Systems of sets and their representatives *

Jiří Fiala[1] Jan Kratochvíl[1†] Andrzej Proskurowski[2‡]

[1] Institute for Theoretical Computer Science[§]
and Department of Applied Mathematics,
Charles University, Prague
{fiala,honza}@kam.mff.cuni.cz
[2] Department of Computer and Information Science,
University of Oregon, Eugene
andrzej@cs.uoregon.edu

Abstract

We introduce a new notion of Systems of Distant Representatives of families of subsets of a metric space. We are in particular interested in the computational complexity of deciding the existence of such systems, for different distance parameters and for various metric spaces. The problem contains as a subproblem the well known polynomial time solvable problem of Systems of Distinct Representatives (for discrete metric and distance parameter 1). We prove several NP-hardness results, e.g., for discrete metric and distance parameter 2, or for Euclidean metric spaces. We also show a direct connection to practically motivated and previously studied problems such as scheduling, distance constrained graph labeling, map labeling, disjoint representatives of hypergraphs and independent sets in graphs.

1 Introduction

Consider a universe X and a family of its subsets $\mathcal{M} = \{M_i | \; i \in I, M_i \subseteq X\}$. The *System of Distinct Representatives* (SDR) for $\mathcal{M}$ selects from each M_i an element (its *representative*), such that distinct sets are represented by

*The authors acknowledge support of joint Czech U.S. grants KONTAKT ME338 and NSF-INT-9802416 during visits of the first two authors to Eugene, OR, and of the third author to Prague.

†Research partially supported by Czech Research grant GAUK 158/99.

‡Supported in part by the grant NSF-ANI-9977524.

§Supported by the Ministry of Education of the Czech Republic as project LN00A056.

distinct elements. Such assignment can be formally described by an injective mapping $f : I \to X$ satisfying the property $f(i) \in M_i$ for all $i \in I$.

An equivalent interpretation of SDR is by a matching in a bipartite graph, where vertices of one block of the bi-partition correspond to the elements of X, the other block represents the family $\mathcal{M}$ and edges describe the membership of elements in sets of $\mathcal{M}$.

The theory of Systems of Distinct Representatives is well known and very important for discrete optimization problems. When the family $\mathcal{M}$ and all its sets M_i are finite, the elegant Hall theorem [12] describes the necessary and sufficient condition for the existence of an SDR, which – if it exists – can be found by a polynomial time algorithm (augmenting paths or the matching algorithm of Edmonds [5]).

Though several generalizations of the concept have been studied, we believe that the concept of *distant* representatives is new. We assume that X is equipped by a metric $\mathrm{dist}(a, b)$ and therefore we may ask for representatives that are sufficiently spaced in X. More formally:

Definition 1 *Given a parameter $q > 0$ and a family $\mathcal{M} = \{M_i | \ i \in I\}$ of subsets of X, a mapping $f : I \to X$ is called a* System of q-Distant Representatives *(shortly an Sq-DR) if*

(1) $f(i) \in M_i$ *for every* $i \in I$,

(2) $\mathrm{dist}(f(i), f(j)) \geq q$ *for every* $i, j \in I, i \neq j$.

The metric on the space X could be defined in several ways, for example:

- The *trivial* metric: For arbitrary universe X we set $\mathrm{dist}(a, b) = 0$ if $a = b$, and $\mathrm{dist}(a, b) = 1$ otherwise.
- The *integral* metric: $X = \mathbb{N}$, $\mathrm{dist}(a, b) = |a - b|$.
- The *graph* metric: Take $X = V(G)$ and set $\mathrm{dist}(a, b)$ to the length of the shortest path connecting vertices a and b in a graph G.
- The L_i metric in the d-dimensional Euclidean space $X = \mathbb{R}^d$. In particular we consider the Euclidean plane $X = \mathbb{R}^2$ and
  - the L_1 (Manhattan) metric defined as the sum of coordinate differences $|a_x - b_x| + |a_y - b_y|$.
  - the L_2 metric. It is the usual Euclidean metric equal to the length of the segment joining the points a and b in the plane.

– the L_∞ metric defined as maximum coordinate difference, i.e. $\mathrm{dist}(a, b) = \max\{|a_x - b_x|, |a_y - b_y||\}$

Observe that in the case of trivial, integral or graph metric the System of q-Distant Representatives is equivalent to System of Distinct Representatives as long as $q \leq 1$. This holds also in a slightly more general setting: Consider any universe X with a metric for which $\mathrm{dist}(a, b) \geq 1$ whenever $a \neq b$. Then for $q \leq 1$ the condition (2) merely says that, for $i \neq j$, $f(i) \neq f(j)$, i.e., a System of q-Distant Representatives is a System of Distinct Representatives. For this case any further metric structure on X becomes irrelevant and in such a case an Sq-DR can be found by a polynomial-time algorithm.

A more general concept of *Systems of Disjoint Representatives on Hypergraphs* (SDRH) was studied by Aharoni and Haxell [1]. They provide neccesary and sufficient conditions for the existence of such systems for finite families $\mathcal{M}$. Although these conditions can be directly translated for instances of the Sq-DR problem with finite sets M_i, we show that in general they cannot be verified in polynomial time (unless *P*=*NP*).

The aim of this paper is to show that for the ranges $q \leq 1$ and $q > 1$ the problem Sq-DR behaves quite differently. This is shown in the case of integral metric in Section 2 and for unit diameter balls in the plane (with all three plane metrics) in Section 3. We show that in many cases, Systems of q-Distant Representatives can be modeled as *Systems of Independent Representatives* (SIR) in graphs and we characterize the computational complexity of this problem for various graph classes in Section 4.

In Section 5 we relate our work to the previous work on disjoint representatives on hypergraphs ([1]). Finally, in Section 6 we discuss the application of Sq-DR in the λ-coloring problem on precolored trees.

2 Systems of distant representatives on $\mathbb{N}$

We first focus our attention on the integral distance. We show that for $q > 1$, the linear ordering of the elements of X becomes the determining factor.

We will show that deciding whether a given family has an Sq-DR is *NP*-complete for $q > 1$. This fact is interesting on its own but for the purpose of λ-colorings (the application discussed in Section 6) we need a somewhat stronger result. We call a set of numbers *t-sparse* if $|x - y| \geq t$ for every two distinct members x, y of the set.

