PPP-Completeness and Extremal Combinatorics

ENS de Lyon, France

Lukáš Folwarczný ⊠ ©

Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic Institute of Mathematics of the Czech Academy of Sciences, Prague, Czech Republic

Pavel Hubáček ⊠ [®]

Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic

Alon Rosen □ □

Bocconi University, Milano, Italy Reichman University, Herzliya, Israel

Nikolaj I. Schwartzbach ⊠®

Department of Computer Science, Aarhus University, Denmark

Abstract

Many classical theorems in combinatorics establish the emergence of substructures within sufficiently large collections of objects. Well-known examples are Ramsey's theorem on monochromatic subgraphs and the Erdős-Rado sunflower lemma. Implicit versions of the corresponding total search problems are known to be PWPP-hard under randomized reductions in the case of Ramsey's theorem and PWPP-hard in the case of the sunflower lemma; here "implicit" means that the collection is represented by a poly-sized circuit inducing an exponentially large number of objects.

We show that several other well-known theorems from extremal combinatorics – including Erdős-Ko-Rado, Sperner, and Cayley's formula – give rise to *complete* problems for PWPP and PPP. This is in contrast to the Ramsey and Erdős-Rado problems, for which establishing inclusion in PWPP has remained elusive. Besides significantly expanding the set of problems that are complete for PWPP and PPP, our work identifies some key properties of combinatorial proofs of existence that can give rise to completeness for these classes.

Our completeness results rely on efficient encodings for which finding collisions allows extracting the desired substructure. These encodings are made possible by the tightness of the bounds for the problems at hand (tighter than what is known for Ramsey's theorem and the sunflower lemma). Previous techniques for proving bounds in TFNP invariably made use of structured algorithms. Such algorithms are not known to exist for the theorems considered in this work, as their proofs "from the book" are non-constructive.

2012 ACM Subject Classification Theory of computation \rightarrow Problems, reductions and completeness; Theory of computation \rightarrow Complexity classes

Keywords and phrases total search problems, extremal combinatorics, PPP-completeness

Digital Object Identifier 10.4230/LIPIcs.ITCS.2023.22

Related Version Full Version: https://arxiv.org/abs/2209.04827 [3]

Funding Lukáš Folwarczný: Supported by the Grant Agency of the Czech Republic under the grant agreement no. 19-27871X and by the Charles University grant SVV-2020-260578.

Pavel Hubáček: Supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101019547), by the Cariplo CRYPTONOMEX grant, by the Grant Agency of the Czech Republic under the grant agreement no. 19-27871X, and by the Charles University project UNCE/SCI/004.

Alon Rosen: Supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101019547) and Cariplo CRYPTONOMEX grant.

© Romain Bourneuf, Lukáš Folwarczný, Pavel Hubáček, Alon Rosen, and Nikolaj I. Schwartzbach; licensed under Creative Commons License CC-BY 4.0

14th Innovations in Theoretical Computer Science Conference (ITCS 2023).

Editor: Yael Tauman Kalai; Article Ño. 22; pp. 22:1–22:20

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Nikolaj I. Schwartzbach: This project is funded by VILLUM FONDEN under the Villum Kann Rasmussen Annual Award in Science and Technology under grant agreement no 17911.

1 Introduction

A well-known theorem by Ramsey gives a lower bound on the size of the largest monochromatic clique in any edge-coloring of the complete graph using two colors.

Ramsey [25] Any edge-coloring of the complete graph on n vertices with two colors contains a monochromatic clique of size at least $\frac{1}{2} \log n$.

Ramsey's theorem gives rise to a natural computational search problem RAMSEY [18, 16]: given a description of an edge-coloring, output the vertices of a monochromatic clique of size $\frac{1}{2} \log n$. Since the theorem guarantees the existence of a monochromatic clique of this size, RAMSEY belongs to the complexity class TFNP, consisting of efficiently verifiable search problems to which a solution is guaranteed to exist [20].

The computational complexity of RAMSEY very much depends on its representation. On the one hand, it is efficiently solvable when the graph is given *explicitly*; a folklore proof of Ramsey's theorem gives an efficient algorithm to find such a subgraph. On the other hand, the situation is less clear when the graph is represented *implicitly*, e.g., via a small Boolean circuit that, for any pair of vertices, outputs the corresponding color of the edge-coloring of the graph.¹

Another TFNP problem considered in the literature that is motivated by a result in extremal combinatorics arises from the well-known Erdős-Rado sunflower lemma.

Erdős-Rado [10] Any family of n-sets of cardinality greater than $n^n n!$ contains an n-sunflower of size n+1, i.e., subsets $A_1, A_2, \ldots, A_{n+1} \in \mathcal{F}$ such that, for some Δ , $A_i \cap A_j = \Delta$ for every distinct A_i, A_j .

Similarly to RAMSEY, an instance of the total search problem SUNFLOWER [16] can be represented implicitly, e.g., by a Boolean circuit which outputs a characteristic vector of a set in the family when given the index of this set.

In general, little is known of the complexity of the implicit variants of RAMSEY or SUNFLOWER – the proofs of the corresponding theorems are either non-constructive or result in inefficient (i.e., superpolynomial-time) algorithms. RAMSEY is known to be PWPP-hard under randomized reductions, as shown by Krajíček [18] and Komargodski, Naor, and Yogev [16], and SUNFLOWER is known to be PWPP-hard, as shown by Komargodski, Naor, and Yogev [16]. This means that finding the desired substructure is at least as hard as finding collisions in an arbitrary poly-sized shrinking circuit and, hence, hard in the worst-case if collision-resistant hash functions exist. However, they are not known to be complete for the class PWPP and the intriguing question of whether they give rise to a complexity class distinct from PWPP has remained open for years.

1.1 Our Results

We explore new connections between classical theorems in extremal combinatorics and the complexity classes PPP [23] and PWPP [14], i.e., the classes of search problems with totality guaranteed by the (weak) pigeonhole principle. We show that PPP and PWPP can be characterized via a number of new TFNP problems based on the following theorems.

Note that, given such a representation, it might be even hard to compute the degree of a node with respect to one of the two colors.

Erdős-Ko-Rado [9]. Any family of distinct pairwise-intersecting k-sets on a universe of size m has size at most $\binom{m-1}{k-1}$.

Sperner [27]. The largest antichain, i.e., a family of subsets such that no member is contained in any other, on a universe with 2n elements is unique and consists of all subsets of size n.

Cayley [4]. There are exactly n^{n-2} spanning trees of the complete graph on n vertices.

Just as for Ramsey and Sunflower, the corresponding search problems are efficiently solvable when given explicit access to the family of objects and, again, their computational complexity is open when we consider implicit access to the structure, e.g., where the instance is given by a circuit that on input i returns an encoding of the ith object in the collection.² The totality of the problems we define follows from a common principle – the instances are given via an implicit representation of a sufficiently large collection of objects (e.g., subsets for Erdős-Ko-Rado) such that, by the corresponding theorem, there exists a small subset of these objects satisfying some efficiently verifiable property (e.g, a pair of disjoint subsets for Erdős-Ko-Rado).

In addition to the above completeness results, we define TFNP problems arising from the following results in extremal combinatorics.

Mantel [19]. Any triangle-free graph on n vertices has at most $n^2/4$ edges.

Turán [28]. If G=(V,E) is a graph on n=|V| vertices that does not contain any r+1-clique, then $|E|\leq (1-\frac{1}{r})\frac{n^2}{2}$.

Ward-Szabó [29]. Any edge-coloring of the complete graph on n vertices with $2 \le r \le \sqrt{n}$ colors must contain a bichromatic triangle.

In the case of the Ward-Szabó theorem, we define three variants of the corresponding computational problem. We prove that all three variants are PWPP-hard, the first one is in PWPP and the second in in PPP. Proving the inclusion for the third variant remains open and it joins RAMSEY and SUNFLOWER as a candidate problem that might define a new class above PWPP. Turán's theorem is a generalization of Mantel's theorem. We define a weak version of a problem based on Turán's theorem for every r, and prove that these problems form a hierarchy between PWPP and PPP. Furthermore, we define strong versions which, by a reduction from [24], form a hierarchy above PPP. It is open whether these hierarchies colllapse. An overview of our results in terms of weak and strong problems (see Section 1.3) is given in Table 1.

