}_v$. Any edge in $\mathcal{R}_v$ is opposite w in a triangle; since there are edges in $\mathcal{R}_v$, w is in a triangle involving itself and two elements of $\mathcal{R}_v$. w is also in the triangle (w, a, v). w meets the conditions imposed on v in the next subcase: it has two disjoint edges in its neighborhood. Subcase 2.2.2 finishes the proof that $\theta(G) \leq \overline{n-2}$.

Subcase 2.2.2: There are edges (a, b) and (c, d) on 4 unique vertices in $\mathcal{N}_v$, or there is a triangle (a, b, c) in $\mathcal{N}_v$. In the prior case, the edges (v, a) and (v, b) can be covered by triangle (v, a, b), as can (v, c) and (v, d) by (v, c, d), so v's $(n+1)/2$ edges can be covered by $(n-3)/2$ cliques. Similarly, in the case of a triangle in $\mathcal{N}_v$, three of v's edges can be covered by a single 4-clique (consisting of v and this triangle), which results in the same upper bound. Let G' be G without v or its edges. G' has at least $\overline{n+1} - (n+1)/2$ edges. That is, G' is on $n-1$ vertices and at least $\overline{n}$ edges. By the hypothesis, $\theta(G') \leq \overline{n-3}$. As such, $\theta(G) \leq \overline{n-3} + (n-3)/2 = \overline{n-2}$.

Subcase 2.3: $d_v = (n-1)/2$.

Subcase 2.3.1: v is in a triangle. Let G' be G without v or its $(n-1)/2$ edges. G' has $n-1$ vertices and more than $\overline{n}$ edges, so $\theta(G') \leq \overline{n-3}$ by the hypothesis. v's $(n-1)/2$ edges can be covered by $(n-3)/2$ cliques because v is in a triangle, so $\theta(G) \leq \overline{n-3} + (n-3)/2 = \overline{n-2}$.

Subcase 2.3.2: v is not in a triangle. Then there are no edges within $\mathcal{N}_v$. As such, every vertex in $\mathcal{N}_v$ is connected to at least $(n-3)/2$ of the $(n-1)/3$ vertices in $\mathcal{R}_v$. This accounts for $(n^2 - 4n + 3)/4$ edges. Adding the $(n-1)/2$ from v to $\mathcal{N}_v$ raises this total to $(n^2 - 2n + 1)/4$; there are least n more edges. There is only room for $(n-1)/2$ additional edges between $\mathcal{N}_v$ and $\mathcal{R}_v$, so there are at least $(n+1)/2$ edges in $\mathcal{R}_v$.

If there are any additional edges between $\mathcal{N}_v$ and $\mathcal{R}_v$, let $w \in \mathcal{N}_v$ be adjacent to one of these edges. $d_w = (n+1)/2$, and every edge in $\mathcal{R}_v$ is in $\mathcal{N}_w$. There are at least $(n+1)/2$ edges on the $(n-1)/2$ vertices in $\mathcal{R}_v$; only $(n-3)/2$ of these edges can be adjacent to a single vertex, so there is either a pair of edges (a, b) and (c, d) on four unique vertices in $\mathcal{R}_v$ or a triangle in $\mathcal{R}_v$. w meets the conditions imposed on v in subcase 2.2.2, which completes the proof that that $\theta(G) \leq \overline{n-2}$.

If there are no additional edges between $\mathcal{N}_v$ and $\mathcal{R}_v$, then there are at least n edges on $\mathcal{R}_v$'s $(n-1)/2$ vertices. Consider $w \in \mathcal{N}_v$; it has degree $(n-1)/2$ and is connected to all but one of the $(n-1)/2$ vertices in $\mathcal{R}_v$. Let $a \in \mathcal{R}_v$ be this vertex. Clearly, the n edges in $\mathcal{R}_v$ cannot all be adjacent to a. The rest of $\mathcal{R}_v$ is in $\mathcal{N}_w$, so there is an edge in $\mathcal{N}_w$. We have found a vertex of degree $(n-1)/2$ which is in a triangle. Subcase 2.3.1 shows that $\theta(G) \leq \overline{n-2}$.