Theorem 2 *For every $q > 1$ and every t, it is NP-complete to decide if a family $\mathcal{M}$ of t-sparse sets of integers has a System of q-Distant Representatives.*

Proof We reduce from 3-Satisfiability of Boolean formulas in conjunctive normal form. This problem is known to be *NP*-complete even when restricted to formulas whose each clause contains 2 or 3 literals and every variable occurs in exactly 3 clauses — once positive and twice negated [9]. Suppose $\Phi = (V, C)$ is such a formula (where V is its variable set and C its clause set). We assume that the variables are numbered $x_1, x_2, \ldots, x_n$.

Fix a number $s > t + q \geq 2q$ (we may assume without loss of generality that $t \geq q$). For a variable x_i, let clause c contain the positive occurrence of x_i and let clauses d, e contain $\neg x_i$. Denote $x_i(c) = (i-1)s + 2$, $x_i(d) = (i-1)s + 1$ and $x_i(e) = (i-1)s + q + 1$. For every clause $c \in C$ create a 3-element set $M_c = \{x_i(c) | x_i \in c \text{ or } \neg x_i \in c\}$.

Observe first that the sets M_c, $c \in C$ are t-sparse. This is because the smallest difference of any two numbers $x_i(c), x_j(d)$, $i \neq j$ is at least $s - q > t$.

We claim that $\mathcal{M} = \{M_c\}_{c \in C}$ contains an Sq-DR if and only if Φ is satisfiable. Suppose first that $\mathcal{M}$ contains an Sq-DR f. We define a truth valuation ϕ of the variables V so that

- $\phi(x_i) = \mathsf{true}$ if $f(c) = x_i(c)$ for some clause c such that $x_i \in c$,
- $\phi(x_i) = \mathsf{false}$ if $f(c) = x_i(c)$ for some clause c such that $\neg x_i \in c$,
- $\phi(x_i)$ is arbitrary if none of the above applies.

Obviously every clause is satisfied by this assignment. We have to show, though, that ϕ is defined correctly. Assume to the contrary that for some variable x_i, $\phi(x_i)$ should be both true and false. That means that $(i-1)s+2$ and at least one of $(i-1)s+1$ and $(i-1)s+q+1$ are both representatives of some sets. This is, however, impossible, since their difference $(i-1)s + 2 - ((i-1)s+1) = 1$ or $(i-1)s + q + 1 - ((i-1)s+2) = q - 1$ would be less than q.

On the other hand, suppose that Φ is satisfied by a truth valuation ϕ. For every clause c, pick a variable (say x_i) which satisfies c and set $f(c) = x_i(c)$. This is clearly a set of representatives of $\mathcal{M}$. To see that the system is q-distant, note that if different clauses c and d are satisfied by different variables then the difference of $f(c)$ and $f(d)$ is at least $s - q > t \geq q$, while if c and d are satisfied by the same variable, say x_i, then this variable must occur negated in both clauses, and hence $f(c) = (i-1)s+1$ and $f(d) = (i-1)s+q+1$ (or vice versa), and thus their difference is q. $\square$

The problem Sq-DR on integers becomes tractable if we include an additional requirement that all sets in the family $\mathcal{M}$ are intervals. In such a situation we can extend the universe X to real numbers and formulate the problem in terms of scheduling: Let I be a set of tasks to be executed on one processor without preemption. Each task's processing time has the same length $(\frac{q}{2})$ which is to be scheduled within the given release time r_i and due time d_i. The corresponding sets of the family $\mathcal{M}$ are then $M_i = (r_i + \frac{q}{2}, d_i - \frac{q}{2})$. The question is whether for a given instance a feasible schedule exists.

Clearly, the polynomial-time algorithm solving the scheduling problem due to Simons [19] provides also a solution for the Sq-DR problem on $\mathbb{N}$ (or $\mathbb{R}$) with the integral metric.

3 Systems of distant representatives in the plane

In this section we consider the Euclidean plane as the universe X and sets of its points as the family $\mathcal{M}$. If all sets M_i are finite, we can transform such an instance to a finite graph G as will be described in Section 4. We disregard for the moment this case and focus our attention on infinite sets M_i, namely closed, unit diameter balls in $\mathbb{R}^2$ (we call them simply *unit balls*). We note that each such an infinite set can be described in constant space, e.g., by the coordinates of the center of the ball.

Observe that for all three considered plane metrics, two points a and b are at distance smaller than q if the two balls of diameter q, one centered at a and the other at b, intersect. The Sq-DR problem for closed unit balls in $\mathbb{R}^2$ has thus a nice geometric representation: replace each closed unit ball M_i of $\mathcal{M}$ by an open ball M'_i of diameter $1+q$ and ask, whether each M'_i can be assigned a representative, a ball of diameter q completely placed inside M'_i, such that all these representatives are pairwise disjoint. In the rest of this section we use this representation $\mathcal{M}'$ of the Sq-DR problem, rather than the original sets $\mathcal{M}$. The reason is, that in this representation all essential properties can be captured by the inclusion relation and we do not need concern ourselves with the underlying metric.

It is clear that such problems can model practical applications like, for instance, map labeling, where sets M_i correspond to the possible label placement, and the representative of diameter q corresponds to the place for a label.

We note here that the one-dimensional version of finding a Sq-DR for (unit) balls on $\mathbb{R}$ is equivalent to the scheduling problem mentioned above, and thus admits an algorithm running in polynomial time. We show in this section that in the two-dimensional space problem becomes *NP*-complete for all of the L_1, L_2 and L_∞ metrics.

In the Euclidean metric a ball corresponds to a disk of the same diameter. In the L_1 and L_∞ metrics balls are squares. More precisely, in the Manhattan metric the diameter is the length of its diagonals (which are parallel with the coordinate axes) while in the L_∞ metric the diameter is the square's side length and squares are oriented in the ususal way. Hence, balls in the L_1 and L_∞ metrics are topologically equivalent and the S1-DR problem for unit balls in $\mathbb{R}^2$ with the Manhattan metric reduces to the corresponding problem with L_∞ metric by rotating all squares by $\frac{\pi}{4}$ and scaling the coordiantes by the factor $\sqrt{2}$.