1.2 Techniques and Ideas

A long-standing open problem regarding RAMSEY and SUNFLOWER has been to determine their status with respect to the classes PWPP and PPP. For the most part, the most challenging part in establishing completeness for some syntactic subclass of TFNP lies in proving hardness (see, e.g., [7, 21, 11]). For subclasses of TFNP such as PPAD, PPA, and PLS,

² Note that an implicit representation of the collection might not necessarily satisfy the assumptions of the underlying theorem. For instance, representing sets via characteristic vectors for Erdős-Ko-Rado does not ensure that they are actually k-sets or that they are distinct. Importantly, such a violation could allow evading the totality of the search problem. Nevertheless, we can ensure totality by allowing locally verifiable evidence of a malformed representation as a solution, e.g., an index not corresponding to a k-set or two indices corresponding to the same set.

Table 1 Summary of the complexity of problems we consider. Except for RAMSEY and SUNFLOWER, all problems were introduced in this work. The containment results for WEAK-GENERAL-ERDŐS-KO-RADO₂ and GENERAL-ERDŐS-KO-RADO₂ rely on the efficient Baranyai assumption (Assumption 4.12). The PWPP-hardness of RAMSEY is known only under randomized reductions.

Problem	Hardness	Containment
RAMSEY	PWPP [18, 16]	TFNP
Sunflower		
Ward-Szabó	PWPP	
WEAK-MANTEL		PPP
Weak-Turán $_r$	[Theorems 7.5, 8.4, and 8.14]	
Ward-Szabó-Colorful-Collisions		[Theorems 7.8, 8.5, and 8.15]
Ward-Szabó-Collisions		
weak-Erdős-Ko-Rado		
weak-general-Erdős-Ko-Rado $_k$	PWPP [Theorems 4.5, 4.14, 5.3, 6.3, 7.5, and 7.6]	
WEAK-SPERNER-ANTICHAIN		
WEAK-CAYLEY		
Erdős-Ko-Rado	PPP [Theorems 4.7, 4.16, 5.5, and 6.5]	
general-Erdős-Ko-Rado $_k$		
Sperner-Antichain		
CAYLEY		
Mantel	PPP [Theorem 8.9]	TFNP

the inclusion in a subclass mostly follows from the existence of an inefficient yet structured algorithm for the problem at hand; for example, the chessplayer algorithm for PPA [23] or the steepest descent algorithm for PLS [15]. However, this methodology seems inapplicable for proving inclusion in PWPP or PPP as these classes do not exhibit any characterizing graph-theoretic structure that could capture some class of natural algorithms.

In contrast to many existing bounds in TFNP, our work does not make use of structured algorithms but instead makes use of encodings that translate between substructures and collisions in circuits. In order to establish inclusion in PWPP, we encode the objects of the collection using a "property-preserving encoding" that encodes the objects in a way that translates some specific relation into collisions. More precisely, we want an encoding function that is efficiently computable and (nearly) optimal, such that whenever two elements have the same encoding, these two elements give a solution to the original problem. While this technique is quite general, it is not always clear how to instantiate the encoding to get the desired collisions.

Consider, for example, the total search problem corresponding to the Erdős-Ko-Rado theorem for intersecting families of n-sets on a universe of size 2n. An instance can be given by a Boolean circuit $C: \{0,1\}^{\lceil \log \binom{2n-1}{n-1} \rceil \rceil + 1} \to \{0,1\}^{2n}$ representing a family of subsets of [2n], i.e., C(i) is the characteristic vector of the i-th n-set in the family. Suppose the outputs of C define distinct n-sets. Since there are more than $\binom{2n-1}{n-1}$ of them, then, by the Erdős-Ko-Rado theorem, there must exist a pair of inputs mapped to disjoint n-sets by C. We define any such pair of inputs to be a solution. To ensure totality of the corresponding search problem, circuits that do not represent distinct n-sets give rise to additional solutions of the form (a) an i such that C(i) is not an n-set, or (b) $i \neq j$ such that C(i) = C(j).

When proving that the above total search problem is contained in the complexity class PWPP, at a high level, we want to encode the *n*-sets of the family using a shrinking circuit, in such a way that collisions correspond to disjoint sets. Observe that for *n*-sets in a universe

of size 2n, the only disjoint sets are complements and, hence, we get an equivalent instance of the problem if we map each set to either itself or its complement, arbitrarily. In our construction, we map each set S to the representative not containing the element 1. That is, if $1 \notin S$, the set is left unchanged and, otherwise, it is mapped to its complement \overline{S} . Note that, by the pigeonhole principle, two sets that do not contain 1 must have a non-empty intersection since we work with n-subsets of [2n]. To obtain a shrinking circuit, we make use of Cover encodings (Section 3.1) that give an optimal encoding of all n-sets by considering their lexicographic order. Notice that if the input S is not an n-set, we may map it arbitrarily to any n-set, as a collision, in this case, yields a solution to the instance of the above problem motivated by the Erdős-Ko-Rado theorem.

In contrast, the PWPP-hardness results for RAMSEY and SUNFLOWER follow an elegant and rather direct (compared to other hardness results for subclasses of TFNP) technique of the graph-hash product [18, 16], which we illustrate on RAMSEY. Recall that there are known randomized constructions of edge-colorings of the complete graph $K_{2^{n/4}}$ on $2^{n/4}$ vertices that do not contain a monochromatic clique of size n/2 [8]. Given such an underlying edge-coloring of $K_{2^{n/4}}$ and a hash function h mapping n-bit strings to n/4-bit strings, one can construct an edge-coloring of the complete graph on 2^n vertices by assigning to every edge $(u,v) \in \{0,1\}^n \times \{0,1\}^n$ the color of the edge $(h(u),h(v)) \in \{0,1\}^{n/4} \times \{0,1\}^{n/4}$ from the underlying coloring. Since the underlying edge-coloring of $K_{2^{n/4}}$ does not contain a monochromatic clique of size n/2, it is easy to see that any monochromatic clique of size n/2 in the resulting edge-coloring of K_{2^n} (guaranteed to exist by Ramsey's theorem) must have been introduced via a collision in the hash h.

As noted by [16], the structure of a PWPP-hardness proof using the graph-hash product is not restricted to total search problems corresponding to graph-theoretic theorems of existence; indeed, [16] used the graph-hash product to prove also PWPP-hardness of SUNFLOWER. On a high level, for a problem to be amenable to the graph-hash product technique, it is sufficient to be able to construct a collection of objects such that 1) it does not contain the desired substructure, 2) the logarithm of its size is at least a constant fraction of the logarithm of the threshold necessary for the existential theorem to apply³ and 3) it can be efficiently indexed. Then, we can interpret the output of an appropriately shrinking hash h as an index into the small collection of objects, and, for each index, we can efficiently compute and output the corresponding element in the collection. Again, since the small collection does not contain the desired substructure, all solutions of the instance constructed via graph-hash product must in some way result from a collision in the hash h.

For example, consider the total search problem arising from Sperner's theorem on antichains – here, the threshold size is $\binom{2n}{n}$, meaning that if we have a family with strictly more than $\binom{2n}{n}$ distinct subsets of [2n] then one subset from the family must be contained in another member of the family. It is straightforward to construct a family of subsets that does not contain the specific substructure (i.e., a subset that is included in another one) with size equal to the threshold size $\binom{2n}{n}$. It suffices to consider the family of all the n-subsets of [2n]. Similarly, for many other combinatorial problems we study, an adequate collection of objects can be found by looking at a collection of maximum size that does not contain the substructure.

We also show natural reductions between some of the problems we define (for instance, from Erdős-Ko-Rado to Sperner-Antichain), which, in our opinion, highlights the relevance of these new problems and the fact that their definition is the correct one.

This is a technical condition ensuring that we can reduce from a PWPP-complete variant of the problem of finding collisions in a shrinking hash. Note that it is easy to find collisions in functions that exhibit extreme shrinking.

1.3 **PPP-Completeness From Extremal Combinatorics**

Up to this point, our discussion did not explicitly distinguish between the classes PWPP and PPP. However, our work highlights important structural differences between the two complexity classes. Recall that the class PWPP contains the search problems in TFNP whose totality can be proved using the weak pigeonhole principle: "In any assignment of 2n pigeons to n holes there must be two pigeons sharing the same hole."

This statement can be seen as a result in extremal combinatorics bounding the maximum number of pigeons that can be assigned to n holes without two pigeons being sent to the same hole. More generally, we say that a theorem from extremal combinatorics is "weak" if it gives an upper bound (which may or may not be tight) on the maximum size of a collection of objects that does not contain some substructure (above, two pigeons sharing the same hole). On the contrary, we say that a theorem from extremal combinatorics is "strong" if it gives a tight upper bound on the maximum size of a collection of objects that does not contain some substructure, as well as some structural property about the maximum families without the substructure. For instance, the strong pigeonhole principle can be stated as: "In any assignment of n pigeons to n holes there is either a pigeon in the first hole or two pigeons sharing the same hole." Note that it is exactly this formulation of the strong pigeonhole principle that defines the class PPP.