Subcase 2.4: $d_v \leq (n-3)/2$. Let G' be G without v. G has $n-1$ vertices and at least $\overline{n+1} - (n-3)/2 > \overline{n}$ edges. The hypothesis provides that $\theta(G') \leq \overline{n-3}$. Thus, $\theta(G) \leq \overline{n-3} + (n-3)/2 = \overline{n-2}$. □

Note that we have bounded cover size several times using the following method: select a vertex v and cover everything except v and its edges, and then add cliques to cover these ommited edges. The cover of v's edges is equivalent to a vertex cover of $\mathcal{N}_v$. For any v in graph G, we define $\phi(v)$ to be a minimal vertex clique cover of $\mathcal{N}_v$.

Remark 18 *If graph G with vertex set V has no singletons, then* $\theta(G) \leq \min_{v \in V}\{\theta(G-v) + \phi(v)\}$.

Proof. Clearly, $G-v$ can be covered with $\theta(G-v)$ cliques. This covers everything in G except v and its incident edges. Let C be a minimal vertex cover of $\mathcal{N}_v$, and let $C' = \bigcup_{c \in C}\{c \cup \{v\}\}$. C' consists of $\phi(v)$ cliques, which cover v and all of its edges. □

In the proof of Theorem 17 we also bound the cover size by isolating a clique and covering everything which is not adjacent to this clique, then covering it and the edges connecting it to the rest of the graph.

Remark 19 *If a graph G with n vertices has no singletons and contains clique Δ with d vertices, then* $\theta(G) \leq \theta(G - \Delta) + n - d + 1$.

Proof. Clearly, $G - \Delta$ can be covered with $\theta(G - \Delta)$ cliques. All edges between Δ and a vertex $v \notin \Delta$ can be covered by a single clique; there are at most $n - d$ such vertices. Finally, Δ itself may need to be covered (though it may not, if the $n - d$ cliques coincidentally covered Δ as well). □

3.3 The complete upper bound

We list three versions of the upper bound: one with Lovász and Mantel's Theorems along with lemma 10; one with the improvements provided in Theorems 12 and 17; and finally the hypothesized exact upper bound pending proof of conjecture 14. In all three bounds, n is the number of vertices, m the edges, and k the missing edges. All values are assumed to be natural numbers.

With Lovász's Theorem and lemma 10, we can form an upper bound for Θ_n:

$$\Theta_n^{(2)}(m) \begin{cases} = \Theta_{n-1}(m) + 1 & \text{for } m \leq \overline{n-1} \quad \text{(2a)} \\ = m & \text{for } \overline{n-1} < m \leq \overline{n} \quad \text{(2b)} \\ \leq k + \max\{t | t^2 - t \leq k\} & \text{for } \overline{n} < m \leq \binom{n}{2} \quad \text{(2c)} \end{cases}$$

With Theorems 12 and 17, we can improve the previous bound:

$$\Theta_n^{(3)}(m) \begin{cases} = \Theta_{n-1}(m) + 1 & \text{for } m \leq \overline{n-1} \quad \text{(3a)} \\ = m & \text{for } \overline{n-1} < m \leq \overline{n} \quad \text{(3b)} \\ = \overline{n-1} & \text{for } \overline{n-1} > k \geq \overline{n-2} \quad \text{(3c)} \\ = \overline{n-2} & \text{for } \overline{n-2} > k \geq \overline{n-3} \quad \text{(3d)} \\ \leq k + \max\{t | t^2 - t \leq k\} & \text{for } \overline{n-3} > k \geq 0 \quad \text{(3e)} \end{cases}$$

If conjecture 14 is proven true, the bound can be simplified and made exact for all m:

$$\Theta_n^{(4)}(m) = \begin{cases} \Theta_{n-1}(m) + 1 & \text{for } m \leq \overline{n-1} \quad (4a) \\ m & \text{for } \overline{n-1} < m \leq \overline{n} \quad (4b) \\ \overline{\min\{t | \overline{t} > k\}} & \text{for } m > \overline{n} \quad (4c) \end{cases}$$

Note that these three formulations form a refinement of the upper bound; $\Theta_n^{(4)}(m) \leq \Theta_n^{(3)}(m) \leq \Theta_n^{(2)}(m)$ for all m. Conjecture 14 is sufficient to show that $\Theta_n = \Theta_n^{(4)}$.

