European Journal of Operational Research 242 (2015) 693–709

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Invited Review

A review on algorithms for maximum clique problems

Qinghua Wu[a], Jin-Kao Hao[b,*]

[a] School of Management, Huazhong University of Science and Technology, No. 1037, Luoyu Road, Wuhan, China
[b] LERIA, Université d'Angers, 2 bd Lavoisier, 49045 Angers, Cedex 01, France

ARTICLE INFO

Article history:
Received 23 November 2013
Accepted 28 September 2014
Available online 17 October 2014

Keywords:
Maximum clique problems
Exact algorithms
Heuristics
Applications

ABSTRACT

The maximum clique problem (MCP) is to determine in a graph a clique (i.e., a complete subgraph) of maximum cardinality. The MCP is notable for its capability of modeling other combinatorial problems and real-world applications. As one of the most studied NP-hard problems, many algorithms are available in the literature and new methods are continually being proposed. Given that the two existing surveys on the MCP date back to 1994 and 1999 respectively, one primary goal of this paper is to provide an updated and comprehensive review on both exact and heuristic MCP algorithms, with a special focus on recent developments. To be informative, we identify the general framework followed by these algorithms and pinpoint the key ingredients that make them successful. By classifying the main search strategies and putting forward the critical elements of the most relevant clique methods, this review intends to encourage future development of more powerful methods and motivate new applications of the clique approaches.

1. Introduction

The maximum clique problem (MCP) is to find a complete subgraph of maximum cardinality in a general graph. Its decision version is among the first 21 NP-complete problems presented in Karp's seminal paper on computational complexity (Karp, 1972). The MCP is among the most studied combinatorial problems.

The MCP has a wide range of practical applications in numerous fields. Early applications can be found for instance in Ballard and Brown (1982); Barahona, Weintraub, and Epstein (1992) and Christofides (1975) and are surveyed in Bomze, Budinich, Pardalos, and Pelillo (1999) and Pardalos and Xue (1994). Nowadays, more and more practical applications of clique problems arise in a number of domains including bioinformatics and chemoinformatics (Malod-Dognin, Andonov, & Yanev, 2010; Ravetti & Moscato, 2008), coding theory (Etzion & Östergård, 1998), economics (Boginski, Butenko, & Pardalos, 2006), examination planning (Carter, Laporte, & Lee, 1996; Carter & Johnson, 2001), financial networks (Boginski et al., 2006), location (Brotcorne, Laporte, & Semet, 2002), scheduling (Dorndorf, Jaehn, & Pesch, 2008; Weide, Ryan, & Ehrgott, 2010), signal transmission analysis (Chen, Zhai, & Fang, 2010), social network analysis (Balasundaram, Butenko, & Hicks, 2011; Pattillo, Youssef, & Butenko, 2012), wireless networks and telecommunications (Balasundaram & Butenko, 2006; Jain, Padhye, Padmanabhan, & Qiu, 2005). In addition to these applications, the MCP is tightly related to some important combinatorial optimization problems such as clique partitioning (Wang, Alidaee, Glover, & Kochenberger, 2006), graph clustering (Schaeffer, 2007), graph vertex coloring (Chams, Hertz, & Werra, 1987; Wu & Hao, 2012a), max-min diversity (Croce, Grosso, & Locatelli, 2009), optimal winner determination (Shoham, Cramton, & Steinberg, 2006; Wu & Hao, 2015), set packing (Wu, Hao, & Glover, 2012) and sum coloring (Wu & Hao, 2012b). These problems can either be directly formulated as a maximum clique problem or have a sub-problem which requires to find a maximum clique.

Given its theoretical importance and practical relevance, considerable effort has been devoted to the development of various solution methods for the MCP. On the one hand, effective exact methods have been designed mainly based on the general branch-and-bound (B&B) framework. These methods have the theoretical advantage of guaranteeing the optimality of the solution found. However, due to the inherent computational complexity of the MCP, exact methods can require a prohibitive computing time in the general case and are often applicable only to problems of limited sizes. On the other hand, to handle problems whose optimal solutions cannot be reached within a reasonable time, various heuristic and metaheuristic algorithms have been devised with the purpose of providing sub-optimal solutions as good as possible to large problems within an acceptable time. It is clear that exact and heuristic methods constitute two complementary solution approaches which can be applied to face different situations and fulfill different objectives. These two approaches can even be combined to create more powerful search methods.

* Corresponding author. +33 2 41 73 50 76.
E-mail addresses: qinghuawu1005@gmail.com (Q. Wu), hao@info.univ-angers.fr, jin-kao.hao@univ-angers.fr (J. K. Hao).

http://dx.doi.org/10.1016/j.ejor.2014.09.064
0377-2217/

Since the Second DIMACS Implementation Challenge dedicated to Maximum Clique, Graph Coloring, and Satisfiability organized during 1992–1993 (Johnson & Trick, 1996), studies on these NP-hard problems are becoming more and more intense. In particular, significant progresses have been achieved regarding the MCP, its important generalizations (e.g., maximum vertex weight clique and maximum edge weight clique) and relaxations (e.g., quasi-clique and densest k-subgraph). Advances on new algorithms have helped to find improved results to benchmark problems and deliver effective solutions to new applications (social network analysis, protein structure alignment, wireless network etc.).

At the same time, we observe that the two most influential surveys on the MCP date back to 1994 and 1999 respectively (Bomze et al., 1999; Pardalos & Xue, 1994). To the best of our knowledge, there is no updated review to report the newest advances achieved during the past 15 years. This paper thus aims to fill this gap by providing a detailed review of different solution approaches proposed in the recent literature for maximum clique problems. We will not only make a general and large survey of the most representative exact and heuristic algorithms, but also carry out an in-depth analysis of the studied methods to identify their most relevant ingredients that make these methods successful.