Many results in extremal combinatorics have both a weak statement and a strong statement. For such results, we can define a problem corresponding to the weak statement, which often is related to PWPP, and a problem corresponding to the strong statement, which often is related to PPP. In this paper, all PWPP-hard problems correspond to a weak theorem in extremal combinatorics, while PPP-hard problems correspond to a strong theorems in extremal combinatorics. As an example, consider Cayley's formula and note that the bound n^{n-2} is tight. Hence, if we are given a collection of exactly n^{n-2} distinct graphs on n vertices, then either one of the graphs is not a spanning tree, or every spanning tree is in the collection. This observation induces a TFNP problem that we show to be PPP-complete.

1.4 Related Work

In independent and concurrent work, Pasarkar, Yannakakis, and Papadimitriou [24] explored the connections between extremal combinatorics and the class PPP. They defined a new subclass of TFNP called PLC inspired by the proof of Ramsey's theorem and proved that this class contains RAMSEY, a version of SUNFLOWER and the whole PPP. Compared to our work, they also considered the Erdős-Ko-Rado theorem, yet, in the setup where the universe has size 2^n and the subsets of this universe have size 2; they proved that this variant is also PPP-complete. Importantly, our setup does not allow subsets of constant size. Pasarkar et al. also considered problems based on Turán's theorem and problems called BAD-k-COLORING and show that these problems form a hierarchy above PPP. Their problems based on Turán's theorem are equivalent to our $TURÁN_r$ and hence their reduction implies $\text{TURÁN}_r \leq \text{TURÁN}_{r+1}$.

Another subclass of TFNP, based on approximate counting, containing RAMSEY was defined by Kołodziejczyk and Thapen [17].

Compared to the majority of subclasses of TFNP that have been extensively studied and are known to capture various total search problems from diverse domains of mathematics, PPP and PWPP might seem less expressive and the first non-trivial completeness results appeared only recently. Sotiraki, Zampetakis, and Zirdelis [26] and Ban, Jain, Papadimitriou, Psomas, and Rubinstein [1] demonstrated that PPP contains computational problems from

number theory and the theory of integral lattices. In particular, Sotiraki et al. showed PPP-completeness of a computational problem related to Blitchfeld's theorem and PPP-completeness (resp. PWPP-completeness) of a problem motivated by the Short Integer Solution problem. Hubáček and Václavek [13] showed that some general formalizations of the discrete logarithm problem are complete for PWPP and PPP and, motivated by classical constructions of collision-resistant hashing, they characterized PWPP via the problem of breaking claw-free (pseudo-)permutations.

1.5 Open Problems

Our work suggests various interesting directions for future research:

- We exploit the power of strong statements in extremal combinatorics for establishing PPP-completeness. The notorious lack of tight bounds for the Erdős-Rado sunflower lemma and Ramsey's theorem implies that we have no strong version of these theorems, which may explain why showing the inclusion of the corresponding problems in, e.g., PPP has eluded researchers.
- We introduced total search problems corresponding to Mantel's theorem, Turán's theorem, and Ward-Szabó's theorem. In this work, we only prove hardness results for these problems but no inclusion results. Hence, it is still open whether they are complete for the classes PPP and PWPP, or whether they could define a new subclass of TFNP.
- Another exciting question is whether the efficient Baranyai assumption (Assumption 4.12) holds, as well as whether it is possible to prove the inclusion results of the problems associated to the general version of Erdős-Ko-Rado's theorem without that assumption. Showing reductions between GENERAL-ERDŐS-KO-RADO_k and GENERAL-ERDŐS-KO-RADO_l for $k \neq l$ without the efficient Baranyai assumption would also be interesting.
- The Turá N_r problem is defined in a similar fashion to Mantel and it holds Turá $N_r \le \text{Turá}N_{r+1}$. It is then open whether these problems are of the same complexity or whether they form a hierarchy above PPP.
- Finally, we believe that the problems between PWPP and PPP deserve a more thorough investigation to further our understanding of the classes. In particular, WEAK-TURÁN_r and GENERAL-PIGEON_{k(n)} (see Section 2.1) are two potential hierarchies between these two classes.

2 Preliminaries

If Ω is a set of size n, we associate the set 2^{Ω} with the characteristic vectors from $\{0,1\}^n$ for some arbitrary (but fixed) order on Ω . We denote by \subseteq the partial order on $\{0,1\}^n$ where $x \subseteq y$ iff $x_i \leq y_i$ for every $i = 1 \dots n$. If $x \in \{0,1\}^n$ is a string, we denote by $\overline{x} := \overline{x_1} \overline{x_2} \cdots \overline{x_n}$ the complement of x, defined by $\overline{x}_i = 1 - x_i$. We also use other set-theoretic operators \cap, \cup, \setminus that are defined in a natural way. We also denote by $|x| = \sum_{i=1}^n x_i$ the number of 1s in x when the length is implicit from the context.

2.1 Total Search Problems

A search problem is defined by a binary relation $R \subseteq \{0,1\}^* \times \{0,1\}^* - \text{a string } s \in \{0,1\}^*$ is a *solution* for an *instance* $x \in \{0,1\}^*$ if $(x,s) \in R$. A search problem defined by relation R is *total* if for every x, there exists an s such that $(x,s) \in R$. We define TFNP as the class of all total search problems that can be efficiently verified, i.e., there is a deterministic polynomial-time Turing machine that, given (x,s), outputs 1 if and only if $(x,s) \in R$ and, for every instance x, there exists a solution s of polynomial length in the size of x.

To avoid unnecessarily cumbersome phrasing throughout the paper, we define TFNP relations implicitly by presenting the set of valid *instances* $X \subseteq \{0,1\}^*$ recognizable in polynomial time (in the length of an instance) and, for each instance $i \in X$, the set of admissible *solutions* $Y_i \subseteq \{0,1\}^*$ for the instance i. It is then implicitly assumed that, for any invalid instance $i \in \{0,1\}^* \setminus X$, we define the corresponding solution set as $Y_i = \{0,1\}^*$.

We say that $A \in \mathsf{TFNP}$ reduces to $B \in \mathsf{TFNP}$ (written also as $A \leq B$) if there exist two poly-time computable functions f and g satisfying

```
\forall x, y \in \{0, 1\}^* \quad (f(x), y) \in B \Rightarrow (x, g(x, y)) \in A.
```

Next, we recall the definitions of the complexity classes PWPP and PPP via their canonical complete problems WEAK-PIGEON and PIGEON.

▶ Definition 2.1 (WEAK-PIGEON and PWPP [14]). The problem WEAK-PIGEON is defined by the relation

```
Instance: A Boolean circuit C: \{0,1\}^n \to \{0,1\}^{n-1}.
```

Solution: $x_1 \neq x_2 \text{ s.t. } C(x_1) = C(x_2).$

The class of all TFNP problems reducible to WEAK-PIGEON is called PWPP.

▶ **Definition 2.2** (PIGEON and PPP [23]). The problem PIGEON is defined by the relation **Instance**: A Boolean circuit $C: \{0,1\}^n \to \{0,1\}^n$.

Solution: *One of the following:*

- (i) $x \ s.t. \ C(x) = 0^n$,
- (ii) $x \neq y \text{ s.t. } C(x) = C(y).$

The class of all TFNP problems reducible to PIGEON is called PPP.

On several occasions, we need to change the domain size of the WEAK-PIGEON circuit. The following lemma was proved by Krajíček [18], and later by Jeřábek using the well-known Merkle-Damgård construction [22, 6].

▶ Lemma 2.3 ([14]). Let p be a polyonimal satisfying p(n) > n for every n. If we modify the definition of the problem WEAK-PIGEON by having a circuit $C: \{0,1\}^{p(n)} \to \{0,1\}^n$, the new problem is equivalent to the original WEAK-PIGEON.

More generally, we can define the problem $\operatorname{General-Pigeon}_{k(n)}$ as follows.

▶ **Definition 2.4** (GENERAL-PIGEON_{k(n)}). The problem GENERAL-PIGEON_{k(n)} is defined by the relation

Instance: A Boolean circuit $C: \{0,1\}^n \to \{0,1\}^n$.

- (i) $x \neq y \in \{0,1\}^n$ s.t. C(x) = C(y),
- (ii) $x \in \{0,1\}^n$ s.t. C(x) is one of the first k(n) elements of $\{0,1\}^n$.