4 Finding covers

The authors of algorithm 1 in [3] provide a process which finds a clique cover in polynomial time ($O(n^4)$) on the number of vertices. It works by assigning symbols to sets of vertices; each symbol corresponds to a clique, and each vertex is in a symbol's clique if an only if it has been assigned that symbol. The algorithm's purpose is to construct an indeterminate string from its associated graph, but this is identical to covering said graph. We paraphrase this process in algorithm 1. It produces different results for isomorphic graphs based on the order in which the vertices are presented. In [3, conjecture 12], it is proposed that there is an ordering of vertices which results in an optimal (i.e., minimal) cover.

Algorithm 1 Labelling [3, algorithm 1]

Require: Graph $G = (V, E)$
$\lambda \leftarrow 1$
For each $v \in V$, label$(v) = \{\}$
for $v \in V$ **do**
  if $d_v = 0$ **then**
    label$(v) \leftarrow \{\lambda\}$
    $\lambda \leftarrow \lambda + 1$
  else
    for $w \in \mathcal{N}_v$ **do**
      if label$(v) \cap$ label$(w) = \emptyset$ **then**
        label$(v) \leftarrow$ label$(v) \cup \{\lambda\}$
        label$(w) \leftarrow$ label$(w) \cup \{\lambda\}$
        clique $\leftarrow \{w\}$
        for $q \in \mathcal{N}_v - \{w\}$ **do**
          if *clique* $\subseteq \mathcal{N}_q$ **then**
            label$(q) \leftarrow$ label$(q) \cup \{\lambda\}$
            clique $\leftarrow$ *clique* $\cup \{q\}$
        $\lambda \leftarrow \lambda + 1$

In this section, we present an original heuristic, which we call `CliqueRank` in tribute to its inspiration, `PageRank`. We show that `CliqueRank` reduces the size of algorithm 1's output covers, particularly in dense graphs. Figure 10 displays the results of applying `CliqueRank` to algorithm 1 with several different methods; the relevant methods will be explained in section 4.2.

4.1 `CliqueRank`

`CliqueRank` assigns a value to all vertices and edges in a graph. It operates as follows:

1. Every vertex is given an initial value of 1.
2. The value of each vertex is redistributed uniformly among the edges in its neighborhood. An edge (v, w) is in u's neighborhood if $v, w \in \mathcal{N}_v$. Recall that v itself is not in $\mathcal{N}_v$; this value is being redistributed among those edges which are opposite v in triangles. So if there are m edges in $\mathcal{N}_v$, each of these edges receives $(1/m)$ of v's value. An edge's value for this iteration is the sum of such inputs from vertices.
3. Each edge then splits its value evenly between its two vertices.

For a visual demonstration of an iteration of `CliqueRank`, see figure 9. Steps 2 and 3 are intended for iteration, as their descriptions imply. Note that when an object "redistributes its value", it loses this value; no value is being created other than the initial assignment of 1 to every vertex. As such, at the end of an iteration, the edges all have 0 value. When we reference an edge's value after n iterations, however, we will actually be referring to its value *during* the n'th iteration, after it has been given value by vertices and before it has redistributed this value to vertices.

During the first iteration, any vertices which are not in triangles lose all of their value; it is redistributed among 0 edges, so it ceases to exist. Moreover, these triangle-less vertices share this property with their edges, so these edges never gain value. Thus, all edges and vertices which are not in triangles have value equal to 0 after the any positive number of iterations.