Theorem 3 *The S1-DR problem is NP-hard for closed unit balls in the Euclidean plane with the L_∞ metric.*

Proof We show a reduction from the planar 3-SAT problem, whose *NP*-completeness was proved in [16]. An instance for this version of the satisfiability problem consists of a formula $\Phi = (V, C)$ in the conjunctive normal form, where each variable has one positive and two negative occurrences, each clause consists of two or three literals, and the incidence graph G of Φ is planar. The graph G has vertex set $V \cup C$ and an edge (x, c) belongs to $E(G)$ iff the clause c contains x or $\neg x$ as a literal.

We reduce Φ to an instance of S1-DR in the plane. The corresponding family $\mathcal{M}'$ contains squares of side length 2 (=$q + 1$). It consists of three parts: one part represents clauses ("clause gadgets"), the next variables ("variable gadgets"), and the third part are connectors joining clause and variable gadgets together.

A three-literal clause gadget consists of two squares both with the same placement. It is clear, that only two of the four quadrants (unit squares) could not be occupied by representatives whenever an S1-DR exists. Similarly, for a clause with only two literals we use three squares in a common position.

The variable gadget is schematically depicted in Fig. 1. The dashed lines represent the unit grid in the plane. Solid lines show the 16 squares which we call active, the bold lines indicate two squares in the same position. The gray area cannot be used for a possible representative of an active square —

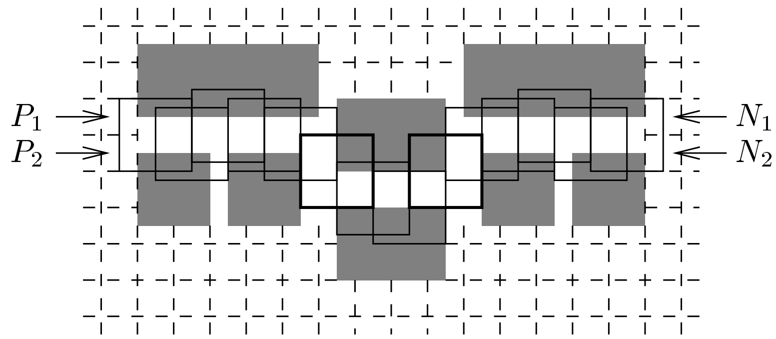

Figure 1: Variable gadget.

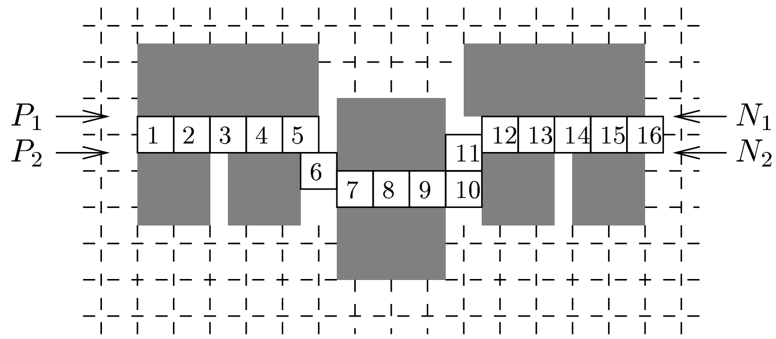

Figure 2: S1-DR for the variable gadget corresponding to the truth assignment.

it is fully occupied by four squares in the same position (thus each of them will be represented by a different quadrant).

The important property of the variable gadget is, that if any S1-DR exists, then either squares P_1 and P_2 or N_1 and N_2 (but not both pairs) will be partially occupied by the representatives for the variable gadget. For example, if neither P_1 nor P_2 is hit by the representatives, then N_1 and N_2 are, as is shown in Fig. 2. The small squares are representatives for the active squares of the gadget, their numbering corresponds to left-to-right ordering in Fig. 1.

The connector consists of a chain of squares. These squares come in triples so that all squares in a triple have the same position (thus, Fig. 3 actually represents 18 squares of side length 2). Two consecutive triples in the chain share one quadrant. Of the two quadrants at the ends of the chain, like A and B in the Fig. 3, at least one is fully occupied, if the chain has a

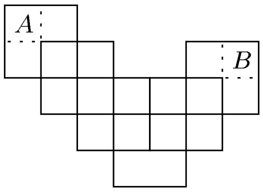

Figure 3: Connector gadget.

S1-DR.

In the collection $\mathcal{M}'$ we include for each clause $c \in C$ in Φ a disjoint copy of the clause gadget placed on a grid according to the planar drawing of G. Similarly we place the variable gadgets for each variable $x \in V$. If a variable x occurs positively in a clause c, we include a connector joining the unique quadrant of the clause gadget representing c with a lower right quadrant of the square P_1 related to the variable gadget of x. Similarly, we use left corners of N_1 and N_2 for the two possible occurrences of $\neg x$.

The placement of all these objects in the plane will follow the planar drawing of the graph G in a unit grid. All squares except those in the construction of the variable gadget coincide with unit grid, so the entire collection $\mathcal{M}$ can be constructed so that its size is polynomial in the size of the formula Φ.

Assume that an S1-DR for $\mathcal{M}'$ exists. For no clause gadget all its literals can be represented by a free quadrant, and thus at least one these literals is forced to be represented by the square P_1 (for a positive literal) or by some N_i (otherwise) of the corresponding variable gadget. As indicated above, it is impossible for a representative to occupy both P_1 and N_i at the same time. Hence we may define the truth assignment $x = \mathsf{true}$ if P_1 is not occupied by the representatives for active squares of the variable gadget for x and $x = \mathsf{false}$ otherwise. Such an assignment is well-defined and is valid for Φ. The construction of an S1-DR from the truth assignment of Φ is then straightforward.

□

In the proof of Theorem 3 only few squares were placed out of the unit grid. We show that when all unit squares respect this grid, the problem is solvable by a polynomial-time algorithm.