This problem is mentioned by Goldberg and Papadimitriou [12] and they denote the class of all problems reducible to General-Pigeon_{k(n)} by $\mathsf{PPP}_{k(n)}$. Note that this problem gets harder as the growth-rate of k(n) decreases. It is trivial for $k(n) = 2^n$, equivalent to Weak-Pigeon for $k(n) = 2^{n-1}$ and to Pigeon for k(n) = 1.

This problem induces an entire family of problems that interpolates between WEAK-PIGEON and PIGEON. It is not clear how many distinct problems appear in the hierarchy. It is also unclear whether each PWPP-hard problem that is in PPP is in fact equivalent to one of these. This is relevant to us in Section 7, where we reduce to GENERAL-PIGEON_{2ⁿ⁻¹-2^{n/2-1}}, and in Section 8, where we show a hierarchy of problems WEAK-TURÁN_r between PWPP and PPP.

3 Property-Preserving Encodings

A key ingredient to our proofs of inclusion in PWPP and PPP is the use of efficient encodings. We rely on two different types of encodings. The first one simply consists of bijections between two different representations of the same set of objects, the first one being more natural and more convenient to work with, and the second one being more concise. The second type of encodings, which we call property-preserving encodings, consists of shrinking functions, in the sense that the range of the encoding is smaller than the domain, whose collisions correspond to elements sharing some property of interest. The following definition gives a precise description of the features we require from these encodings.

- ▶ **Definition 3.1** (Property-preserving encoding). Let $\mathcal{X} \subseteq \{0,1\}^k$, \mathcal{Y} be sets, and let \sim be an equivalence relation on \mathcal{X} . Let $E: \{0,1\}^k \to \mathcal{Y}$ be a surjection. We say that E constitutes a property-preserving encoding for \sim on \mathcal{X} if it satisfies.
- (Efficiency). E can be computed in polynomial time.
- \blacksquare (Compression). $|\mathcal{Y}| \leq |\mathcal{X}|$.
- (\sim -correctness). $\forall x, x' \in \mathcal{X}, [E(x) = E(x') \Rightarrow x \sim x'].$

Note that bijections are a form of property-preserving encoding, where the equivalence relation we want to preserve is equality. We describe several property-preserving encodings, including some bijective encodings.

3.1 Cover Encodings

Our reductions in Section 4 make use of *Cover encodings* [5] that efficiently encode subsets of a specified size in optimal space: namely, we may encode every subset $S \subseteq \{0,1\}^m$ such that |S| = k by considering the lexicographic order of all $\binom{m}{k}$ such sets (in fact we consider the lexicographic order over their characteristic vectors $\in \{0,1\}^m$), and mapping this into binary strings: this requires $\lceil \log \binom{m}{k} \rceil$ bits which is optimal. We denote the encoding and decoding functions as follows, with $\alpha(k,m) = \lceil \log \binom{m}{k} \rceil$.

$$\begin{split} E_{\mathsf{Cover}}^{k,m} \colon \{0,1\}^m &\to \{0,1\}^{\alpha(k,m)} \\ D_{\mathsf{Cover}}^{k,m} \colon \{0,1\}^{\alpha(k,m)} &\to \{0,1\}^m \end{split}$$

We set $E_{\mathsf{Cover}} = E_{\mathsf{Cover}}^{n,2n}$ and $D_{\mathsf{Cover}} = D_{\mathsf{Cover}}^{n,2n}$, and $\alpha = \alpha(n,2n)$. As described in [5], these functions can be made efficient.

▶ **Lemma 3.2.** For every $k \leq m$, $D_{\mathsf{Cover}}^{k,m} \circ E_{\mathsf{Cover}}^{k,m}$ is the identity over all k-subsets of $\{0,1\}^m$. Similarly, $E_{\mathsf{Cover}}^{k,m} \circ D_{\mathsf{Cover}}^{k,m}$ is the identity over the first $\binom{k}{m}$ elements in the lexicographic order of $\{0,1\}^{\alpha(k,m)}$.

Note that the behavior of $D_{\mathsf{Cover}}^{k,m}$ is undefined for the last $2^{\alpha(k,m)} - \binom{m}{k}$ inputs. Furthermore, by design, $E_{\mathsf{Cover}}^{k,m}$ is well-defined on any subset of [m] (even if this subset does not have size k), but the encoding only makes sense for subsets of size k. We also note the following identity which will be useful later when dealing with n-subsets of [2n].

$$D_{\mathsf{Cover}}(0^{\alpha}) = 0^{n} 1^{n} = \overline{[n]} \tag{1}$$

- ▶ Remark 3.3. Consider the case of encoding n-subsets of [2n]. Since we encode sets according to the rank of their characteristic vector in the lexicographic order, any set that does not contain element 1 is one of the $\binom{2n-1}{n-1} = \frac{1}{2} \binom{2n}{n} \le 2^{\alpha-1}$ first ones in the lexicographic order, hence its encoding starts with a 0. Conversely, if we decode an element whose first two bits are 0's, this means that the corresponding n-subset of [2n] is one of the first $2^{\alpha-2} \le \binom{2n-1}{n-1}$ in the lexicographic order, hence that it does not contain the element 1.
- ▶ Remark 3.4. Let T_1 be the tree composed of the edges $(1,2),(1,3),\ldots,(1,n)$. Then, $\tilde{E}_{\mathsf{Prüfer}}(T_1) = 0^\beta$ and $\tilde{D}_{\mathsf{Prüfer}}(0^\beta) = T_1$.

4 Erdős-Ko-Rado Theorem on Intersecting Families

In this section, we define total search problems motivated by the well-known Erdős-Ko-Rado theorem on intersecting families and study their computational complexity. First, we present a PWPP-complete variant of the problem. Next, we modify the problem using a strong statement of the Erdős-Ko-Rado theorem to get a PPP-complete variant. Recall the definition of an intersecting family and the statement of the Erdős-Ko-Rado theorem.

- ▶ **Definition 4.1** (Intersecting family). Let Ω be any set. A family of sets $\mathcal{F} \subseteq 2^{\Omega}$ is an intersecting family if no two sets are disjoint, i.e., if for any $A, B \in \mathcal{F}$, it holds that $A \cap B \neq \emptyset$.
- ▶ Classical Theorem 4.2 (Erdős-Ko-Rado [9]). Assume $m \ge 2k$. Any intersecting family where each set has k elements on a universe of size m contains at most $\binom{m-1}{k-1}$ sets, and this bound is tight.

We start by defining a total search problem motivated by a special case of the Erdős-Ko-Rado theorem for families of n-sets in a universe of size 2n presented in the following corollary.

▶ Corollary 4.3. Any intersecting family where each set has n elements on a universe of size 2n contains at most $\binom{2n-1}{n-1}$ sets, and this bound is tight. Furthermore, if \mathcal{F} is an intersecting family of maximum size, then for every n-subset S, exactly one of S and \overline{S} is in \mathcal{F} .

Suppose that we have a collection containing more than $\binom{2n-1}{n-1}$ sets of size n on 2n elements. Then, by Classical Theorem 4.2, there must be two sets that do not intersect. This induces a total search problem of finding two such disjoint sets. We consider an implicit representation of such a collection by a circuit C whose inputs serve as indices in the collection. The output of the circuit is a representation of the corresponding set as a characteristic vector of the 2n elements. Of course, this representation does not guarantee that C satisfies the conditions required for Classical Theorem 4.2 to apply, which would make the problem not total; in this case, we allow evidence of this fact to be a solution to the problem. Namely, if for a given input x, we do not have |C(x)| = n, or two distinct indices x, y represent the same set, i.e., C(x) = C(y), we allow such inputs as solutions.

▶ Definition 4.4 (WEAK-ERDŐS-KO-RADO). The problem WEAK-ERDŐS-KO-RADO is defined by the relation

Instance: A Boolean circuit $C: \{0,1\}^{\left\lceil \log\left(\binom{2n-1}{n-1}\right)\right\rceil + 1} \to \{0,1\}^{2n}$.

Solution: *One of the following:*

- (i) $x \ s.t. \ |C(x)| \neq n$,
- (ii) $x \neq y \ s.t. \ C(x) = C(y),$
- (iii) $x, y \ s.t. \ C(x) \cap C(y) = \emptyset.$

As we discussed in the introduction, the totality of this problem is proved using a "weak" statement in extremal combinatorics, namely the first part of Corollary 4.3, hence the name Weak. However, the analogy with Weak-Pigeon goes further. Indeed, our first main theorem is the following.

► Theorem 4.5. WEAK-ERDŐS-KO-RADO is PWPP-complete.

Due to space limitations, we present all the proofs from this and following sections and some additional discussions in the full version of our work [3].

