Journal of Discrete Algorithms 48 (2018) 42–56

Contents lists available at ScienceDirect

Journal of Discrete Algorithms

www.elsevier.com/locate/jda

An improved upper bound and algorithm for clique covers

Ryan McIntyre, Michael Soltys *

California State University Channel Islands, Dept. of Computer Science, One University Drive, Camarillo, CA 93012, USA

A R T I C L E I N F O

Article history:
Available online 27 March 2018

Keywords:
Clique covers
Indeterminates
Lovász bound
Mantel's Theorem

A B S T R A C T

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 of 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.

* Corresponding author.
E-mail addresses: ryan.mcintyre466@myci.csuci.edu (R. McIntyre), michael.soltys@csuci.edu (M. Soltys).

https://doi.org/10.1016/j.jda.2018.03.002
1570-8667/

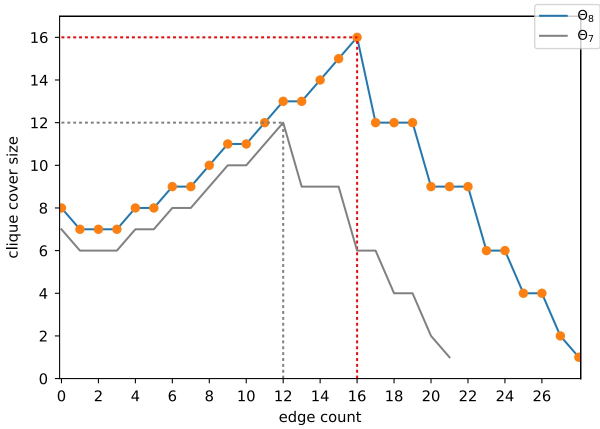

Fig. 1. $\Theta_8(m)$ and $\Theta_7(m)$. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

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 $0 \le m \le \binom{n}{2}$, 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, Fig. 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).

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] (following work in [1]) 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 $n \le 8$.

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 $\overline{\text{exp}} = \lfloor \text{exp}^2/4 \rfloor$.