Corollary 4 *For any q of the form $\frac{1}{k}, k \in \mathbb{N}$, the Sq-DR problem for closed unit squares can be solved in polynomial time if all sides of squares in $\mathcal{M}$ coincide with the grid of span q.*

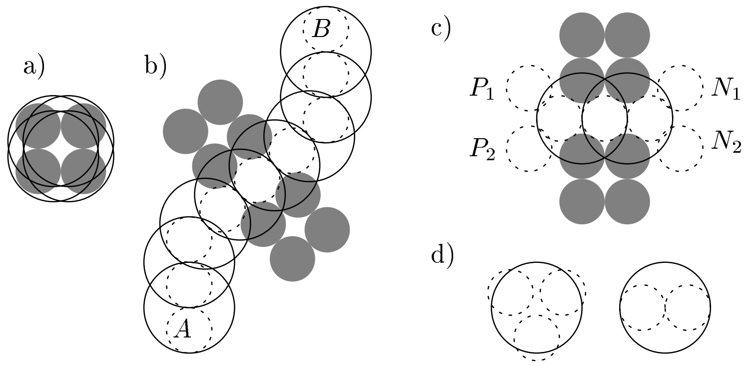

Figure 4: Gadgets for the S1-DR problem on disks.

Proof If any Sq-DR exists, then shift all the representatives to the left- and down-most position. The resulting Sq-DR has the property that all representatives coincide with the grid points.

Thus, we can without loss of generality restrict the universe X to the grid points and solve the problem by finding a System of Distinct Representatives. □

We conclude this section by proving an analogous results for the Euclidean metric.

Theorem 5 *The S1-DR problem is NP-hard for unit disks in the Euclidean metric.*

Proof The proof mimics the proof of Theorem 3. We represent a planar formula Φ by a collection of disks consisting of clause and variable gadgets, and connectors. As above we use the representation $\mathcal{M}'$, where representatives correspond to disks of diameter 1, while the disks in $\mathcal{M}'$ have diameter 2.

The first tool involved in the construction is the reserved gray area. Observe that the four disks depicted in Fig. 4 (a) allow an S1-DR only as the four gray disks.

The connectors are depicted in Fig. 4 (b). In no S1-DR of the connector gadget disks A and B may remain empty. The variable gadget is illustrated in the part (c). As in the previous proof, it is impossible to find representatives for the variable gadget where some P_i remains free together with some N_i. If the disk P_1 remains empty, the corresponding variable will be assigned the `true` value.

Finally, clause gadgets are shown in Fig. 4 (d). At least one of the three (or two) dashed disks cannot be left free to saturate the adjacent connector's end.

The remainder of the prof follows in the same manner as in the case of the L_∞ metric. $\square$

We believe that both of the above constructions can be adjusted for other fixed values of q.

Problem 1 Classify the computational complexity for the Sq-DR geometric problem with the L_1, L_2 and L_∞ metrics and q in the range $(0, \infty)$.

Our *NP*-completeness proof of the S1-DR problem can be adjusted also for higher dimensions d of the Euclidean space. Here a d-dimensional unit ball corresponds to

- the d-dimensional ball of radius $\frac{1}{2}$ for the Euclidean metric,
- the convex hull of $2d$ vectors in the case of Manhattan metric where each of these vectors has only one non-zero coordinate equal to $\pm\frac{1}{2}$,
- the d-dimensional cube of unit side length for the case of L_∞ metric.

The proof is based on a similar reduction from the satisfiability problem. It uses similar gadgets to represent variables and clauses, similar connectors and also alike geometric arguments of possible placement of the representants.

Problem 2 Find a uniform proof for the *NP*-hardness of the Sq-DR problem for unit balls in the d-dimensional Euclidean space for all L_i metrics ($q > 0, d \geq 2, i \geq 1$), where the L_i metric is defined as

$$\mathrm{dist}_{L_i}(a, b) = \sqrt[i]{\sum_{j=1}^{d} |a_j - b_j|^i}.$$

Problem 3 It remains as an open question whether the Sq-DR problem for unit balls belongs to the class *NP*. It is not straightforward to see whether the coordinates of the representatives — if they exist — could be always described and verified in polynomial space and time.

4 Systems of independent representatives in graphs

We have shown a significant difference in the computational complexity of Sq-DR for the integral metric when $q \leq 1$ and when $q > 1$.

In this section we observe that for any metric space X, the Sq-DR problem can be reduced to the S2-DR problem on graphs as follows: Given an instance $\mathcal{M}$ of Sq-DR, construct (as an instance of the S2-DR) a graph G on the (possibly infinite) vertex set $V(G) = X$ with $a, b \in X$ adjacent in G if and only if $\mathrm{dist}_X(a, b) < q$. The family $\mathcal{M}$ remains unchanged.

Then, vertices of G corresponding to an Sq-DR in X are independent (non-adjacent) in G and vice-versa. This leads naturally to the following version of Systems of q-Distant Representatives, a problem parametrized by a class $\mathcal{A}$ of graphs. The instance of this S$\mathcal{A}$-IR problem (*System of $\mathcal{A}$-Independent Representatives*) is a graph $G \in \mathcal{A}$ and a family $\mathcal{M}$ of subsets of the vertex set $V(G)$. The question is whether $\mathcal{M}$ allows a System of Distinct Representatives such that the representatives form an independent set in G. L. Lovász asked [personal communication] for a possible complete characterization of polynomial and NP-complete instances of the S$\mathcal{A}$-IR problem with respect to the parameter class $\mathcal{A}$.

Proposition 6 *For any graph class $\mathcal{A}$, the S$\mathcal{A}$-IR problem is at least as difficult as the Independent Set problem restricted to graphs of $\mathcal{A}$.*

Proof Given a graph $G \in \mathcal{A}$ and a number k, we take G and the multiset of k copies of $V(G)$ as the family $\mathcal{M}$. This family allows a System of Independent Representatives if and only if $\alpha(G) \geq k$. $\square$

This easy observation gives many examples of hard instances of our problem (e.g., planar graphs, triangle-free graphs etc.), but all of these are actually covered by the following corollary of our Theorem 2. For a graph G, we denote by $mv(G)$ the maximum k such that G contains k disjoint copies of $K_{1,2}$ as an induced subgraph.

Proposition 7 *If a graph class $\mathcal{A}$ is such that for every k, there is a $G \in \mathcal{A}$ with $mv(G) \geq k$, then the S$\mathcal{A}$-IR problem is NP-complete.*

Proof Take an instance of the S2-DR problem on 2-sparse sets of integers, as constructed in the proof of Theorem 2. Then the graph G corresponding to this situation is just the disjoint union of n copies of $K_{1,2}$, n being the number of variables in the formula the reduction started with. $\square$

We immediately see that Proposition 6 does not provide a tight lower bound on the complexity of the S$\mathcal{A}$-IR problem. Disjoint unions of $K_{1,2}$'s, and more generally the so called complement-reducible graphs (*cographs*) [4], allow a polynomial time algorithm for the Independent Set problem, but provide hard instances for our S$\mathcal{A}$-IR problem. It would be very interesting to know whether Proposition 7 completely characterizes the hard instances of S$\mathcal{A}$-IR. At present we are able to prove this dichotomy for classes $\mathcal{A}$ which are subclasses of cographs.

