Identifiability of Graphs with Small Color Classes by the Weisfeiler-Leman Algorithm

Frank Fuhlbrück
Institut für Informatik, Humboldt-Universität zu Berlin, Germany
fuhlbfra@informatik.hu-berlin.de

Johannes Köbler
Institut für Informatik, Humboldt-Universität zu Berlin, Germany
koebler@informatik.hu-berlin.de

Oleg Verbitsky
Institut für Informatik, Humboldt-Universität zu Berlin, Germany
verbitsky@informatik.hu-berlin.de

───── **Abstract** ─────

It is well known that the isomorphism problem for vertex-colored graphs with color multiplicity at most 3 is solvable by the classical 2-dimensional Weisfeiler-Leman algorithm (2-WL). On the other hand, the prominent Cai-Fürer-Immerman construction shows that even the multidimensional version of the algorithm does not suffice for graphs with color multiplicity 4. We give an efficient decision procedure that, given a graph G of color multiplicity 4, recognizes whether or not G is identifiable by 2-WL, that is, whether or not 2-WL distinguishes G from any non-isomorphic graph. In fact, we solve the more general problem of recognizing whether or not a given coherent configuration of maximum fiber size 4 is separable. This extends our recognition algorithm to directed graphs of color multiplicity 4 with colored edges.

Our decision procedure is based on an explicit description of the class of graphs with color multiplicity 4 that are *not* identifiable by 2-WL. The Cai-Fürer-Immerman graphs of color multiplicity 4 distinctly appear here as a natural subclass, which demonstrates that the Cai-Fürer-Immerman construction is not ad hoc. Our classification reveals also other types of graphs that are hard for 2-WL. One of them arises from patterns known as (n_3)-*configurations* in incidence geometry.

1 Introduction

Over 50 years ago, Weisfeiler and Leman [34] described a natural combinatorial procedure that since then constantly plays a significant role in the research on the graph isomorphism problem. The procedure is now most often referred to as the 2-dimensional Weisfeiler-Leman algorithm (2-WL). It generalizes and improves the classical color refinement method (1-WL)

and has an even more powerful k-dimensional version (k-WL) for any $k > 2$. The original 2-dimensional version and the logarithmic-dimensional enhancement are important components in Babai's quasipolynomial-time isomorphism algorithm [4].

Even on its own, 2-WL is a quite powerful tool in isomorphism testing. For instance, it solves the isomorphism problem for several important graph classes, in particular, for interval graphs [16]. Also, it is successful for almost all regular graphs of a fixed degree [5]. On the other hand, not every pair of non-isomorphic graphs is distinguishable by 2-WL. For example, it cannot detect any difference between two non-isomorphic strongly regular graphs with the same parameters.

We call a graph G *amenable to* k-WL if the algorithm distinguishes G from any non-isomorphic graph. An efficient characterization of the class of graphs amenable to 1-WL is obtained by Arvind et al. in [2], where it is given also for vertex-colored graphs. Independently, Kiefer et al. [26] give an efficient criterion of amenability to 1-WL in a more general framework including also directed graphs with colored edges. Similar results for 2-WL are currently out of reach. A stumbling block here is the lack of understanding which strongly regular graphs are uniquely determined by their parameters. Note that a strongly regular graph is determined by its parameters up to isomorphism if and only if it is amenable to 2-WL.

A general strategy to approach a hard problem is to examine its complexity in the parameterized setting. We consider vertex-colored graphs with the *color multiplicity*, that is, the maximum number of equally colored vertices, as parameter. If this parameter is bounded, the graph isomorphism problem is known to be efficiently solvable. More specifically, it is solvable in time polynomial in the number of vertices and quasipolynomial in the color multiplicity [4, Corollary 4], and it is solvable in polylogarithmic parallel time [27]. Graph Isomorphism is known to be in the Mod_kL hierarchy for any fixed color multiplicity [3], and even in $\oplus\text{L} = \text{Mod}_2\text{L}$ for color multiplicity at most 5 [1].