PPP-completeness using the tight bound

We remark that Corollary 4.3 gives a tight upper bound on the size of the collection. Furthermore, we know some structure of any collection whose size is exactly one $\binom{2n-1}{n-1}$: it must either not be an intersecting family, or it must contain either [n] or [n]. This is an example of a "strong" theorem in extremal combinatorics. As discussed in the introduction, this observation allows us to modify the problem to be create a variant of WEAK-ERDŐS-KO-RADO that is to WEAK-ERDŐS-KO-RADO what PIGEON is to WEAK-PIGEON. The idea is to let Cencode a collection whose size exactly matches the threshold. We then let C represent a collection of exactly $\binom{2n-1}{n-1}$ sets, and also allow preimages of [n] and $\overline{[n]}$ as solutions. We show that modifying the problem in this manner makes it PPP-complete, thus strengthening the analogy with PIGEON. This technique is quite general, and we utilise it again in later sections.

▶ Definition 4.6 (Erdős-Ko-Rado). The problem Erdős-Ko-Rado is defined by the

Instance: A Boolean circuit $C: \{0,1\}^{\lceil \log \binom{2n-1}{n-1} \rceil} \to \{0,1\}^{2n}$.

Solution: *One of the following:*

- (i) $x \text{ s.t. } |C(x)| \neq n \text{ and } x < \binom{2n-1}{n-1},$

- $\begin{array}{l} \text{(ii)} \ \ x \neq y \ \ s.t. \ \ C(x) = C(y) \ \ and \ \ x,y < \binom{2n-1}{n-1}, \\ \text{(iii)} \ \ x,y \ \ s.t. \ \ C(x) \cap C(y) = \emptyset \ \ and \ \ x,y < \binom{2n-1}{n-1}, \\ \text{(iv)} \ \ x \ \ s.t. \ \ C(x) = [n] \ \ or \ \ C(x) = [\overline{n}] \ \ and \ \ x < \binom{2n-1}{n-1}. \end{array}$
- ▶ Theorem 4.7. Erdős-Ko-Rado is PPP-complete.

A Generalized Erdős-Ko-Rado Problem 4.1

For the previous problems, we were only considering a very restricted version of the Erdős-Ko-Rado theorem, namely for an intersecting family of n-subsets of [2n]. We now consider a more general version where we consider an intersecting family of n-subsets of [kn] for some k > 2.

We now fix some k > 2 for the rest of this section. The Erdős-Ko-Rado theorem states that if \mathcal{F} is an intersecting family where each set has n elements on a universe of size kn, then \mathcal{F} contains at most $\binom{kn-1}{n-1}$ sets. Then, we can define the following TFNP problem, very similar to WEAK-ERDŐS-KO-RADO.

 \blacktriangleright Definition 4.8 (Weak-General-Erdős-Ko-Rado_k). The problem

Weak-general-Erdős-Ko-Rado $_k$ is defined by the relation

Instance: A Boolean circuit $C: \{0,1\}^{\lceil \log \left(\binom{kn-1}{n-1}\right) \rceil + 1} \to \{0,1\}^{kn}$.

Solution: *One of the following:*

- (i) $x \ s.t. \ |C(x)| \neq n,$
- (ii) $x \neq y \ s.t. \ C(x) = C(y),$
- (iii) x, y s.t. $C(x) \cap C(y) = \emptyset$.

To prove that WEAK-GENERAL-ERDŐS-KO-RADO $_k \in \mathsf{PWPP}$, we present some useful definitions and results related to the Erdős-Ko-Rado theorem.

- ▶ **Definition 4.9.** If k divides m, a (k, m)-parallel class is a set of m/k k-subsets of [m] which partition [m].
- ▶ Classical Theorem 4.10 (Baranyai, [2]). If k divides m, we can define $\binom{m-1}{k-1}$ (k,m)-parallel classes $A_1, \ldots, A_{\binom{m-1}{k-1}}$ such that each k-subset of [m] appears in exactly one A_i .
- ▶ Remark 4.11. Note that Baranyai's theorem proves the Erdős-Ko-Rado theorem in the case where the size of the subsets divides the size of the universe. Note also that, up to renaming of the elements, we can assume that A_1 consists exactly of the sets $\{1, 2, ..., n\}, \{n+1, n+2, ..., 2n\}, ...,$ and $\{(k-1)n+1, (k-1)n+2, ..., kn\}$.

However, all known proofs of this theorem are inefficient, in the sense that there is no known way to define $\mathcal{A}_1, \ldots, \mathcal{A}_{\binom{m-1}{k-1}}$ such that given a k-subset of [m], we can find in polynomial time the only i such that this subset appears in \mathcal{A}_i . We make this assumption explicit.

- ▶ Assumption 4.12 (efficient Baranyai assumption). There is an efficient procedure to define $A_1, \ldots, A_{\binom{m-1}{k-1}}$ and a circuit $Bar: \{0,1\}^m \to [\binom{m-1}{k-1}]$ which takes as input a k-subset of [m] and returns the only index i such that this subset appears in A_i . Furthermore, we assume that A_1 consists exactly of the sets $\{1,2,\ldots,n\}, \{n+1,n+2,\ldots,2n\},\ldots$, and $\{(k-1)n+1,(k-1)n+2,\ldots,kn\}$.
- ▶ Remark 4.13. Let \mathcal{X} be the set of n-subsets of [kn]. We define an equivalence relation \sim on \mathcal{X} by saying that two n-subsets X and Y of [kn] are equivalent if and only Bar(X) = Bar(Y), meaning that they are in the same (n,kn)-parallel class in the partition induced by Bar. Then, we have that Bar is a property-preserving encoding for \sim on \mathcal{X} . Note that two equivalent subsets are either equal or disjoint. Hence, the property that is preserved by Bar is such that if two of its inputs collide, they form a solution to our problem. Then, to prove the inclusion of WEAK-GENERAL-ERDŐS-KO-RADO $_k$ into PPP, it suffices to compose our instance of WEAK-GENERAL-ERDŐS-KO-RADO $_k$ with Bar.

The previous two lemmas establish the following result.

▶ Theorem 4.14. Under Assumption 4.12, WEAK-GENERAL- $ERD\H{O}$ S-KO- $RADO_k$ is PWPP-complete.

PPP-completeness using the tight bound

Like for the case of n-subsets of [2n], we can define a "tight" version of the previous problem, which is very similar to Erdős-Ko-Rado.

▶ **Definition 4.15** (GENERAL-ERDŐS-KO-RADO $_k$).

The problem GENERAL-ERDŐS-KO-RADO_k is defined by the relation

Instance: A Boolean circuit $C: \{0,1\}^{\lceil \log(\binom{kn-1}{n-1}) \rceil} \to \{0,1\}^{kn}$.

Solution: *One of the following:*

- (i) $x \ s.t. \ |C(x)| \neq n \ and \ x < \binom{kn-1}{n-1},$
- (ii) $x \neq y$ s.t. C(x) = C(y) and $x, y < \binom{kn-1}{n-1}$,
- (iii) x, y s.t. $C(x) \cap C(y) = \emptyset$ and $x, y < \binom{kn-1}{n-1}$,
- (iv) x s.t. $C(x) = \{1, 2, ..., n\}$ or $\{n+1, n+2, ..., 2n\}$, or..., or $\{(k-1)n+1, (k-1)n+2, ..., kn\}$ and $x < \binom{kn-1}{n-1}$.

First, let us see why this problem is total. Suppose that we have a list of $\binom{kn-1}{n-1}$ subsets of [kn]. If one of the sets does not have n elements, if two of the sets are equal, or if two of the sets do not intersect, we have a solution. Now, suppose that we have an intersecting family of $\binom{kn-1}{n-1}$ distinct n-subsets of [kn] and consider a collection of (n,kn)-parallel classes $\mathcal{A}_1,\ldots,\mathcal{A}_{\binom{kn-1}{n-1}}$ such that each n-subset of [kn] appears in exactly one \mathcal{A}_i (which exists by Classical Theorem 4.10). Up to renaming the elements, we can assume that \mathcal{A}_1 is composed of the k n-subsets $\{1,2,\ldots,n\}$, $\{n+1,n+2,\ldots,2n\}$, \ldots and $\{(k-1)n+1,(k-1)n+2,\ldots,kn\}$. Since we have an intersecting family of distinct subsets, no two subsets can be in the same \mathcal{A}_i , and we have as many subsets as \mathcal{A}_i s, which means that one of the subsets is in \mathcal{A}_1 , hence that it is one of the particular subsets we are looking for. This proves that GENERAL-ERDŐS-KO-RADO $_k \in \mathsf{TFNP}$. We then have the following result.