Before concluding this section with a formal statement, we mention a somewhat similar situation for the Independent Set problem. It has been showed by Tuza [6] that if a graph does not contain k disjoint copies of K_2 (denoted $k \cdot K_2$) as an induced subgraph, then the number of maximal (wrt. set inclusion) independent sets is polynomial in the number of vertices of the graph (the exponent depending on k). Consequently, since one can list all maximal independent sets in time polynomial in their number, for every fixed k, one can solve the Independent Set problem for graphs without k induced copies of K_2 in polynomial time.

For the *System of independent representatives* (SIR), we need to go a little further than enumerating the maximal independent sets, since e.g., if a graph G is the disjoint union of $\frac{n}{2}$ copies of K_2, G has n vertices and $\mathcal{O}(2^{\frac{n}{2}})$ maximal independent sets, but still the SIR problem on G can be solved in polynomial time, as lays the following observation.

Definition 8 *For a set A of vertices of a graph G, we say that the subgraph $G[A]$ induced in G by A is a* ducs *(disjoint union of complete subgraphs) if every connected component of $G[A]$ is a complete graph. We call a collection $\mathcal{D}$ of subsets of the vertex set of G an* iducs *if, for every $A \in \mathcal{D}$, $G[A]$ is a ducs and every independent set S in G is contained in some $A \in \mathcal{D}$. The minimum size of an iducs in G will be denoted by $md(G)$.*

Proposition 9 *If G has an iducs which can be constructed in polynomial time, then SIR on G can be solved in polynomial time.*

Proof Let $M = \{M_i : i \in I\}$, $M_i \subseteq V(G)$ be the input to our problem and $\mathcal{D}$ an iducs as stated. We list the elements of iducs $\mathcal{D}$ and for each $A \in \mathcal{D}$, we check if M has a SIR using only vertices of A. This we do as follows. If $A_j, 1 \leq j \leq t$, are the vertex sets of the connected components of $G[A]$, we collapse each A_j into a single vertex a_j, and set $M'_i = \{a_j : A_j \cap M_i \neq \emptyset\}$. Then we solve SDR for $M' = \{M'_i : i \in I\}$. If such an SDR exists for some $A \in \mathcal{D}$, we conclude that M has an SIR. We answer in negative in the opposite case.

This algorithm is clearly polynomial in the size of the problem and the time needed to construct the iducs $\mathcal{D}$. To argue its correctness, note that if M has an SIR, the representatives form an independent set in G, say S, which by the definition of the iducs is a subset of some $A \in \mathcal{D}$. For this particular A our algorithms finds an SDR of M'.

On the other hand, the existence of disjoint representatives on the collapsed instance M' guarantees that a system independent representatives for M exists as well. If $f' : I \to \{a_j : j = 1, \ldots, t\}$ is an SDR for M' based on $A = \bigcup_{j=1}^{t} A_j$, then for each i there is some $v_i \in A_j \cap M_i$, where j is such that $f'(i) = a_j$. Then $f(i) = v_i, i \in I$ is an SIR for M. □

Now we are ready to classify the computational complexity of S$\mathcal{A}$-IR for any $\mathcal{A}$, subclass of cographs. Recall that cographs are defined recursively so that (1) a single vertex is a cograph, (2) the disjoint union of cographs is a cograph and (3) the join (i.e., the disjoint union plus all edges between pairs of components) of cographs is a cograph. Each cograph G is then described by a rooted tree (called the cotree of G) in which every inner node is either a join or a sum node, and join and union nodes alternate on every path. Leaves of the cotree are in a bijective correspondence with the vertices of G. If we assume that every inner node has at least two children, such a tree is uniquely defined.

We will further call a join node a *strong join* node if it has at least one non-leaf child (which must then be a union node). A set of join nodes is called *union independent* if no two of them lie on the same shortest path from the root and the lowest common ancestor of any two of them is a union node.

Proposition 10 *For a cograph G, $mv(G)$ is equal to the maximum number of union independent strong join nodes in the cotree of G.*

Proof Suppose $G[A] = k \cdot K_{1,2}$ for some $A = \bigcup_{i=1}^{k} A_i$, where $k = mv(G)$ and each A_i induces a copy of $K_{1,2}$. Let u_i be the lowest common ancestor of the vertices in A_i (considered as leaves of the cotree). Because $G[A_i]$ is connected, u_i is a join node. The two non-adjacent vertices of $G[A_i]$ belong to the subtree rooted in the same child of u_i. Hence $\{u_i : i = 1, \ldots, k\}$ are strong join nodes. Since vertices of different A_i's are nonadjacent, these nodes are union independent.

On the other hand, suppose that $U = \{u_i : i = 1, \ldots, k\}$ is a set of union independent strong join nodes of the cotree. For each u_i, consider one of its union node children v_i and take two nonadjacent vertices, say x_i, y_i, belonging to the subgraph of G determined by the leaves that lie in the subtree rooted in v_i. Let z_i be a vertex of G belonging to the subgraph determined by the leaves of the subtree rooted in another child of u_i (u_i has at least two children). Then $G[\{x_i, y_i, z_i\}]$ is isomorphic to $K_{1,2}$ and since the nodes of U are union independent, there are no edges between vertices of different copies of $K_{1,2}$. Hence $mv(G) \geq |U|$. $\square$

Theorem 11 *Let $\mathcal{A}$ be a class of cographs. Then the following dichotomy holds:*

1. *for every k, $\mathcal{A}$ contains a graph G such that $mv(G) \geq k$, and the SA-IR problem is NP-complete;*

2. *there is a k such that $mv(G) \leq k$ for every $G \in \mathcal{A}$, and then the SA-IR problem is solvable in polynomial time (the exponent depending on k).*

The first part is just restated Proposition 7, the second part follows from the following lemma and Proposition 9.