If a vertex or edge is in a triangle, however, then it is easy to prove through induction that it has nonzero value after every iteration. Moreover, it is also easy to prove that edges which are in exactly one triangle will have value less than edges in multiple triangles. That is, edges which are "easier to cover", meaning they are in multiple cliques, tend toward larger values. This falls apart when 4-cliques come into play; if an edge is in exactly one 4-clique, and no triangles other than those within this 4-clique, it will still appear to be "in three triangles". That is, its value will not be as low as those edges which are in exactly one triangle, even though it is contained in exactly one maximal clique.

This presents intuitive strategies for covering a graph. First, edges with zero value should be covered; they are not in triangles, so they must be covered individually (as must singleton vertices, which will also be given zero value). Then, a vertex with a low value and uncovered edges can be selected (v in algorithm 1), and its neighbors (w and q) can be considered in any order. There are many ways in which neighbors can be prioritized, and we consider a few of them in section 4.2.

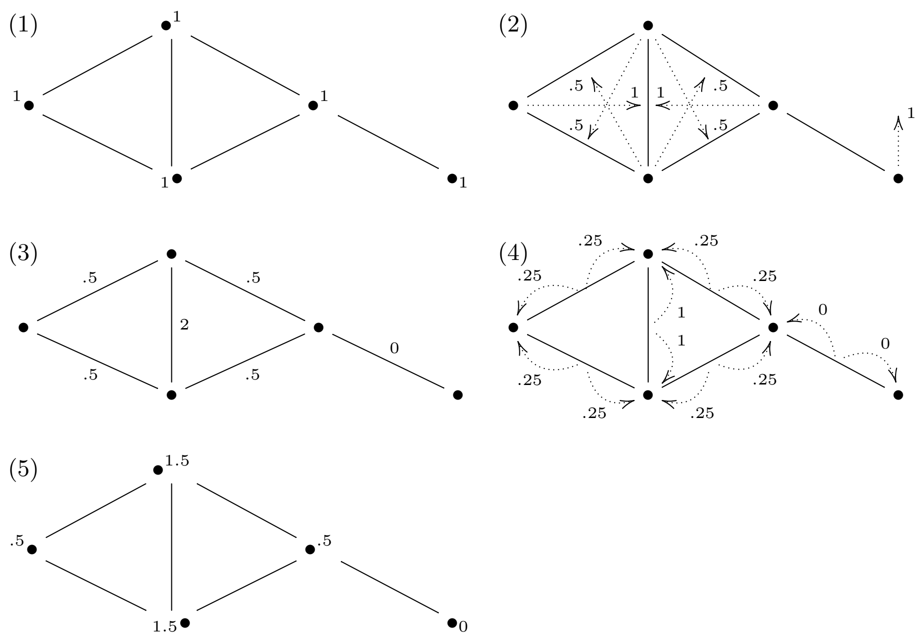


Fig. 9. An iteration of `CliqueRank`

4.2 Applying `CliqueRank` to algorithm 1

`CliqueRank` provides a method of assigning values to edges and vertices; these values can be applied to algorithm 1 in an assortment of ways, some of which are effective and some of which are not. In this section, we define and evaluate the effectiveness of some of these methods of application. Methods will be named in correspondence with the legend in figure 10. All methods can be applied after any positive number of iterations of `CliqueRank`. Surprisingly, it is rare for extra iterations to improve the resulting cover size; the best cover is usually found after a single iteration, but occasionally better covers can be found by iterating to convergence. An iteration of `CliqueRank` is $O(n^3)$, so on large graphs it is prudent to iterate just once.

We next examine a few methods of application of `CliqueRank` to algorithm 1. In figure 10 and the following descriptions, we use *Vscore* to refer to vertex values, *Escore* to refer to edge values, and *ECC* to refer to "edge cover count", i.e., a counter of the number of times each edge has been covered.