2. Definitions, problem formulations and computational complexity

Let $G = (V, E)$ be an undirected graph with vertex set $V = \{1, \ldots, n\}$ and edge set $E \subseteq V \times V$. A *clique* C of G is a subset of V such that every two vertices in C are adjacent, i.e., $\forall u, v \in C, \{u, v\} \in E$. A clique is *maximal* if it is not contained in any other clique, a clique is *maximum* if its cardinality is the largest among all the cliques of the graph. The *maximum clique problem* (MCP) is to find a maximum clique of a given graph in the general case. The clique number $\omega(G)$ of G is the number of vertices in a maximum clique in G.

The maximum clique problem is strictly equivalent to two other well-known combinatorial optimization problems: the *maximum independent set problem* (MIS) and the *minimum vertex cover problem* (MVC). Given $G = (V, E)$, an independent set (also called a stable set) I of G is a subset of V such that every two vertices in I are not connected by an edge, i.e., $\forall u, v \in I, \{u, v\} \notin E$. The MIS is to determine an independent set of maximum cardinality. A vertex cover V' of G is a subset of V, such that every edge $\{i, j\} \in E$ has at least one endpoint in V'. The MVC is to find a vertex cover of minimum cardinality.

Let $\overline{G} = (V, \overline{E})$ be the complementary graph of G such that $\{i, j\} \in \overline{E}$ if $\{i, j\} \notin E$. One observes that C is a maximum clique of G if and only if C is a maximum independent set of $\overline{G}$, and if and only if $V \backslash C$ is a minimum vertex cover of $\overline{G}$. An illustration of the relation between maximum clique, maximum independent set and minimum vertex cover is given in Fig. 1. Due to the close connection between the MCP and MIS, we will operate with both problems while describing the properties and algorithms for the MCP. Clearly a result which holds for the MCP in G will also be true for the MIS in $\overline{G}$. This paper focuses thus on the MCP (and the MIS) while putting aside the MVC which itself has a large body of studies in the literature.

On the other hand, one notices that some generalizations and relaxations of the MCP are attracting increasing attention recently due to their wide applications in some emerging areas like bioinformatics and social network analysis. We will discuss these cases in Section 5. The rest of this section is dedicated to the issue of problem formulations and complexity of the MCP.

There are numerous studies on the formulation of the MCP. These studies are of great interest since they can lead to deep understandings of the problem and the discovery of new results of theoretical and practical nature. For a comprehensive review of the existing formulations of the MCP, the reader is refereed to Bomze et al. (1999); Butenko (2003) and Pardalos and Xue (1994). Below, we review some typical and recently developed formulations.

The simplest formulation is given by the following binary program:

$$\text{maximize} \quad \sum_{i=1}^{n} x_i \tag{1}$$

$$\text{subject to} \quad x_i + x_j \leq 1, \quad \forall \{i, j\} \in \overline{E} \tag{2}$$

$$x_i \in \{0, 1\}, \quad i = 1, \ldots, n. \tag{3}$$

In this edge formulation, any feasible solution defines a clique C in G as follows: vertex i is in the clique if $x_i = 1$ and otherwise $x_i = 0$. The linear relaxation of this formulation is significant as well. If a variable $x_i = 1$ holds for an optimal solution to the linear relaxation of the above formulation, then $x_i = 1$ holds for at least one optimal solution to the integer formulation (Nemhauser & Trotter, 1975). Clearly this result can be used by an algorithm to reduce the explored search space when seeking an optimal clique.

Let S denote the set of all maximal independent sets in G, an alternative formulation based on independent sets imposes that any clique of G can contain no more than a single vertex from any maximal independent set of G:

$$\text{maximize} \quad \sum_{i=1}^{n} x_i \tag{4}$$

$$\text{subject to} \quad \sum_{i \in s} x_i \leq 1, \quad \forall s \in S \tag{5}$$

$$x_i \in \{0, 1\}, \quad i = 1, \ldots, n. \tag{6}$$

This formulation has the advantage that the expected gap between the optimal solution and its linear relaxation is small. However, it is difficult to enumerate all independent sets in an arbitrary graph. Furthermore, as the number of independent sets in the graph grows exponentially with the graph size, it is not trivial to solve the relaxation of the independent set formulation.

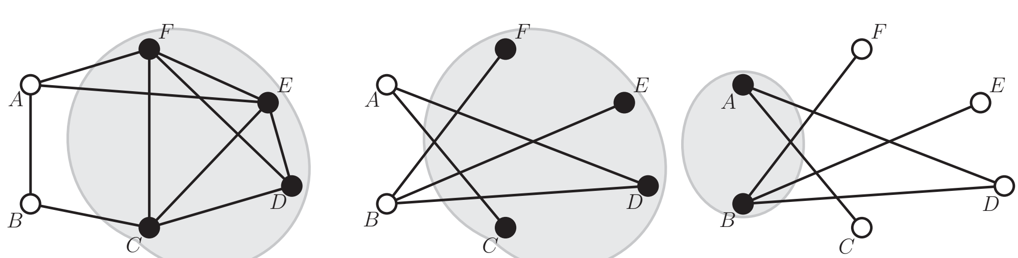

Fig. 1. An illustration of the relation between maximum clique, maximum independent set and minimum vertex cover. Given the initial graph G with $V = \{A, B, C, D, E, F\}$ (left) and its complementary graph $\overline{G}$ (middle/right), the set of vertices $\{C, D, E, F\}$ is a maximum clique of G and an maximum independent set of $\overline{G}$ while $\{A, B\} = V \setminus \{C, D, E, F\}$ is a minimum vertex cover of $\overline{G}$.