Lemma 12 *For a cograph G, the minimum size of an iducs*

$$md(G) \leq |V(G)|^{mv(G)}$$

and such iducs can be constructed in $O(|V(G)|^{mv(G)})$ time.

Proof Let T be the cotree of G. We construct the optimum iducs $\mathcal{D}$ recursively as follows.

1. If T is a single leaf, G is a single vertex, $\mathcal{D}$ consists of the single one-element set $V(G)$, $mv(G) = 0$ and indeed $md(G) = |\mathcal{D}| = |V(G)|^0$.

2. If T has just one non-leaf node (the root) then G is either an edgeless graph (if the root is a union node), or G is a complete (if the root is a join node). In both cases $\mathcal{D}$ consists of the single independent set $V(G)$, $mv(G) = 0$ and $md(G) = |\mathcal{D}| = |V(G)|^0$.

3. If the root of T is a strong join node, G is the join of the subgraphs G_i corresponding to the children u_i of the root, say $i = 1, \dots, t$. In this case every independent set in G is fully contained in one of the G_i's, and hence $\mathcal{D} = \bigcup_{i=1}^{t} \mathcal{D}_i$ is an iducs for G, provided each $\mathcal{D}_i$ is an iducs for G_i. Hence
$$md(G) \leq \sum_{i=1}^{t} md(G_i).$$

Since the root itself forms a union indepondent set of strong join nodes (with one element only) and every other union independent set of strong join nodes lies in one of the subtrees rooted in the children of the root, we have

$$mv(G) = \max\{1, mv(G_1), \ldots, mv(G_t)\}.$$

Then we have

$$md(G) \leq \sum_{i=1}^{t} |V(G_i)|^{mv(G_i)} \leq \sum_{i=1}^{t} |V(G_i)|^{mv(G)} \leq |V(G)|^{mv(G)}$$

since $mv(G) \geq 1$.

4. If the root is a union node with at least one non-leaf child, G is the disjoint union of subgraphs G_i corresponding to the children u_i of the root, say $i = 1, \ldots, t$. In this case every independent set in G is itself the disjoint union of independent sets in particular G_i's. It follows that

$$\mathcal{D} = \{\bigcup_{i=1}^{t} D_i : D_i \in \mathcal{D}_i\}$$

(where $\mathcal{D}_i$ is an iducs for G_i) is an iducs for G, and hence

$$md(G) \leq \prod_{i=1}^{t} md(G_i).$$

Straightforwardly,

$$mv(G) = \sum_{i=1}^{t} mv(G_i),$$

and hence

$$md(G) \leq \prod_{i=1}^{k} |V(G_i)|^{mv(G_i)} \leq |V(G)|^{\sum_{i=1}^{t} mv(G_i)} = |V(G)|^{mv(G)}.$$

□

We have fully classified the computational complexity of the S$\mathcal{A}$-IR problem for all subclases of cographs.

We conjecture that the Proposition 7 gives good characterization for all *NP*-complete cases. To prove or disprove the conjecture one has to solve the following problem:

Problem 4 Find an polynomial time algorithm for all classes of graphs having bounded number of disjoint induced $K_{1,2}$. or construct such class of graphs $\mathcal{A}$ and show that the S$\mathcal{A}$-IR problem is *NP*-complete.

5 Relation to hypergraphs

We will show that the Systems of $\mathcal{A}$-Independent Representatives are related to Systems of Disjont Representatives on Hypergraphs (SDRH) of Aharoni and Haxell[1] by the line graph construction.

An instance of the SDRH problem consists of a family $\mathcal{H}$ of hypergraphs $H_i (i \in I)$ (over a common vertex set). The question is whether each hypergraph can be represented by one of its edges, such that the representatives are pairwise disjoint [1].

Proposition 13 *The SDRH problem is polynomially equivalent to the S$\mathcal{G}$-IR problem for the class $\mathcal{G}$ of all graphs.*

Proof We fist show a reduction from SDRH to S$\mathcal{G}$-IR:

For an instance $\mathcal{H}$ we construct a graph G as follows: we represent hyperedges of all H_i ($i \in I$) by vertices of G with two vertices adjacent whenever the corresponding hyperedges intersect. The set M_i (of the family $\mathcal{M} = M_i$) contains vertices that represent edges of H_i.

It is straightforward to check that any system of independent representatives for G and $\mathcal{M}$ is a system of disjoint representatives for $\mathcal{H}$ and vice-versa.

For the implication in the opposite direction, for each graph $G \in \mathcal{G}$, construct $\mathcal{H}$ by taking as its vertices the edges of G. For each $i \in I$, each hyperedge of the hypergraph H_i contains all edges incident with the vertex u, for each vertex $u \in M_i$.

Disjoint representatives of $\mathcal{H}$ correspond naturally to independent representatives in G: if the representatives for $H_i, H_j \in \mathcal{H}$ were to correspond to adjacent vertices $u, u' \in V_G$, then (u, u') would appear (as a vertex) in both of the representatives. $\square$

Observe that in the second part of the proof we have constructed a hypergraph, in which the maximum size of a hyperedge coresponds to the maximum degree in G. In general, this hypergraph may not be a graph. But the SDRH problem remains hard when restricted to graphs (viewed as 2-uniform hypergraphs), because $n \cdot K_{1,2}$ is a line graph of $n \cdot P_4$ for arbitrary n. Hence we may conclude with the following computational complexity charcterization of the SDRH problem on k-uniform hypergraphs.

Corollary 14 *The SDRH problem is*

1. *polynomially solvable for 1-uniform hypergaphs (i.e., set systems);*
2. *NP-complete when restricted to k-uniform hypergraphs, for every $k \geq 2$.*

This corollary shows that SDRH is indeed a more difficult problem than SDR. Though the necessary and sufficient condition for the existence of an SDR provided by the Hall theorem involves checking exponentially many inequalities, the existence of an SDR can be decided in polynomial time by well known techniques. Our result shows the SDRH problem, though characterized by the generalized Hall condition by Aharoni and Haxell[1], does not allow a polynomial time decision algorithm (unless P=NP).