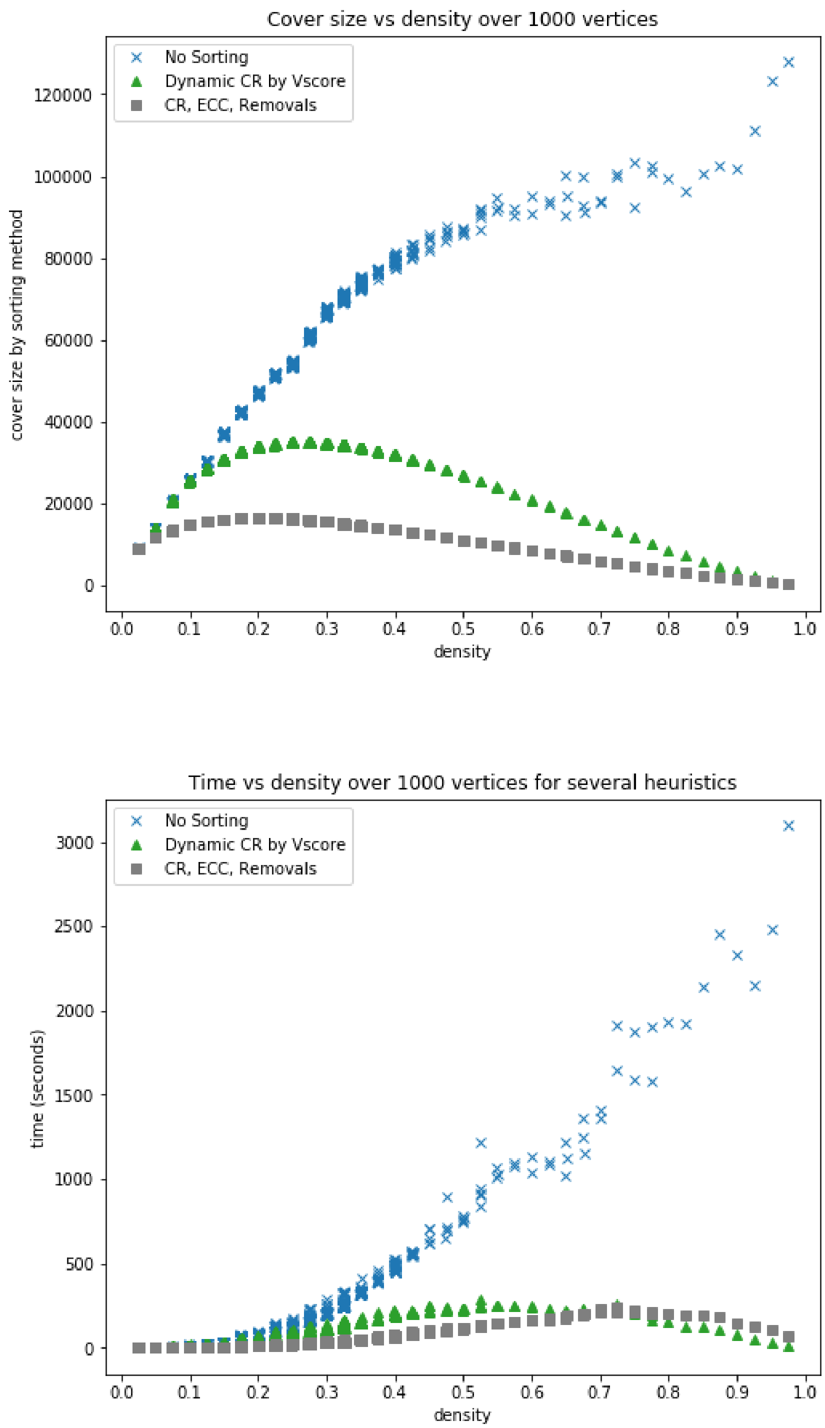


Fig. 10. Cover size vs edge density and time vs edge density for Algorithm 1

`CR` **by Vscore:** As the name implies, this method of application of `CliqueRank` to algorithm 1 works simply by sorting the vertices in ascending order with respect to their `CliqueRank` values; that is, low valued vertices are considered first in lines 3, 8, and 13 of algorithm 1. This method is not shown in figures, but the following method is nearly identical to it, with one small variation.

Dynamic `CR` by Vscore: This method sorts vertices in non-decreasing `CR` score as well, but whenever a clique is added to the cover-in-progress, this vertex's score is increased by 1.