Every graph of color multiplicity at most 3 is amenable to 2-WL (Immerman and Lander [24]). Starting from color multiplicity 4, the amenability concept is non-trivial: The prominent Cai-Fürer-Immerman construction [9] shows that for any k, there exist graphs with color multiplicity 4 that are not amenable to k-WL.

We design an efficient decision procedure that, given a graph G with color multiplicity 4, recognizes whether or not G is amenable to 2-WL. Note that an a priori upper complexity bound for this decision problem is coNP, as a consequence of the aforementioned fact that Graph Isomorphism for graphs of bounded color multiplicity is in P. From now on, amenability is meant with respect to 2-WL, unless stated otherwise.

We actually solve a much more general problem. 2-WL transforms an input graph G, possibly with colored vertices and directed and colored edges, into a *coherent configuration* $\mathcal{C}(G)$, which is called the *coherent closure* of G. The concept of a coherent configuration has been discovered independently in statistics [6] and algebra [22] and, playing an important role in diverse areas, has been developed to the subject of a rich theory; see a recent monograph [10], that we will use in this paper as a reference book. A coherent configuration $\mathcal{C}$ is called *separable* if the isomorphism type of $\mathcal{C}$ is determined by its regularity parameters in a certain strong sense; see the definition in Section 2. The separability of the coherent closure $\mathcal{C}(G)$ implies the amenability of the graph G. This was the approach undertaken in [16], where it was shown that the coherent closure of any interval graph is separable. Somewhat less obviously, the converse relation between amenability of G and separability of $\mathcal{C}(G)$ is also true: For every graph G,

$$G \text{ is amenable if and only if } \mathcal{C}(G) \text{ is separable;} \tag{1}$$

see Theorem 2.1 in Section 2. Equivalence (1) reduces the amenability problem for graphs to the separability problem for coherent configurations. This reduction works as well for

directed graphs with colored vertices and colored edges, that is, essentially for arbitrary binary relational structures. If G has color multiplicity b, then the maximum *fiber* size of $\mathcal{C}(G)$ is also bounded by b (see Section 2 for the definitions). While all coherent configurations with fibers of size at most 3 are known to be separable [10], the separability property for coherent configurations with fibers of size 4 is non-trivial, and our first result is this.

▶ **Theorem 1.1.** *The problem of deciding whether a given coherent configuration with maximum fiber size 4 is separable is solvable in* $\oplus\mathrm{L}$.

Since $\oplus\mathrm{L} \subseteq \mathrm{NC}^2$ (which follows from the inclusion $\#L \subseteq \mathrm{NC}^2$ in [35]), Theorem 1.1 implies that the separability problem is solvable in parallel polylogarithmic time. Using the reduction (1), we obtain our result for graphs.

▶ **Theorem 1.2.** *The problem of deciding whether a given vertex-colored graph with maximum color multiplicity 4 is amenable to* 2-WL *is solvable in* P. *This holds true also for vertex- and edge-colored directed graphs.*

More precisely, the proof of Theorem 1.2 yields an algorithm deciding amenability of graphs of color multiplicity at most 4 with running time $O(n^{2+\omega})$, where $\omega < 2.373$ is the exponent of fast matrix multiplication [19]. Using randomization, the running time can be improved to $O(n^4 \log^2 n)$.

Theorems 1.1 and 1.2 are proved in Section 5. The proof is based on a combinatorial Cut-Down Lemma in Section 3, which reduces deciding separability of a coherent configuration $\mathcal{C}$ with maximum fiber size 4 to deciding separability of a subconfiguration $\mathcal{C}'$ of $\mathcal{C}$ belonging to a class of well-structured coherent configurations which we call *irredundant*. Separability of irredundant configurations is studied in Section 4, where it is recast as a question about a certain permutation group.

Our results have the following consequences, which we discuss in Section 6.