▶ Theorem 4.16. Under Assumption 4.12, GENERAL-ERDŐS-KO-RADO_k is PPP-complete.

5 Sperner's Theorem on Largest Antichains

We now turn our attention to a different existence theorem from extremal combinatorics, concerning antichains. We say a family of sets $\mathcal{F} \subseteq 2^{\Omega}$ is an *antichain* if for every $A \neq B \in \mathcal{F}$, it holds that $A \not\subseteq B$. A well-known theorem by Sperner gives a characterization of the largest antichain. As before, for an appropriate input size, this induces a total search problem of finding two distinct sets A, B for which $A \subseteq B$. As in the previous section, we consider both a weak and a strong version, and prove the weak version to be PWPP-complete, and the strong one PPP-complete.

▶ Classical Theorem 5.1 (Sperner [27]). The largest antichain on a universe of 2n elements is unique and consists of all subsets of size n.

Like before, we consider an implicit representation of the collection of subsets via a circuit C whose input corresponds to an index into the collection, and whose output is the characteristic vector of the corresponding set.

▶ **Definition 5.2** (WEAK-SPERNER-ANTICHAIN). The problem WEAK-SPERNER-ANTICHAIN is defined by the relation

Instance: A Boolean circuit $C: \{0,1\}^{\lceil \log \binom{2n}{n} \rceil + 1} \to \{0,1\}^{2n}$.

Solution: $x \neq y$ s.t. $C(x) \subseteq C(y)$.

▶ Theorem 5.3. WEAK-SPERNER-ANTICHAIN is PWPP-complete

PPP-completeness using the tight bound

As with Erdős-Ko-Rado, we observe that the bound in theorem is tight, and we know the unique antichain of size $\binom{2n}{n}$, so we have some structural information about any collection of size $\binom{2n}{n}$. From that strong theorem, employing the same technique as before, we modify the problem to let the circuit represent a collection of that exact size. By Classical Theorem 5.1, we observe that if \mathcal{F} is an antichain with $|\mathcal{F}| = \binom{2n}{n}$, then \mathcal{F} must contain [n]. This leads us to define the following problem.

▶ Definition 5.4 (Sperner-Antichain). The problem Sperner-Antichain is defined by the relation

Instance: A Boolean circuit $C: \{0,1\}^{\lceil \log \binom{2n}{n} \rceil \rceil} \to \{0,1\}^{2n}$.

Solution: *One of the following:*

- (i) $x \neq y$ s.t. $C(x) \subseteq C(y)$ and $x, y < \binom{2n}{n}$, (ii) x s.t. $C(x) = \overline{[n]}$ and $x < \binom{2n}{n}$.
- ▶ **Theorem 5.5.** *SPERNER-ANTICHAIN is PPP-complete.*

6 Cayley's Tree Formula

We consider yet another classic theorem from combinatorics, related to spanning trees. A classic result by Cayley establishes the number of spanning trees of the complete graph on n vertices. We observe then that if we have a collection of sufficiently many such graphs, either one of the graphs is not a spanning tree, or two spanning trees collide. Note that two isomorphic trees on distinct vertices are not considered a collision. This allows us to define a total search problem of either finding a collision or finding an index not corresponding to a spanning tree. We represent trees using a bitmap on all possible edges, ordered arbitrarily. We show that this problem is equivalent to WEAK-PIGEON, in a more direct way than for the previous results. As before, the problem can be modified using the same technique as previously to become equivalent to Pigeon, and thus PPP-complete.

- ▶ Classical Theorem 6.1 (Cayley [4]). There are exactly n^{n-2} spanning trees of the complete graph on n vertices.
- ▶ Definition 6.2 (WEAK-CAYLEY). The problem WEAK-CAYLEY is defined by the relation **Instance:** A Boolean circuit $C: \{0,1\}^{\lceil (n-2)\log(n)\rceil+1} \to \{0,1\}^{\binom{n}{2}}$.

Solution: *One of the following:*

- (i) x s.t. C(x) is not a spanning tree (i.e., is not spanning, not connected or contains a cycle),
- (ii) $x \neq y \text{ s.t. } C(x) = C(y).$
- ▶ **Theorem 6.3.** WEAK-CAYLEY is PWPP-complete.

PPP-completeness using the tight bound

Again, we observe that Classical Theorem 6.1 gives an exact bound, namely that there are exactly n^{n-2} labelled spanning trees on n vertices. As before, this leads us to defining the following problem.

▶ **Definition 6.4** (CAYLEY). The problem CAYLEY is defined by the relation

Instance: A Boolean circuit $C: \{0,1\}^{\lceil (n-2)\log(n) \rceil} \to \{0,1\}^{\binom{n}{2}}$.

- (i) x s.t. C(x) is not a spanning tree and $x < n^{n-2}$,
- (ii) $x \neq y$ s.t. C(x) = C(y) and $x < n^{n-2}$,
- (iii) x s.t. $C(x) = T_1$ and $x < n^{n-2}$, with T_1 defined as in Remark 3.4.
- **► Theorem 6.5.** CAYLEY is PPP-complete.

7 Ward-Szabó Theorem on Swell Colorings

We now focus on a different theorem from extremal combinatorics, and more precisely from extremal graph theory. Let G = (V, E) be the complete graph on N vertices. An edge-coloring $c \colon E \to [r]$ for some r is called a *swell coloring* of G if it uses at least 2 colors and if every triangle is either monochromatic or trichromatic. It is rather straightforward to see that in any 2-coloring of G, there must exist a bichromatic triangle. On the contrary, if we color each edge with a different color, we trivially get a swell coloring. The natural question that appears is then to determine the minimal number of colors required to swell-color the complete graph on N vertices. This was solved in some cases by Ward and Szabó in 1994.

▶ Classical Theorem 7.1 (Ward-Szabó [29]). The complete graph on N vertices cannot be swell-colored with fewer than $\lceil \sqrt{N} \rceil + 1$ colors, and this bound is tight when $N = p^{2k}$ for some prime p and integer k.

From that theorem, we can define a TFNP problem as follows: the input is a coloring C of the edges of the complete graph on 2^{2n} vertices with 2^n colors, as well as three vertices a, b, c such that $C(a, b) \neq C(a, c)$ to guarantee that at least 2 colors are used in the coloring. A solution is then the vertices of a bichromatic triangle (which is guaranteed to exist by Classical Theorem 7.1). We also allow extra solutions if the coloring of the graph is not consistent.

- ▶ Definition 7.2 (WARD-SZABÓ). The problem WARD-SZABÓ is defined by the relation Instance: The following:
 - **1.** A Boolean circuit $C: \{0,1\}^{2n} \times \{0,1\}^{2n} \to \{0,1\}^n$; and,
 - **2.** Distinct $a, b, c \in \{0, 1\}^{2n}$ s.t. $C(a, b) \neq C(a, c)$.

Solution: *One of the following:*

- (i) $x, y \ s.t. \ C(x, y) \neq C(y, x),$
- (ii) Distinct x, y, z s.t. $C(x, y) = C(y, z) \neq C(x, z)$.

We also define two variants of this problem, whose totality is a consequence of the totality of WARD-SZABÓ.

In the first one, we allow an extra type of solution, namely the vertices of two distinct triangles with the same "color profile".

▶ Definition 7.3 (WARD-SZABÓ-COLLISIONS). The problem WARD-SZABÓ-COLLISIONS is defined by the relation

Instance: The following:

- **1.** A Boolean circuit $C: \{0,1\}^{2n} \times \{0,1\}^{2n} \to \{0,1\}^n$; and,
- **2.** Distinct $a, b, c \in \{0, 1\}^{2n}$ s.t. $C(a, b) \neq C(a, c)$.

- (i) $x, y \text{ s.t. } C(x, y) \neq C(y, x),$
- (ii) Distinct x, y, z s.t. $C(x, y) = C(y, z) \neq C(x, z)$,
- (iii) Two triples, (x, y, z), (x', y', z'), each with 3 distinct elements, s.t. $\{x, y, z\} \neq \{x', y', z'\}$ and C(x, y) = C(x', y'), C(x, z) = C(x', z'), C(y, z) = C(y', z').

In the second variant, we allow the same extra type of solution, namely the vertices of two distinct triangles with the same "color profile", with the additional constraint that these triangles should be trichromatic.

▶ Definition 7.4 (Ward-Szabó-Colorful-Collisions). The problem Ward-Szabó-Colorful-Collisions is defined by the relation

Instance: The following:

- 1. A Boolean circuit $C: \{0,1\}^{2n} \times \{0,1\}^{2n} \to \{0,1\}^n$; and,
- **2.** Distinct $a, b, c \in \{0, 1\}^{2n}$ s.t. $C(a, b) \neq C(a, c)$.

