Cardinality Constrained Graph Partitioning into Cliques with Submodular Costs

J.R. Correa, [a] N. Megow, [b] R. Raman, [b] K. Suchan [c]

[a] *Departamento de Ingeniería Industrial, Universidad de Chile, Chile*
correa@dim.uchile.cl

[b] *Max-Planck-Institut für Informatik, Saarbrücken, Germany*
{nmegow, rraman}@mpi-inf.mpg.de

[c] *Universidad Adolfo Ibáñez, Facultad de Ingeniería y Ciencias, Chile*
karol.suchan@uai.cl

Key words: graph partitioning, cardinality constraint, submodular functions

1. Introduction

We consider the problem of partitioning a graph into cliques of bounded cardinality. The goal is to find a partition that minimizes the sum of clique costs where the cost of a clique is given by a set function on the nodes. We present a general algorithmic solution based on solving the problem variant without the cardinality constraint. We yield constant factor approximations depending on the solvability of this relaxation for a large class of submodular cost functions. We give optimal algorithms for special graph classes.

More formally, we are given a simple graph $G = (V, E)$, a set function $f : 2^V \to \mathbb{R}^+$, and a bound $B \in \mathbb{Z}^+$. The problem is to find a partition of the graph G into cliques $K_1, \ldots, K_\ell$ (the value of ℓ is not part of the input) of size at most B, that is, $|K_i| \leq B, i = 1, \ldots, \ell$, such that the objective function $\sum_{i=1}^{\ell} f(K_i)$ is minimized. We denote our problem as *partition into cliques of bounded size* PCliq(G, f, B).

Let V be a finite set. The function $f : 2^V \to \mathbb{R}$ is called *submodular* if for all subsets $A, B \subseteq V$ holds $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$. We consider non-negative submodular functions that satisfy the following *exchange properties*. For all subsets $A, B \subseteq V$ with $f(A) \geq f(B)$ and elements $u, v \in V \setminus (A \cup B)$ with $f(u) \geq f(v)$ holds

$$f(A + u) + f(B + v) \;\leq\; f(A + v) + f(B + u)\,. \qquad \text{(E1)}$$

Moreover, for all sets $A, B \subseteq V$ such $f(A) \geq f(B)$ and all elements $u \in V \setminus (A \cup B)$, holds that

$$f(A + u) \geq f(B + u). \tag{E2}$$

This class of set functions contains well-known cost functions such as maximum function [7], chromatic entropy [2], probabilistic coloring [3], etc.

The problem PCliq(G, f, B) is generally $\mathcal{NP}$-hard because it contains the classic $\mathcal{NP}$-hard problems *partition into cliques* (or *clique cover*) and *graph coloring* [5]. If the bound on the clique size B equals 2 then the problem corresponds to a maximum cardinality matching problem in G and can be solved optimally in polynomial time.

Graph partitioning and coloring problems (without cardinality constraints) are among the fundamental problems in combinatorial optimization. Various results are known for particular cost functions and graph classes. Recently, also generalized set functions have been considered in this context. Gijswijt, Jost, and Queyranne [6] introduce *value-polymatroidal* set functions which are *polymatroid rank functions* (i.e. nondecreasing, submodular, and $f(\emptyset) = 0$) that satisfy a slightly weaker exchange property than we require above. For every $A, B \subseteq V$ such that $f(A) \geq f(B)$ and every $u \in V \setminus (A \cup B)$, holds that $f(A+u) + f(B) \leq f(A) + f(B+u)$. They consider the problem PCliq(G, f, ∞) and derive polynomial time algorithms for interval graphs and circular arc graphs. Fukunaga, Halldorsson, and Nagamochi [4] consider monotone concave cost functions and provide a general algorithmic scheme which yields a factor 4 approximation for perfect graphs. They also give a result for general graphs depending on the solvability of the *maximum independent set* problem on the complement of the graph.

Graph partitioning (or coloring) with a constraint on the clique size has been addresses rarely so far. Bodlaender and Jansen [1] investigated a special case of PCliq(G, f, B) with the objective to minimize the number of cliques, that is, $f(S) = 1$ for all $S \subseteq 2^V$. They show that the decision variant of the problem on co-graphs is $\mathcal{NP}$-complete whereas it is polynomial time solvable on split graphs, bipartite graphs and interval graphs.

2. Our Results

2.1 Optimal algorithm for complete graphs and proper interval graphs

Consider the problem PCliq(K, f, B) on a complete graph K. The following simple algorithm solves the problem optimally.

Algorithm PARTK

Order elements $u \in K$ in non-increasing order of $f(u)$ and group them greedily into sub-cliques of size B beginning with the elements of largest value.

Theorem 2.1. PARTK solves the problem PCliq(K, f, B) on a clique K for submodular functions with exchange property optimally in polynomial time.

The exchange properties are indeed substantial to the result above. The problem of partitioning a complete graph is generally $\mathcal{NP}$-hard for submodular functions, and even for polymatroid rank functions.