Highlighting the inherent structure of the Cai-Fürer-Immerman graphs. The essence of our proof of Theorem 1.2 is an explicit description of the class of graphs with color multiplicity 4 that are *not* amenable to 2-WL. The Cai-Fürer-Immerman graphs of color multiplicity 4 distinctly appear here as a natural subclass, which demonstrates that the Cai-Fürer-Immerman construction is not ad hoc. In a sense, the famous CFI gadget [9, Fig. 3] (or [25, Fig. 13.24]) appears in our analysis inevitably "by itself".[1]

While the CFI graphs have many automorphisms, Gurevich and Shelah [21] came up with a construction of (non-binary) *multipede* structures that are rigid and yet not identifiable by k-WL. Neuen and Schweitzer [30, 31] combined both approaches to construct *multipede graphs* and to give sufficient conditions ensuring that these graphs are not amenable to k-WL (see also a recent related paper [12]). The multipede graphs are vertex-colored and the results of [30, 31] make perfect sense if the color multiplicity is bounded by 4. An irredundant coherent configuration typically admits a natural representation by a multipede graph and vice versa; see Remark 6.6. Though non-amenability to k-WL for higher dimensions implies non-amenability to 2-WL, the results obtained in [12, 30, 31] and in our paper are incomparable as we provide both sufficient and *necessary* conditions for 2-WL-non-amenability.

---

[1] More precisely, this concerns a simplified version of the CFI gadget, where each vertex in a cubic pattern graph is replaced with a quadruple of new vertices and two quadruples are connected by edges directly, and not via two extra pairs of auxiliary vertices as in the original version; cf. Fig. 4. The simplified gadget appears in an algebraic analog of the CFI result by Evdokimov and Ponomarenko [15]; see also Fürer's survey paper [18]. This gadget comes out also in the *shrunken* multipede graphs [30].

More graphs hard for 2-WL. Our analysis reveals new types of non-amenable graphs. A particularly elegant construction is based on the well-studied (n_3)-configurations of lines and points [20, 32]. For example, the 7-point Fano plane and the 9-point Pappus configuration give rise to non-amenable graphs of color multiplicity 4 with, respectively, 28 and 36 vertices.

Classification of small graphs. Our amenability criteria are easy to apply in many cases. In particular, they imply that all graphs of color multiplicity 4 with no more than 15 vertices are amenable. Among graphs of color multiplicity 4 with 16 vertices there are 434 non-amenable graphs, which are split into 217 pairs of 2-WL indistinguishable non-isomorphic graphs.

All proofs omitted in this version of the paper can be found in [17].

2 Basic definitions and facts

Let V be a set, whose elements are called *points*. Let $\mathcal{C} = \{R_1, \ldots, R_s\}$ be a partition of the Cartesian square V^2, that is, $\bigcup_{i=1}^{s} R_i = V^2$ and any two R_i and R_j are disjoint. An element R of $\mathcal{C}$ will be referred to as a *basis relation*. $\mathcal{C}$ is called a *coherent configuration* on $V = V(\mathcal{C})$ if it has the following properties:

(A) If a basis relation $R \in \mathcal{C}$ contains a loop vv, then all pairs in R are loops;

(B) For every $R \in \mathcal{C}$, the *transpose relation* $R^* = \{uv : vu \in R\}$ is also in $\mathcal{C}$.

(C) For every triple $R, S, T \in \mathcal{C}$, the number $p(uv) = |\{w : uw \in R, wv \in S\}|$ is the same for all $uv \in T$.

For a coherent configuration $\mathcal{C}$, the number $p(uv)$ in (C) does not depend on the choice of uv in T and is denoted by p_{RS}^T. The entries of this 3-dimensional matrix are called *intersection numbers* of $\mathcal{C}$.

Two coherent configurations $\mathcal{C}$ and $\mathcal{D}$ are *combinatorially isomorphic* if there is a bijection $\phi : V(\mathcal{C}) \to V(\mathcal{D})$, called a *combinatorial isomorphism* from $\mathcal{C}$ to $\mathcal{D}$, such that $\phi(R) \in \mathcal{D}$ for every $R \in \mathcal{C}$. We write $\mathcal{C} \cong_{comb} \mathcal{D}$ for this relationship. Here $\phi(R) = \{\phi(u)\phi(v) : uv \in R\}$.