6 The tree labeling problem

We show an application of Sq-DR in the distance constrained labelings of precolored trees. This notion stems from the radio frequency (or channel) assignment problem. Its graph theoretical model [13] asks for a labeling of the vertices of an input graph by nonnegative integers so that labels of vertices at distance at most i differ by at least p_i, for every $i \leq k$, where k and $p_1, \ldots, p_k$ are parameters of the problem. As a particular subproblem, Roberts proposed the problem of assigning integers (frequencies) to vertices (transmitters) such that vertices that are "fairly close" to each other (at distance two) receive different labels and vertices that are very close (adjacent) receive labels that are at least two apart. This corresponds to the case of $k = 2$ and $(p_1, p_2) = (2, 1)$, referred to as $(2, 1)$-labelings of graphs [3, 7, 11, 10, 15, 18, 20]. The more general two-parameter problem with $(p_1, p_2) = (p, q)$, $p \geq q > 1$, was considered in [2, 7, 8].

The minimum λ such that a graph G allows a (p, q)-labeling by integers from the range $\{0, 1, \ldots, \lambda\}$ is denoted by $\lambda_{(p,q)}(G)$. It was shown in [11, 20] that determining $\lambda_{(2,1)}(G)$ is an *NP*-complete problem even for graphs G with diameter two. The complexity of deciding $\lambda_{(2,1)}(G) \leq \lambda$ for fixed λ was shown *NP*-complete for every $\lambda \geq 4$ in [7]. It was also shown in [7] that for every $p \geq q \geq 1$, there is a λ (dependent on p, q) such that deciding $\lambda_{(p,q)}(G) \leq \lambda$ is *NP*-complete.

As concerns special graph classes, $\lambda_{(2,1)}(G)$ can be determined efficiently for paths, cycles and wheels [11], and for cographs and trees [3] (disproving the conjecture of [11] that the problem is *NP*-complete for trees). D. Welsh suggested [personal communication, 1999] that, by an algorithm similar to Chang and Kuo's, it should be possible to determine $\lambda_{(p,q)}(T)$ for a tree T for arbitrary p, q. The crucial step of the algorithm uses bipartite matchings (or Systems of Distinct Representatives, SDR), and an analogous algorithm for $q > 1$ would need to be able to decide existence of Systems of q-Distant Representatives.

As follows from Theorem 2 this problem is *NP*-complete in general. It is therefore plausible to conjecture that determining $\lambda_{(p,q)}(T)$ is *NP*-hard for trees, when $q > 1$. Note also that the complexity of determining $\lambda_{(2,1)}(G)$ for graphs of bounded tree-width is not known.

For graph coloring problems, it is natural to consider the *precoloring extension* variants of the problems where some vertices of the input graph are given as already (pre)colored, and the question is if this precoloring can be extended to a proper coloring of the entire graph using a given number of colors. (For several results on the complexity of precoloring extension for special graph classes see e.g. [14, 17].) The aim of this section is to consider the precoloring extension variant of (p, q)-labelings of trees. We show that at least in this setting the cases $q = 1$ and $q > 1$ behave quite differently. Chang and Kuo's algorithm can be easily extended to precolored trees and parameters $(p, 1)$, for any p. On the other hand, the problem is *NP*-complete for every $p > q > 1$.

Theorem 15 *For every fixed $p \geq q > 1$, it is NP-complete to decide if a precoloring of a tree can be extended to a (p, q)-labeling whose labels do not exceed a given λ.*

Proof We show a reduction from Systems of q-Distant Representatives. Let $t \geq q^2 + 2p$ and let a family $\mathcal{M} = \{M_i\}_{i=1}^n$ of t-sparse sets be given. We may assume that $\min(\bigcup_{i=1}^n M_i) \geq t$ and we set $\lambda = t + \max(\bigcup_{i=1}^n M_i)$.

We build a tree T with a root v_0 and children of the root $v_i, i = 1, 2, \ldots, n$ (and another level of nodes $v_{i,j,m}$, as defined later). The root will be precolored by 0, vertices v_i will not be precolored, but each of them will have a certain number of children precolored so that the only admissible colors for v_i will be exactly the elements of M_i. This can be attained as follows.

For a vertex v_i, let $M_i = \{a_{i,1} < a_{i,2} < \ldots < a_{i,k_i}\}$. Now consider the interval $[a_{i,j} + 1, a_{i,j+1} - 1]$. Since M_i is t-sparse, $a_{i,j+1} - a_{i,j} \geq q^2 + 2p$ and one can choose numbers $c_{i,j,1} = a_{i,j} + p < c_{i,j,2} < \ldots < c_{i,j,l_j} = a_{i,j+1} - p$ so that $q \leq c_{i,j,m+1} - c_{i,j,m} \leq 2p - 1$ for every $m = 1, 2, \ldots, l_j - 1$. (In more detail, if $a_{i,j+1} - a_{i,j} - 2p = \alpha q + \beta$ then $\alpha \geq q$ and we set $l_j = \alpha + 1$. Then take $c_{i,j,m+1} = c_{i,j,m} + q + 1$ for $m = 1, 2, \ldots, \beta$ and take $c_{i,j,m+1} = c_{i,j,m} + q$

for $m = \beta + 1, \dots, \alpha$. Note that $2p - 1 \geq q + 1 > q$.) Finally for all $m = 1, 2, \dots, l_j$ add leaves $v_{i,j,m}$ precolored by $c_{i,j,m}$ adjacent to the vertex v_i. In every (p, q)-labeling f, it must be $f(v_i) \notin [a_{i,j} + 1, a_{i,j+1} - 1]$, since any integer from this interval differs by at most $p - 1$ from some $c_{i,j,m}$.

We perform this construction for every vertex v_i and for each interval $[a_{i,j} + 1, a_{i,j+1} - 1]$. Because we have assumed that $a_{i,1} \geq t$ and $\lambda - a_{i,k_i} \geq t$, the initial and terminal intervals $[0, a_{i,1} - 1]$ and $[a_{i,k_i} + 1, \lambda]$ are handled in the same way, with dummy boundary values $a_{i,0} = -1$ and $a_{i,k_i+1} = \lambda + 1$.

It follows that T allows a (p, q)-labeling which extends the precoloring (and uses labels from $\{0, 1, \dots, \lambda\}$) if and only if $\mathcal{M}$ has a System of q-Distant Representatives.

