Maximum k-intersection, edge labeled multigraph max capacity k-path, and max factor k-gcd are all NP-hard

Staal A. Vinterbo[a,b]

[a] *Decision Systems Group, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, USA*

[b] *Harvard Medical School, Boston, MA, USA*

staal@dsg.harvard.edu

DSG Technical Report DSG TR 2002/12

Abstract

Proofs of NP-hardness of the maximum k-intersection, edge labeled multigraph max capacity k-path, and max factor k-gcd problems are presented.

Keywords: time complexity, NP-hardness, optimization

1 Introduction

The problem of finding k out of a collection of sets such that their intersection is maximized has previously been investigated by Vinterbo [1] in the context of disclosure control. Here we restate the hardness proof of this problem, and use it as a problem from which we find polynomial time reductions to two other NP-optimization problems. These two problems are the edge labeled multigraph max capacity k-path problem and the max factor k-gcd problem.

2 The Problems

Ausiello et al. [2] formally define an optimization problem P to consist of a quadruple of objects $(\mathcal{I}_P, \mathcal{S}_P, m_P, g_P)$ where

1. $\mathcal{I}_P$ is the set of instances of problem P,

2. $\mathcal{S}_P$ is a function returning for each instance $x \in \mathcal{I}_P$ the set of feasible solutions $\mathcal{S}_P(x)$ to I_P,
3. the measure $m_P : \mathcal{I}_P \times \mathcal{S}_p(\mathcal{I}_P) \to \mathbf{Z}^+$, that for a given instance $x \in \mathcal{I}_P$ and a feasible solution $s \in \mathcal{S}_P(I)$ returns a positive integer $m_P(x, s)$, indicating the quality of the solution, and
4. $g_P \in \{\min, \max\}$, the goal, which indicates whether we wish to minimize the measure or maximize the measure over the feasible solutions of a particular instance.

Further, given an instance x of P, we denote by $m_P^*(x)$ the optimal value of instance x:

$$m_P^*(x) = g_P\{m_P(x, s) | s \in \mathcal{S}_P(x)\},$$

and the set of optimal solutions of instance x as the set

$$\mathcal{S}_P^*(x) = \{s \in \mathcal{S}_P(x) | m_P(x, s) = m_P^*(x)\},$$

i.e., the subset of the feasible solutions that exhibit the optimal value under m_P.

In general, we are interested in the constructive version of an optimization problem, that is to find both $m_P^*(x)$ and an element $s \in \mathcal{S}_P^*(x)$ given an instance x.

Using the definitions presented above, we now present the problems in turn.

2.1 Maximum k-intersection

The maximum k-intersection problem consists of finding k sets from a finite collection of sets such that the size of the intersection is maximal. This problem is essentially equivalent to the problem of finding k sets out of a collection that their union is minimal.

The related problem of maximizing the union is NP-hard and its approximation properties are discussed by Hochbaum et al. [3].

Problem 1 *(MkI) A maximization problem with*

- *instance: collection $[C_i]_{i \in J}$ where $C_i \subseteq U$, and positive integer $k \leq |J|$,*
- *feasible solutions: $\{I \subseteq J \mid |I| \geq k\}$, and*
- *measure: $| \cap_{i \in I} C_i|$.*

2.2 Max Factor k-gcd

The max factor k-gcd problem is the problem of finding k integers from a given set of integers such that the gcd of these k integers has the most positive divisors.

Let $\tau : \mathbf{Z}^+ \to \mathbf{Z}^+$ count the number of positive divisors of its argument. As an example, $\tau(4) = 3$ as 4,2,1 are the positive divisors of 4. Further, let $\gcd : 2^{\mathbf{Z}^+} \to \mathbf{Z}^+$ return the greatest common divisor of the positive integers in its argument. As an example, $\gcd(\{6, 9, 12\}) = 3$.

Problem 2 *(MF-k-gcd) A maximization problem with:*

- *instance: $S \subseteq \mathbf{Z}^+$, $|S| = n$, $k \in \mathbf{Z}^+$, $k \leq n$,*
- *feasible solutions $\{S' \subseteq S || S'| = k\}$,*
- *measure: $\tau(\gcd(S'))$.*

2.3 Edge labeled Multigraph max capacity k-path

An edge labeled multigraph is defined to be a multigraph where each edge is labeled with a label from a set L. An edge can only be connected to another edge if they have the same label. We assign each vertex an edge capacity and wish to find a path of length k that maximizes the combined capacity through the vertices.

Problem 3 *(**MGMC-k-path**) A maximization problem with:*

- *instance: Edge labeled multigraph $M = (V, E \subseteq V \times V \times S, S)$, vertice capacity function $c\colon V \times S \to \mathbf{N}$, positive integer $k < |V|$.*
- *feasible solutions: $\{(v_1, v_2, \dots, v_{k+1}) \in V^{k+1} | v_i = v_j \Rightarrow i = j \wedge (\exists s \in S)\ (v_i, v_{i+1}, s) \in E\}$,*
- *measure: $\sum_{s \in S} \min_{i=1}^{k+1} c(v_i, s)$.*

3 NP-hardness

We show NP-hardness of the problems by initially finding a polynomial time reduction to the maximum k-intersection problem from the well known NP-complete balanced complete bipartite subgraph problem. We then find polynomial time reductions from maximum k-intersection to edge labeled multigraph max capacity k-path and max factor k-gcd in turn. A graphical representation of the reduction process can be seen in Figure 1.

