ARTICLE IN PRESS

Discrete Applied Mathematics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

On comparing algorithms for the maximum clique problem

Alexandre Prusch Züge [a,*], Renato Carmo [b]

[a] Campus Avançado de Jandaia do Sul, UFPR — Rua Dr. João Maximiano, 426, Bloco UFPR, Jandaia do Sul, PR, 86900-000, P.O. Box 123, Brazil

[b] Departamento de Informática, UFPR — Centro Politécnico da Universidade Federal do Paraná, Curitiba, PR, 81531-990, P.O. Box 19081, Brazil

ARTICLE INFO

Article history:
Received 31 December 2015
Received in revised form 4 December 2017
Accepted 8 January 2018
Available online xxxx

Keywords:
Maximum clique
Experimental algorithms
Random graphs
Branch and bound

ABSTRACT

Several algorithms for the exact solution of the maximum clique problem are available in the literature. Some have been proposed with the aim of bounding the worst case complexity of the problem, while others focus on practical performance as evaluated experimentally. These two groups of works are somewhat independent, in the sense that little experimental investigation is available in the former group, and little theoretical analysis exists for the latter. Moreover, the experimental results seem to be much better than could be expected from the theoretical results. We show that a broad class of branch and bound algorithms for the maximum clique problem display sub-exponential average running time behavior, and also show how this helps to explain the apparent discrepancy between the theoretical and experimental results. We also propose a more structured methodology for the experimental analysis of algorithms for the maximum clique problem, which takes into account the peculiarities of cliques in random graphs, bringing the theoretical and experimental approaches closer together in the search for better algorithms. As a proof of concept, we apply the proposed methodology to thirteen algorithms from the literature.

1. Introduction

The maximum clique problem is the problem of finding a clique of maximum size on a given graph. In addition to being a fundamental $\mathcal{NP}$-hard problem, it is also used to model important applications in several domains [5].

There has been increasing interest in the development of exact (albeit exponential time) algorithms for $\mathcal{NP}$-hard problems [13,17,45–47]. In the case of the maximum clique problem, some works can be found in the literature (see Section 2) whose main concern is to better characterize the complexity of the problem, and others which propose and study the performance of particular algorithms by implementing them, subjecting this implementation to experimentation and reporting the results. Curiously enough, the intersection between these two types of work is relatively small, in the sense that, on the one hand, little experimental investigation is available for proposed algorithms in the former group, while on the other hand, no theoretical analysis exists for the algorithms proposed in the latter. Moreover, the results reported in these works may leave the reader with the impression that the performance of the experimental results is much better than could be expected from the theoretical results alone. Indeed, despite the fact that the theoretical results indicate that this problem is among the less tractable ones [3,10], several of the proposed algorithms are reported as solving instances of practical interest and considerable size in various application domains quite satisfactorily [7,11,19,40].

* Corresponding author.
E-mail address: alexandrezuge@ufpr.br (A.P. Züge).

https://doi.org/10.1016/j.dam.2018.01.005
0166-218X/

Please cite this article in press as: A.P. Züge, R. Carmo, On comparing algorithms for the maximum clique problem, Discrete Applied Mathematics (2018), https://doi.org/10.1016/j.dam.2018.01.005.

We offer an explanation for this apparent discrepancy through a closer inspection of the methodology used in the experimental works, considered in the light of certain theoretical results. Based on these same results, we propose a more structured approach to the experimental analysis of algorithms for the maximum clique problem which avoids this discrepancy and, in doing so, contributes towards bringing the theoretical and experimental efforts closer together in the search for better algorithms.

In Section 2, we summarize and briefly discuss some known results, theoretical and experimental, for the maximum clique problem and algorithms for its solution. In doing so, we comment on the methodology used to obtain experimental results for these algorithms, highlighting the main concern of this work.

In Section 3, we discuss a class of branch and bound algorithms for the maximum clique problem. We note that some of the best performing algorithms from an experimental point of view belong to this class. We then show how to derive some bounds on the average running time of branch and bound algorithms in this class, based on an analysis of another family of algorithms introduced in [8] and further studied in [29].

In Section 4, we propose a more structured methodology for the experimental analysis of algorithms for the maximum clique problem, grounded in the results discussed in Section 3. This methodology is insensitive to implementation details and provides a metric against which different algorithms can be ranked and compared to each other. By means of a proof of concept, we show the result of applying this methodology to thirteen different algorithms for the maximum clique problem.