Suppose that $f : V(T) \longrightarrow \{0, 1, \dots, \lambda\}$ is a (p, q)-labeling which extends the precoloring. For each i, $f(v_i) \in M_i$, since (by the argument above) for every j, $f(v_i) \notin [a_{i,j} + 1, a_{i,j+1} - 1]$. Since for $i \neq i'$, v_i and $v_{i'}$ have a common neighbor (the root v_0), we have $|f(v_i) - f(v_{i'})| \geq q$ and f restricted to $\{v_1, v_2, \dots, v_n\}$ yields a System of q-Distant Representatives for $\mathcal{M}$.

Showing that T allows a (p, q)-labeling extending the precoloring, provided $\mathcal{M}$ allows a System of q-Distant Representatives, is straightforward. Note here that the precoloring of T constructed from $\mathcal{M}$ itself satisfies the (p, q)-constraints, since for each i the colors used on $v_{i,j,m}$ are at least q apart, they are at least p (and thus at least q away from the label 0 of the root), and they do not interfere with other $v_{i'}$ or $v_{i',j',m'}$ since such vertices are at distance at least three. $\square$

For the (p, q)-labeling problem of trees without precolored vertices, the complexity is still open when $q > 1$. It is tempting to try to prove *NP*-completeness along the lines above. One possibility would be to replace precolored vertices by trees that allow only certain labels at the root. Of course one cannot ask for trees that would allow only one possible label, since the set of admissible labels for a particular vertex is always symmetric with respect to the interval $[0, \lambda]$ (if f is a (p, q)-labeling then so is $f' = \lambda - f$). This does not cause problems for the desired reduction since one can show that Systems of q-Distant Representatives are *NP*-complete even when all sets $M_i \in \mathcal{M}$ are symmetric. This observation leads to the following open problems (affirmative solution to the first one would imply *NP*-completeness of (p, q)-labelings of trees):

Problem 5 Does there exist, for any relatively prime $p > q > 1$, any sufficiently large λ and any $(q^2 + 2p)$-sparse set M, a construction of trees $T_{x,\lambda}$, $x \in [p, \lambda - p]$, such that

(1) the size of $T_{x,\lambda}$ is polynomial in λ,

(2) $T_{x,\lambda}$ allows a (p,q)-labeling in which the root is labeled by x and all labels of children of the root are at distance at least q from the set M,

(3) in any (p,q)-labeling of $T_{x,\lambda}$, the root is labeled either by x or by $\lambda - x$?

The condition (1) above is imposed to guarantee polynomiality of the desired reduction. However, we do not even know of the existence of any $T_{x,\lambda}$ satisfying at least (2-3), and therefore we deem the following problem interesting on its own:

Problem 5 Prove that for any relatively prime $p > q > 1$, any large enough λ and any $x \in [p, \lambda - p]$, there exists a tree $T_{x,\lambda}$ such that in any (p,q)-labeling of $T_{x,\lambda}$ the root is labeled either by x or by $\lambda - x$, and $T_{x,\lambda}$ has such a labeling.

References

[1] Aharoni, R., P. Haxell, Hall's theorem for hypergraphs, *Journal of Graph Theory* **35**, (2000), pp. 83–88.

[2] Battiti, R., A. A. Bertossi, and M. A. Bonuccelli, Assigning codes in wireless networks: Bounds and scaling properties, *Wireless Networks* **5**, 3, (1999), pp. 195–209.

[3] Chang, G. J. and D. Kuo, The $L(2,1)$-labeling problem on graphs, *SIAM J. Disc. Math.* **9**, (1996), pp. 309–316.

[4] Corneil, D. G., H. Lerchs, L. Stewart, Complement reducible graphs, *Discrete Appl. Math.* **3**, (1981), pp. 163–174.

[5] Edmonds, J., Paths, trees and flowers, *Can. J. Math.* **17**, (1965), pp. 449–467.

[6] Farber, M., Hujter, M. and Tuza, Zs., An upper bound on the number of cliques in a graph. *Networks*, **23** (1993) pp. 207-210.

[7] Fiala, J., T. Kloks, and J. Kratochvíl, Fixed-parameter complexity of λ-colorings, In: Graph Theoretic Concepts in Computer Science, WG'99, LNCS 1665, (1999), pp. 350–363.

[8] Fotakis, D., G. Pantziou, G. Pentaris, and P. Spirakis, Frequency assignment in mobile and radio networks, In: Networks in distributed computing. DIMACS workshop, Ser. Discrete Math. Theor. Comput. Sci. 45, (1997), pp. 73–90.

[9] Garey, M.R. and D.S. Johnson, *Computers and Intractability*, W.H. Freeman, 1979.

[10] Georges, J. P. and D. W. Mauro, On the size of graphs labeled with a condition at distance two, *Journal of Graph Theory* **22**, (1996), pp. 47–57.

[11] Griggs, J. R. and R. K. Yeh, Labelling graphs with a condition at distance 2, *SIAM J. Disc. Math.* **5**, (1992), pp. 586–595.

[12] Hall, P., On representatives of subsets, *J. London Math. Soc.* **10**, (1935), pp. 26–30.

[13] van den Heuvel, J., R. A. Leese, and M. A. Shepherd, Graph labeling and radio channel assignment, *Journal of Graph Theory*, **29**, (1998), pp. 263–283.

[14] Hujter, M. and Zs. Tuza, Precoloring extension III. Classes of perfect graphs, *Combin. Probab. Comput.* **5**, (1996), pp. 35–56.

[15] Jonas, K., *Graph colorings analogues with a condition at distance two: $L(2,1)$-labelings and list λ-labelings*, Ph.D. thesis, University of South Carolina, 1993.

[16] Kratochvíl, J., A special planar satisfiability problem and a consequence of its NP- completeness, *Discrete Appl. Math.* **52**, (1994), pp. 233–252.

[17] Kratochvíl, J. and A. Sebő, Coloring precolored perfect graphs, *J. Graph. Th.* **25**, (1997), pp. 207–215.

[18] Liu, D. D.-F. and R. K. Yeh, On distance two labellings of graphs, *ARS Combinatorica* **47**, (1997), pp. 13–22.

[19] Simons, B., A fast algorithm for single processor scheduling, *Proc. 19th Annnual Symp. Foundations of Computer Science*, (1978), pp. 246–252.

[20] Yeh, K.-Ch., *Labeling graphs with a condition at distance two*, Ph.D. Thesis, University of South Carolina, 1990.