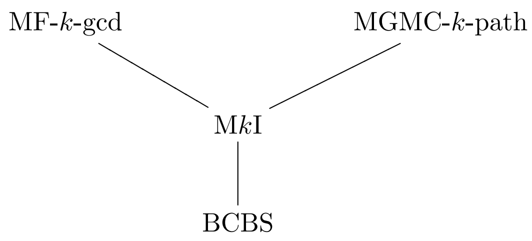

Figure 1: Reductions

Theorem 1 (Vinterbo [1]) *Maximum k-intersection is NP-hard.*

Proof: We prove that Problem 1 is NP-hard via a reduction from the NP-Complete Balanced Complete Bipartite Subgraph (BCBS) problem [4, GT24]: given a bipartite graph $G = (V, E)$ and a positive integer $k \leq |V|$, are there two disjoint subsets $V_1, V_2 \subseteq V$ such that $|V_1| = |V_2| = k$ and $V_1 \times V_2 \subseteq E$?

Let $G = (V, E)$ be a bipartite graph.

We start by proving that for $V_1 \subseteq V$ such that $|V_1| = k$, there exists $V_2 \subseteq V$ such that $|V_2| = k$ and $V_1 \times V_2 \subseteq E$ if and only if $|\cap_{i \in V_1} C_i| \geq k$, where $C_i = \{j | (i, j) \in E\}$.

- $\Rightarrow$ Assume that $|V_1| = |V_2| = k$ and $V_1 \times V_2 \subseteq E$. For each $i \in V_1$ we have that $V_2 \subseteq C_i$. As $|V_2| = k$, we have that $|\cap_{i \in V_1} C_i| \geq k$.
- $\Leftarrow$ Now assume that there exists V_1, $|V_1| = k$, such that $|\cap_{i \in V_1} C_i| \geq k$. As G is bipartite, we have that V_1 and $\cap_{i \in V_1} C_i$ are disjoint. Any set $V_2 \subseteq \cap_{i \in V_1} C_i$ results in $V_1 \times V_2 \in E$ by definition of C_i. We can choose $V_2 \subseteq \cap_{i \in V_1} C_i$ such that $|V_2| = k$ by assumption.

Now, let s be an optimal solution to the instance $x = (\{C_i\}_{i \in V}, k)$ of Problem 1. If $m(x, s) < k$, there exists no $V_1 \subseteq V$ of size k such that $|\cap_{i \in V_1} C_i| \geq k$, and, by the above, no balanced complete bipartite subgraph with $2k$ vertices. If $m(x, s) \geq k$, we can choose any $V_1 \subseteq s$ such that $|V_1| = k$ and any $V_2 \subseteq \cap_{i \in s} C_i$ such that $|V_2| = k$ and have that $V_1 \times V_2 \subseteq E$. The proof is concluded by noticing that we can compute C_i in polynomial time in the size of G for each $i \in V$. ■

Theorem 2 *Max factor k-gcd is NP-hard.*

Proof: We show that the max factor k-gcd is NP-hard by a reduction from the maximum k-intersection problem.

The fundamental theorem of arithmetic states that every positive integer $n > 1$ can be expressed as a product of primes uniquely up the the order of the factors. A positive integer is square free if each prime factor (apart from 1) only occurs once in this product. This means that a positive square free integer can be represented by a ordered set of 0's and 1's, each indicating the absence or presence of a particular prime as a factor. The gcd of a set of square free positive integers can be computed using this representation by the intersection of the corresponding sets. Let b_i be the element in position i in our representation of square free positive integers. For a square free positive integer n, we have that

$$\tau(n) = 2^{(\sum b_i)}.$$

Let $([C_i]_{i \in J}, k)$ be an instance of the maximum k-intersection problem, let $S = \cup_{i \in J} C_i$, and let $\pi : S \to \mathbf{P}$ be any injective function from S to the set of primes $\mathbf{P}$. We can now in a well defined manner associate a square free positive integer $\nu(C_i)$ with each set C_i using the injection $\nu(C_i) = \prod_{s \in C_i} \pi(s)$, in essence meaning that

$$\nu^{-1}(\gcd(\{\nu(j) | j \in J'\})) = \cap_{j \in J'} C_j,$$

and

$$\tau(\gcd(\{\nu(j) | j \in J'\})) = 2^{|\cap_{j \in J'} C_j|}$$

for $J' \subseteq J$. As 2^x is strictly monotonous in x, we have the wanted reduction. ■

Theorem 3 *Edge labeled multigraph max capacity k-path is NP-hard.*

Proof: We show that the edge labeled multigraph max capacity k-path is NP-hard by a reduction from the maximum k-intersection problem. Let $([C_i]_{i \in J}, k + 1)$ be an instance of the maximum k-intersection problem. Let $S = \cup_{i \in J} C_i$, $V = J$, and $E = V \times V \times S$. Define $c(i, s) = 1$ if $s \in C_i$ and 0 otherwise. Since the multigraph $G = (V, E, S)$ is complete, any permutation of the vertices of a length k path is a valid path, and furthermore, shares the same capacity. A k length path has $k + 1$ vertices, and the measures of the two problems coincide as min on the characteristic functions c for each i in the solution is analogous to intersecting the corresponding sets C_i. ■

Acknowledgments

This work was supported by grant R01-LM07273 from the National Library of Medicine.

References

[1] Staal A. Vinterbo, "A note on the hardness of the k-ambiguity problem," Tech. Rep. DSG-TR-2002-006, Decision Systems Group/Harvard Medical School, 75 Francis Street, Boston, MA 02115, USA, May 2002.

[2] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, and M. Protasi, *Complexity and approximation: combinatorial optimization problems and their approximability properties*, Springer-Verlag, 1999.

[3] Dorit Hochbaum and Anu Pathria, "Analysis of the greedy approach in covering problems," *Naval Research Quarterly*, , no. 45, pp. 615–627, 1998.

[4] 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, NY, 1979.