Solution: One of the following:

- (i) $x, y \text{ s.t. } C(x, y) \neq C(y, x),$
- (ii) Distinct x, y, z s.t. $C(x, y) = C(y, z) \neq C(x, z)$,
- (iii) Two triples (x, y, z), (x', y', z'), each with 3 distinct elements, s.t. $\{x, y, z\} \neq \{x', y', z'\}$, C(x, y) = C(x', y'), C(x, z) = C(x', z'), C(y, z) = C(y', z') and the triangle (x, y, z) is trichromatic.
- ▶ Theorem 7.5. It holds that
- 1. Weak-Pigeon < Ward-Szabó-Collisions.
- 2. Ward-Szabó-Collisions ≤ Ward-Szabó-Colorful-Collisions,
- 3. Ward-Szabó-Colorful-Collisions < Ward-Szabó.
- ▶ Theorem 7.6. WARD-SZABÓ-COLLISIONS \in PWPP.
- ▶ Remark 7.7. The last two theorems prove that Ward-Szabó-Collisions is PWPP-complete. However, notice that the proof of inclusion into PWPP does not use solutions of the first three types. Hence, if we call Ward-Szabó-Collisions' the problem similar to Ward-Szabó-Collisions but without the first three types of solutions, this new problem is also PWPP-complete. Indeed, the proof of inclusion into PWPP would be similar, and the proof of hardness too, only with less cases to consider. Thus, it seems (at least that is how we prove it) that what makes Ward-Szabó-Collisions PWPP-complete is only its last type of solutions. Now, one could wonder how hard this problem becomes if we slightly modify this last type of solutions to make them harder to find. This is exactly what Ward-Szabó-Colorful-Collisions does.
- ▶ Theorem 7.8. WARD-SZABÓ-COLORFUL- $COLLISIONS \in PPP$.
- ▶ Remark 7.9. In the last proof, we define a reduction to PIGEON where the circuit C' only has a range of $2^{2n-1}+2^{n-1}$ elements. Indeed, we need exactly $\binom{2^n}{2}=2^{2n-1}-2^{n-1}$ elements to encode the pairs of colors. We also need exactly 2^n elements for the fourth case. However, we can map the x anywhere in that case if $C(x,a) \in \{C(a,b),C(a,c),C(b,c)\}$ because such an x would give us a bichromatic triangle. Hence, we need 2^n-3 colors for this case. We also need 3 extra elements for a,b and c. Hence, overall, we only need a range of $2^{2n-1}+2^{n-1}$ elements. Thus, we get a reduction from Ward-Szabó-Colorful-Collisions to General-Pigeon $2^{n-1}-2^{n/2-1}$.

8 Mantel's Theorem on Triangle-Free Graphs

We move on to another classical theorem in extremal graph theory. It answers the following question: What is the maximum number of edges in a triangle-free graph on N vertices?

▶ Classical Theorem 8.1 (Mantel [19]). If G = (V, E) is a triangle-free graph on N vertices then $|E| \leq N^2/4$, and this bound is tight.

This gives rise to the following search problem. Suppose that we are given a collection of strictly more than $N^2/4$ distinct edges for a graph on N vertices. Then, by Mantel's theorem, there must be three of these edges forming a triangle in the graph. The search problem is then to find them. We can turn this problem into a TFNP problem if we also allow evidence that two edges in the collection are in fact the same, or that an edge is in fact a loop. For practical reasons, we demand that the endpoints of every edge are given in the lexicographic order. When the edges are represented implicitly by a poly-sized circuit, we get the following problem.

▶ **Definition 8.2** (WEAK-MANTEL). The problem WEAK-MANTEL is defined by the relation **Instance**: A Boolean circuit $C: \{0,1\}^{2n-1} \rightarrow \{0,1\}^n \times \{0,1\}^n$.

Solution: *One of the following:*

- (i) Distinct i, j, k s.t. C(i), C(j), C(k) form a triangle,
- (ii) i s.t. C(i) = (u, v) with $u \ge v$ in the lexicographic order,
- (iii) $i \neq j \text{ s.t. } C(i) = C(j).$
- ▶ Remark 8.3. Like in the other problems, the size of the collection we receive (in this case, edges) is twice the threshold size (here, 2^{2n-2}). However, here, we observe that the number of edges we receive as input is greater than the number of possible edges since $2^{2n-1} > {2n \choose 2}$. Thus, in any instance of WEAK-MANTEL, there *must* be solutions of type (ii) or (iii).
- ▶ **Theorem 8.4.** WEAK-MANTEL is PWPP-hard.
- ▶ Theorem 8.5. $WEAK-MANTEL \in PPP$.
- ▶ Remark 8.6. Similarly to the proof that WARD-SZABÓ-COLLISIONS \in PPP, we only use the last two types of solutions, which suggests that what makes this problem reducible to PIGEON is only the fact that we are given more edges than there are different possible edges in a graph on 2^n vertices.
- ▶ Remark 8.7. In fact, the range of the circuit in this last proof has size $2^{2n-1} 2^{n-1}$ and therefore WEAK-MANTEL reduces to GENERAL-PIGEON_{2(n-1)/2}.

Mantel's theorem states that there is a unique triangle-free graph on 2N vertices that has N^2 edges, it is the complete bipartite graph $K_{N,N}$. Now, consider any labelling of the vertices of $K_{N,N}$. If for every label x, the vertices labelled x and $x+1 \mod 2N$ were on the same side of the bipartition, then all the vertices would be on the same side of the bipartition, which is impossible. Hence, there must be 2 vertices labelled x and $x+1 \mod 2N$ on different sides of the bipartition, and therefore there must be an edge between them. Thus, the following problem is total.

▶ **Definition 8.8** (Mantel). The problem Mantel is defined by the relation

Instance: A Boolean circuit $C: \{0,1\}^{2n-2} \rightarrow \{0,1\}^n \times \{0,1\}^n$.

- (i) Distinct i, j, k s.t. C(i), C(j), C(k) form a triangle,
- (ii) i s.t. C(i) = (u, v) with $u \ge v$ in the lexicographic order,
- (iii) $i \neq j$ s.t. C(i) = C(j),
- (iv) i s.t. C(i) = (u, v) with $v = u + 1 \mod 2^n$ when we consider u and v as integers.
- ▶ Theorem 8.9. MANTEL is PPP-hard.

8.1 Generalization with Turán's Theorem

Mantel's theorem investigates the maximum number of edges in a triangle-free graph on N vertices. Similarly, one could wonder about the maximum number of edges in a graph on N vertices that does not contain a clique on r vertices, where $r \geq 3$ is an arbitrary constant. This problem was solved by Turán in 1941.

▶ Classical Theorem 8.10 (Turán [28]). If G = (V, E) is a graph on N = |V| vertices that does not contain any r+1-clique, then $|E| \le (1-\frac{1}{r})\frac{N^2}{2}$ and this bound is tight when r divides N.

Now, suppose that we are given a list of strictly more than $(1-\frac{1}{r})\frac{N^2}{2}$ edges for a graph on N vertices. It follows from Turán's theorem that if all these edges are distinct, the graph must contain an r+1-clique. This induces a total search of finding the vertices of such a clique. If the edges are given implicitly via a Boolean circuit which on input i returns the endpoints of the i-th edge, we get the following TFNP problem.

- ▶ Definition 8.11 (WEAK-TURÁN_r). The problem WEAK-TURÁN_r is defined by the relation Instance: A Boolean circuit $C: \{0,1\}^{2n-1} \to \{0,1\}^n \times \{0,1\}^n$. Solution: One of the following:
 - (i) Distinct $i_1, i_2, \ldots i_{\binom{r+1}{2}}$ such that $C(i_1), C(i_2), \ldots C(i_{\binom{r+1}{2}})$ are the edges of an r+1-clique.
 - (ii) i s.t. C(i) = (u, v) with $u \ge v$ in the lexicographic order,
 - (iii) $i \neq j \text{ s.t. } C(i) = C(j).$
- \triangleright Remark 8.12. Note that r could be any polynomial in n in the previous definition and it would still define a TFNP problem.
- ▶ Theorem 8.13. For every $r_1 < r_2$, WEAK-TURÁ N_{r_1} is reducible to WEAK-TURÁ N_{r_2} .
- ▶ **Theorem 8.14.** For every $r \ge 2$, WEAK-TURÁN_r is PWPP-hard.
- ▶ **Theorem 8.15.** For every r > 2, WEAK-TURÁN $_r \in PPP$.