Coherent configurations $\mathcal{C}$ and $\mathcal{D}$ are *algebraically isomorphic* if their 3-dimensional matrices of intersection numbers, p_{RS}^T and $p_{R'S'}^{T'}$, are isomorphic, that is, there is a bijection $f : \mathcal{C} \to \mathcal{D}$ such that

$$p_{RS}^T = p_{f(R)f(S)}^{f(T)}.$$

In this case we write $\mathcal{C} \cong_{alg} \mathcal{D}$. Such a bijection f is called an *algebraic isomorphism* from $\mathcal{C}$ to $\mathcal{D}$. Note that combinatorially isomorphic coherent configurations are also algebraically isomorphic. Indeed, any combinatorial isomorphism ϕ from $\mathcal{C}$ to $\mathcal{D}$ gives rise to the algebraic isomorphism f defined by $f(R) = \phi(R)$.

To allow a uniform treatment of ordinary graphs, vertex-colored graphs, and even edge-colored directed graphs, we formally define a *colored graph* G on a vertex set $V = V(G)$ as a function $c_G : V^2 \to C$ such that $c_G(vv) \neq c_G(uw)$ whenever $u \neq w$. For each color $c \in C$, the set $\{uv : c_G(uv) = c\}$ is called a *color class* of G. Two colored graphs G and H are isomorphic if there is a bijection $\phi : V(G) \to V(H)$ such that $c_H(\phi(u)\phi(v)) = c_G(uv)$ for all $u, v \in V(G)$. In the context of the isomorphism problem, we can always assume that

$$c_G(uv) = c_G(u'v') \text{ if and only if } c_G(vu) = c_G(v'u'), \tag{2}$$

that is, if arrows have the same color, then the inverse arrows must also be equally colored. This condition can be ensured by modifying the coloring as follows. Suppose that an arrow uv is colored *red* in G, and the inverse arrow vu is colored *blue*. Then uv is recolored a new color *redblue*, and vu is recolored a new color *bluered*. The new colored graph $\hat{G}$ satisfies the condition (2). Note that $\hat{G} \cong \hat{H}$ exactly when $G \cong H$.

We remark that ordinary graphs are covered by this setting as adjacency and non-adjacency can be seen as two distinct colors of vertex pairs. The vertex-colored graphs are covered as well as a vertex v can be seen as having color $c_G(vv)$. A set of vertices of the same color is referred to as *vertex color class*.

Given a colored graph G as input, the *2-dimensional Weisfeiler-Leman algorithm* (2-WL for short) iteratively computes colorings c_G^i of the Cartesian square V^2 for $V = V(G)$. Initially, $c_G^0 = c_G$ and then,

$$c_G^{i+1}(uv) = c_G^i(uv) \mid \left\{\!\!\left\{ c_G^i(uw) \mid c_G^i(wv) \right\}\!\!\right\}_{w \in V},$$

where $\left\{\!\!\left\{\ \right\}\!\!\right\}$ denotes the multiset and $\mid$ denotes the string concatenation (an appropriate encoding is assumed). Denote the partition of V^2 into the color classes of c_G^i by $\mathcal{R}_G^i$. Note that $\mathcal{R}_G^{i+1}$ refines $\mathcal{R}_G^i$. Let $t = t_G$ be the minimum number such that $\mathcal{R}_G^t = \mathcal{R}_G^{t-1}$. The algorithm terminates after the t-th color refinement round. It is easy to verify that $\mathcal{R}_G^t$ is a coherent configuration. Moreover, let $\mathcal{R}_G$ denote the partition of $V(G)^2$ into the color classes of G. It turns out that $\mathcal{R}_G^t$ coincides with the *coherent closure* of $\mathcal{R}_G$, which is the coarsest coherent configuration refining $\mathcal{R}_G$; see [10, Section 2.6.1]. We call this configuration the *coherent closure of the graph G* and denote it by $\mathcal{C}(G)$.