`CR` **by Escore:** This method operates as follows: in algorithm 1, line 3 is sorted by non-decreasing vertex score, and lines 8 and 13 are sorted (again, non-decreasing) by the edge scores of the edge connecting the new vertex to the vertex selected in line 3. Its results are not included in figures as it is not particularly effective; it is of note because we expected it to be a top competitor, and as such we mention it as a possibility which we have found to be ineffective.

`CR`**, ECC, Removals:** This method, shown in grey, operates as follows: vertices are initially put in non-decreasing order by `CR` score in line 3. Vertices in line 8 are sorted primarily by whether the corresponding edge (connecting w to v in the pseudocode) is covered—uncovered edges come first. They are sorted secondarily in non-decreasing order of edge score from `CR`. Finally, vertices in line 13 are sorted by the number of uncovered edges connecting them to the clique in construction, in non-increasing order. When the graph is covered, we then review all cliques in non-decreasing order of size. If every edge in a given clique is covered more than once by remaining cliques, then the clique in question is superfluous and is removed from the cover. This last step rarely finds any redundancy, but occasionally reduces cover size minutely.

5 Conclusion and Future Work

The function $\Theta_n(m)$ is the exact upper bound on the size of a minimal clique cover for a graph with n vertices and m edges. We progress toward an exact characterization of the shape of Θ_n for any n; using theorems from Erdős and Mantel, we fully characterize $\Theta_n(m)$ for $m \leq \overline{n}$ via the recursive properties in lemma 10. Lovász provides an upper bound for $\Theta_n(m)$ when $m > \overline{n}$, and we improve this to an exact characterization for $m \leq \binom{n}{2} - \overline{n-3}$ with Theorems 12 and 17.

If conjecture 14 is true, it completes the characterization of Θ_n for all n.

Conjecture 14 *If $k < \overline{p}$, then $\Theta_n(\binom{n}{2} - k) \leq \overline{p}$.*

Remarks 18 and 19 formalize the strategies used in the proofs of Theorems 12 and 17 to bound cover size; they may be useful in completing the characterization of $\Theta_n(m)$. Remarks 15 and 16 may also prove useful in this pursuit.

We then move on to application; bioinformatics provides motivation to find small clique covers. We develop a method for ordering vertices (`CliqueRank`) and apply it to a recently developed algorithm for indeterminate string construction.

Much of the motivation for graphs and graph algorithms originates in networks situated in space, which (among other applications) are represented by graphs in euclidean space. In figure 11, we show the results of algorithm 1 on metric graphs.

We generate these graphs as follows: points are randomly distributed in the n-dimensional box $[0, 1]^n$. These points are the vertices. Any two vertices which are within a specified distance of eachother, under a given metric, are connected. Figure 11 shows the results of algorithm 1 on 2-dimensional graphs using the euclidean metric.

It is of interest to improve performance on metric graphs. Also, as demonstrated by algorithm 1's motivation in bioinformatics and string processing, it is pertinent to generate graphs via construction of indeterminate strings, and to analyze and improve performance and effectiveness on this particular class of graphs.

References

[1] Manolis Christodoulakis, P.J. Ryan, W.F. Smyth, and Shu Wang. Indeterminate strings, prefix arrays and undirected graphs. *Theoretical Computer Science*, 600:34 – 48, 2015.

[2] Paul Erdos, A.W. Goodman, and Louis Posa. The representation of a graph by set intersections. *Canadian Journal of Mathematics*, 18:106–112, 1966.

[3] Joel Helling, P. J. Ryan, W. F. Smyth, and Michael Soltys. Constructing an indeterminate string from its associated graph. *Accepted for publication in the Journal of Theoretical Computer Science*, 2017.

[4] L. Lovász. On covering of graphs. In G. Katona P. Erdos, editor, *Theory of graphs*. Akad. Kiadó, 1968.

[5] W. Mantel. Problem 28 (solution by H. Gouweniak, W. Mantel, J. Texeira de Mattes, F. Schuh, and W.A Whythoff). *Wiskundige Opgaven*, 10(60–61), 1907.