1.1. Definitions and notation

We use $\lg x$ to denote $\log_2 x$. Given a set S and an integer k we denote by $\binom{S}{k}$ the set of subsets of S of size k and denote the set of subsets (i.e., the power set) of S by 2^S.

A *graph* G is a pair $(V(G), E(G))$ where $V(G)$ is a finite set of *vertices* and $E(G) \subseteq \binom{V(G)}{2}$. Each element of $E(G)$ is called an *edges* of G. Two vertices u and v are *neighbors* in G if $\{u, v\} \in E(G)$. The *neighborhood* of v in G is the set of its neighbors in G and is denoted by $\Gamma_G(v)$. The *common neighborhood* of a non-empty set of vertices $Q \subseteq V(G)$ is the intersection of the neighborhoods of the vertices in Q; that is, it is the set $\bigcap_{v \in Q} \Gamma_G(v)$.

The graph G is *complete* if any two vertices of G are neighbors. The *complement* of graph G is the graph $\overline{G}$ given by $V(\overline{G}) = V(G)$ and $E(\overline{G}) = \binom{V(G)}{2} - E(G)$.

If $X \subseteq V(G)$, then $G[X]$ is the *subgraph of G induced by X* given by $V(G[X]) = X$ and $E(G[X]) = \binom{X}{2} \cap E(G)$. The set X is a *clique* if the graph $G[X]$ is complete and is *independent* if $G[X]$ has no edges. The maximum size of a clique in G is denoted by $\omega(G)$ and the set of all cliques in G is denoted $\mathcal{C}(G)$.

The maximum clique problem, denoted as MC, is the problem of finding a clique of maximum size on a given graph. The maximum independent set problem, denoted as MIS, is the problem of finding an independent set of maximum size on a given graph. Both problems are equivalent in the sense that the set C is a solution for the instance G of MC if and only if C is a solution for instance $\overline{G}$ of MIS.

Given a positive integer n and $p \in [0, 1]$, the $\mathcal{G}_{n,p}$ model of random graphs is the probability space of the graphs on n vertices, where each possible edge exists with probability p. We note that $\mathcal{G}_{n,1/2}$ is the uniform probability space generated by the graphs on n vertices. Throughout the text, when referring to random graphs, we mean the $\mathcal{G}_{n,p}$ model for some chosen values of n and p, where p is a constant (as opposed to a function of n).

2. Exact solution of the maximum clique problem

The maximum clique problem is $\mathcal{NP}$-hard [15] and cannot be approximated in polynomial time to within $|V(G)|^{1-\varepsilon}$, for all $\varepsilon > 0$, unless $\mathcal{P} = \mathcal{NP}$ [50].

The decision problem associated with MC is known as the clique problem; given a graph G and an integer k, it is the problem of deciding whether or not G has a clique of size (at least) k. The clique problem is $\mathcal{NP}$-complete. Moreover, its natural parameterization, where the integer k is the parameter, is $W[1]$-complete [10].

A graph on n vertices can have as many as $3^{n/3}$ distinct maximal cliques [26]. Therefore, any algorithm which enumerates all maximal cliques of a graph on n vertices must have a worst case running time of $\Omega(3^{n/3})$. An algorithm for enumerating all the maximal cliques of a graph matching this bound with a worst case running time of $O(3^{n/3})$ was introduced in [43]. A refined version with better practical performance is introduced in [27]. An $O(|\mathcal{C}(G)| \cdot |E(G)|)$ time algorithm for the same problem is presented in [16].

Finding a maximum clique on a graph does not require one to actually examine all of its maximal cliques; some of the maximal cliques can be discarded in the enumeration process using the technique known as *branch and bound* [20]. A substantial number of cliques can be discarded in this way, as shown in [39], which introduces an algorithm for MIS whose worst case running time is $O(2^{n/3})$. This bound was later improved to $O(2^{0.304n})$ [18] and further to $O(2^{0.276n})$ [31] at the expense of exponential space consumption. The value of this bound is currently set at $O(2^{n/4})$ [32], also with exponential

Please cite this article in press as: A.P. Züge, R. Carmo, On comparing algorithms for the maximum clique problem, Discrete Applied Mathematics (2018), https://doi.org/10.1016/j.dam.2018.01.005.