Theorem 2.2. The problem PCliq(K, f, B) on a clique K is NP-hard even if we restrict f to polymatroid rank functions, that is, non-decreasing, submodular functions with $f(\emptyset) = 0$.

Proof 1. Reduction from *graph partitioning* with unit node weights [5].

Additionally, we devise a dynamic program that solves the problem optimal for another special graph class.

Theorem 2.3. The problem PCliq(G, f, B) on a proper interval graph G can be solved optimally in pseudopolynomial time for submodular functions with exchange property. If the number of distinct values for single elements $f(u)$ for all $u \in V$ is bounded, then this algorithm runs in polynomial time.

2.2 An $(c + 1)-$approximation for general graphs

Our algorithmic framework is based on solving two relaxation of the given problem PCliq(G, f, B). One relaxation concerns ignoring the graph structure, i.e., PCliq(K, f, B) for K being a complete graph as considered above. In the other relaxation we assume that the cardinality of the cliques is not bounded, or $B \geq |V|$. We denote it as PCliq(G, f, ∞). Optimal values for both relaxations considered individually may give arbitrarily bad lower bounds on an optimum solution. Still, we derive constant factor approximation guarantees when combining them. Consider the following polynomial time algorithm.

Algorithm PART

(i) Solve the relaxation PCliq(G, f, ∞) without cardinality restrictions.

(ii) For all cliques K_i: Solve the problem PCliq(K_i, f, B).

Theorem 2.4. Let f be a submodular function with exchange property. If there exists a c-approximation algorithm for the problem PCliq(G, f, ∞) without cardinality constraint, then PART is a $(c + 1)-$approximation for PCliq(G, f, B).

Sketch of Proof: Let $K_1, \dots, K_\ell$ be the solution of the relaxed problem

PCliq(G, f, ∞). For each of the cliques K_i let $K_i^1, \ldots, K_i^{\ell_i}$ denote the partition into subcliques (Step 2). Assume that all sets are indexed such that $f(K_i^j) \geq f(K_i^{j+1})$ for $j = 1, \ldots, \ell_i - 1$. Then the value of the algorithms solution is

$$\begin{aligned} \textsc{Part} &= \sum_{i=1}^{\ell}\sum_{j=1}^{\ell_i} f(K_i^j) = \sum_{i=1}^{\ell} f(K_i^1) + \sum_{i=1}^{\ell}\sum_{j=2}^{\ell_i} f(K_i^j) \\ &\leq c\,\mathsf{Opt}(G, f, \infty) + \sum_{i=1}^{\ell}\sum_{j=2}^{\ell_i} f(K_i^j)\,. \end{aligned}$$

The main effort lies in proving $\sum_{i=1}^{\ell}\sum_{j=2}^{\ell_i} f(K_i^j) \leq \mathsf{Opt}(K, f, B)$. We first observe: For two sets $A, B \subseteq V$ with $|A| \geq |B|$ holds that if each element $v \in B$ can be mapped to a distinctive element in $u \in A$ such that $f(v) \leq f(u)$, then $f(A) \geq f(B)$. This fact combined with Algorithm PARTK and Theorem 2.1 allows us to apply a charging scheme where we map the elements of the cliques K_i^j with $i > 1$ to the optimal solution.

Since the set functions we consider are value-polymatroidal, can employ the optimal algorithm in [6] and yield a quite general result for interval graphs which are of particular interest in applications.

Corollary 2. There is a factor 2 approximation algorithm for PCliq(G, f, B) on interval graphs and circular arc graphs.

References

[1] H. L. Bodlaender and K. Jansen. Restrictions of graph partition problems. Part I. *Theoretical Computer Science*, 148(1):93–109, 1995.

[2] J. Cardinal, S. Fiorini, and G. Joret. Minimum entropy coloring. *J. Comb. Optim.*, 16(4):361–377, 2008.

[3] F. D. Croce, B. Escoffier, C. Murat, and V. T. Paschos. Probabilistic coloring of bipartite and split graphs. In *ICCSA (4)*, volume 3483 of *Lecture Notes in Computer Science*, pages 202–211, 2005.

[4] T. Fukunaga, M. M. Halldórsson, and H. Nagamochi. Robust cost colorings. In *SODA'08*, pages 1204–1212, 2008.

[5] M. R. Garey and D. S. Johnson. *Computers and Intractability, A Guide to the Theory of NP-Completeness*. W.H. Freeman and Company, New York, 1979.

[6] D. Gijswijt, V. Jost, and M. Queyranne. Clique partitioning of interval graphs with submodular costs on the cliques. *RAIRO Oper. Res.*, 41:275–287, 2007.

[7] S. V. Pemmaraju and R. Raman. Approximation algorithms for the max-coloring problem. In *ICALP'05*, volume 3580 of *Lecture Notes in Computer Science*, pages 1064–1075, 2005.