[6] Fred S. Roberts. Applications of edge coverings by cliques. *Discrete Applied Mathematics*, 10:93–109, 1985.

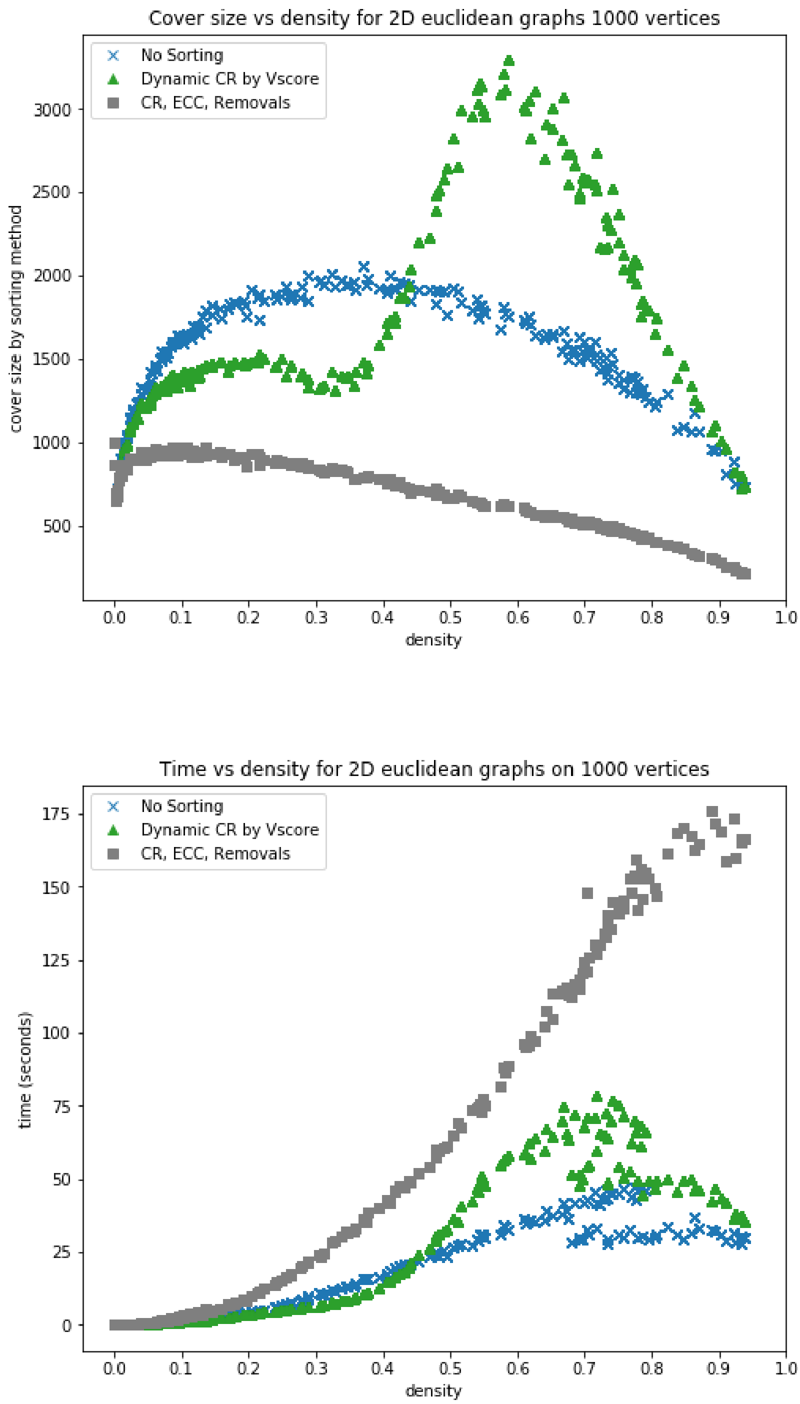


Fig. 11. Cover size and time vs density for 2D metric graphs