The proof is exactly similar to the proof of *Theorem* 8.5. In this case too, it appears that what makes the problem easier than PIGEON is that we are given too many edges.

Turán's theorem states that if r divides N, there is a unique graph on N vertices that does not contain any r+1-clique and that has the maximum number of edges. This graph is the complete r-partite graph, where each part has size N/r. Like previously, there must be 2 vertices labelled x and $x+1 \mod 2N$ with an edge between them. We denote by N the largest multiple of r that is at most 2^n , and set $M=(1-\frac{1}{r})\frac{N^2}{2}$. Thus, the following problem is in TFNP.

- ▶ Definition 8.16 (Turá N_r). The problem Turá N_r is defined by the relation Instance: The following:
 - 1. A Boolean circuit $C: \{0,1\}^{2n-1} \to \{0,1\}^n \times \{0,1\}^n$; and,
- **2.** Two integers N and M satisfying $N \leq 2^n < N+r$, r divides N, and $M = (1-\frac{1}{r})\frac{N^2}{2}$. **Solution:** One of the following:
 - (i) i s.t. C(i) = (u, v) with $u \ge N$ or $v \ge N$, and i < M
 - (ii) Distinct $i_1, i_2, \ldots i_{\binom{r+1}{2}}$ such that $C(i_1), C(i_2), \ldots C(i_{\binom{r+1}{2}})$ are the edges of an r+1-clique, and $i_j < M$ for every j,
 - (iii) i s.t. C(i) = (u, v) with $u \ge v$ in the lexicographic order, and i < M,

- (iv) $i \neq j$ s.t. C(i) = C(j), and i, j < M,
- (v) i s.t. C(i) = (u, v) with $v = u + 1 \mod 2^n$ when we consider u and v as integers, and i < M.

This last problem is in TFNP. The reduction from [24] implies $\text{Tur\'an}_{r_1} \leq \text{Tur\'an}_{r_2}$ for every $r_1 < r_2$. Then, Tur\'an_r is PPP-hard because Tur\'an_2 is exactly Mantel.

- References -

- 1 Frank Ban, Kamal Jain, Christos H. Papadimitriou, Christos-Alexandros Psomas, and Aviad Rubinstein. Reductions in PPP. *Inf. Process. Lett.*, 145:48–52, 2019. doi:10.1016/j.ipl. 2018.12.009.
- Zsolt Baranyai. Infinite and finite sets, vol. 1. proceedings of a colloquium held at Keszthely, June 25 – July 1, 1973. Dedicated to Paul Erdős on his 60th Birthday. J. Symb. Log., 1:91–108, 1973.
- 3 Romain Bourneuf, Lukáš Folwarczný, Pavel Hubáček, Alon Rosen, and Nikolaj Ignatieff Schwartzbach. PPP-completeness and extremal combinatorics, 2022. doi:10.48550/arXiv. 2209.04827.
- 4 Arthur Cayley. A theorem on trees. Quarterly Journal of Mathematics, 23:376–378, 1889.
- 5 Thomas M. Cover. Enumerative source encoding. *IEEE Transactions on Information Theory*, 19(1):73-77, 1973. doi:10.1109/TIT.1973.1054929.
- 6 Ivan Bjerre Damgård. A design principle for hash functions. In Gilles Brassard, editor, Advances in Cryptology – CRYPTO' 89 Proceedings, pages 416–427, New York, NY, 1990. Springer New York.
- 7 Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The complexity of computing a Nash equilibrium. SIAM J. Comput., 39(1):195–259, 2009. doi:10.1137/070699652.
- 8 Paul Erdős. Some remarks on the theory of graphs. Bulletin of the American Mathematical Society, 53(4):292–294, 1947.
- 9 Paul Erdős, Chao Ko, and Richard Rado. Intersection theorems for systems of finite sets. *The Quarterly Journal of Mathematics*, 12(1):313–320, January 1961. doi:10.1093/qmath/12.1.
- Paul Erdős and Richard Rado. Intersection theorems for systems of sets. *Journal of the London Mathematical Society*, s1-35(1):85-90, 1960. doi:10.1112/jlms/s1-35.1.85.
- Aris Filos-Ratsikas and Paul W. Goldberg. Consensus halving is PPA-complete. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 51–64. ACM, 2018. doi:10.1145/3188745.3188880.
- Paul W. Goldberg and Christos H. Papadimitriou. TFNP: an update. In Dimitris Fotakis, Aris Pagourtzis, and Vangelis Th. Paschos, editors, Algorithms and Complexity 10th International Conference, CIAC 2017, Athens, Greece, May 24-26, 2017, Proceedings, volume 10236 of Lecture Notes in Computer Science, pages 3-9, 2017. doi:10.1007/978-3-319-57586-5_1.
- Pavel Hubáček and Jan Václavek. On search complexity of discrete logarithm. In Filippo Bonchi and Simon J. Puglisi, editors, 46th International Symposium on Mathematical Foundations of Computer Science, MFCS 2021, August 23-27, 2021, Tallinn, Estonia, volume 202 of LIPIcs, pages 60:1–60:16. Schloss Dagstuhl Leibniz-Zentrum für Informatik, 2021. doi: 10.4230/LIPIcs.MFCS.2021.60.
- Emil Jeřábek. Integer factoring and modular square roots. J. Comput. Syst. Sci., 82(2):380–394, 2016. doi:10.1016/j.jcss.2015.08.001.
- David S. Johnson, Christos H. Papadimitriou, and Mihalis Yannakakis. How easy is local search? J. Comput. Syst. Sci., 37(1):79–100, 1988. doi:10.1016/0022-0000(88)90046-3.

22:20 PPP-Completeness and Extremal Combinatorics

- Ilan Komargodski, Moni Naor, and Eylon Yogev. White-box vs. black-box complexity of search problems: Ramsey and graph property testing. J. ACM, 66(5), July 2019. doi: 10.1145/3341106.
- 17 Leszek Aleksander Kołodziejczyk and Neil Thapen. Approximate counting and NP search problems. CoRR, abs/1812.10771, 2018. arXiv:1812.10771.
- Jan Krajíček. Structured pigeonhole principle, search problems and hard tautologies. *J. Symb. Log.*, 70(2):619–630, 2005. doi:10.2178/jsl/1120224731.
- Willem Mantel. Problem 28 (Solution by H. Gouwentak, W. Mantel, J. Teixeira de Mattes, F. Schuh and W. A. Wythoff). *Wiskundige Opgaven*, 18:60–61, 1907.
- Nimrod Megiddo and Christos H. Papadimitriou. On total functions, existence theorems and computational complexity. Theor. Comput. Sci., 81(2):317–324, 1991. doi:10.1016/0304-3975(91)90200-L.
- 21 Ruta Mehta. Constant rank two-player games are PPAD-hard. SIAM J. Comput., 47(5):1858–1887, 2018. doi:10.1137/15M1032338.
- 22 Ralph Charles Merkle. Secrecy, Authentication, and Public Key Systems. PhD thesis, Stanford University, Stanford, CA, USA, 1979. AAI8001972.
- Christos H. Papadimitriou. On the complexity of the parity argument and other inefficient proofs of existence. J. Comput. Syst. Sci., 48(3):498–532, 1994. doi:10.1016/S0022-0000(05) 80063-7.
- Amol Pasarkar, Mihalis Yannakakis, and Christos H. Papadimitriou. Extremal combinatorics, iterated pigeonhole arguments, and generalizations of PPP. *CoRR*, abs/2209.07625, 2022. doi:10.48550/arXiv.2209.07625.
- Frank P. Ramsey. On a Problem of Formal Logic. *Proceedings of the London Mathematical Society*, s2-30(1):264–286, January 1930. doi:10.1112/plms/s2-30.1.264.
- 26 Katerina Sotiraki, Manolis Zampetakis, and Giorgos Zirdelis. PPP-completeness with connections to cryptography. In Mikkel Thorup, editor, 59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018, pages 148–158. IEEE Computer Society, 2018. doi:10.1109/FOCS.2018.00023.
- 27 Emanuel Sperner. Ein Satz über Untermengen einer endlichen Menge. *Mathematische Zeitschrift*, 27(1):544–548, 1928.
- 28 Paul Turán. On an extremal problem in graph theory (in Hungarian). Matematikai és Fizikai Lapok, 48:436–452, 1941.
- 29 C. Ward and S. Szabó. On swell-colored complete graphs. *Acta Mathematica Universitatis Comenianae*. New Series, 63(2):303–308, 1994. URL: http://eudml.org/doc/118694.