We say that colored graphs G and H are 2-WL *equivalent* and write $G \equiv_{\text{2-WL}} H$ if

$$\left\{\!\!\left\{ c_G^t(uv) \right\}\!\!\right\}_{uv \in V(G)^2} = \left\{\!\!\left\{ c_H^t(uv) \right\}\!\!\right\}_{uv \in V(H)^2} \tag{3}$$

for $t = t_G$ (equivalently, for $t = t_H$, or for all t).

Suppose that $G \equiv_{\text{2-WL}} H$. Equality (3) implies that there is a one-to-one map $f : \mathcal{C}(G) \to \mathcal{C}(H)$ preserving the 2-WL colors. Note that f is an algebraic isomorphism from $\mathcal{C}(G)$ to $\mathcal{C}(H)$. We, therefore, have the following diagram:

$$
\begin{array}{ccc}
G \cong H & \implies & G \equiv_{\text{2-WL}} H \\
\Downarrow & & \Downarrow \\
\mathcal{C}(G) \cong_{comb} \mathcal{C}(H) & \implies & \mathcal{C}(G) \cong_{alg} \mathcal{C}(H)
\end{array}
$$

We call a colored graph G *amenable (to* 2-WL*)* if 2-WL distinguishes G from any non-isomorphic graph H, that is, $G \equiv_{\text{2-WL}} H$ implies $G \cong H$.

A coherent configuration $\mathcal{C}$ is *separable* if every algebraic isomorphism from $\mathcal{C}$ to any coherent configuration $\mathcal{D}$ is induced by a combinatorial isomorphism from $\mathcal{C}$ to $\mathcal{D}$.

▶ **Theorem 2.1.** *A colored graph G is amenable if and only if its coherent closure $\mathcal{C}(G)$ is separable.*

Let $\mathcal{C}$ be a coherent configuration on the point set $V = V(\mathcal{C})$. A set of points $X \subseteq V$ is called a *fiber* of $\mathcal{C}$ if the set of loops $\{xx : x \in X\}$ is a basis relation of $\mathcal{C}$. Denote the set of all fibers of $\mathcal{C}$ by $F(\mathcal{C})$. By Property (A) of a coherent configuration, $F(\mathcal{C})$ is a partition of V. Property (C) implies that, for every basis relation R of $\mathcal{C}$ there are, not necessarily distinct, fibers X and Y such that $R \subseteq X \times Y$. Thus, if $X, Y \in F(\mathcal{C})$, then the Cartesian product $X \times Y$ is split into basis relations of $\mathcal{C}$. We denote this partition by $\mathcal{C}[X, Y]$. If $X = Y$, we simplify notation to $\mathcal{C}[X] = \mathcal{C}[X, X]$. Note that $\mathcal{C}[X]$ is a coherent configuration on X, with X being its single fiber. Such coherent configurations are called *association schemes*. We will call $\mathcal{C}[X]$ a *cell* of $\mathcal{C}$.

All possible association schemes on at most 4 points are depicted in Figure 1. Basis relations are represented by undirected edges if they are equal to their transposes, and by arrows otherwise. Loops are omitted. The 4-point cells are named K_4, C_4, $\vec{C}_4$, and F_4 according to the graphs underlying their shapes. Here, $\vec{C}_4$ stands for the directed 4-cycle, and F_4 stands for the factorization of K_4 into three matchings $2K_2$.
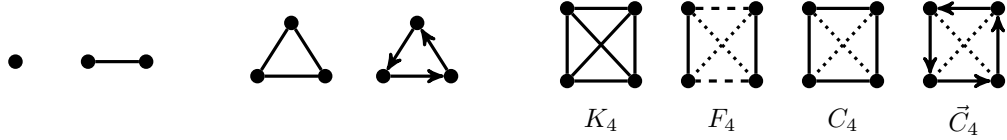
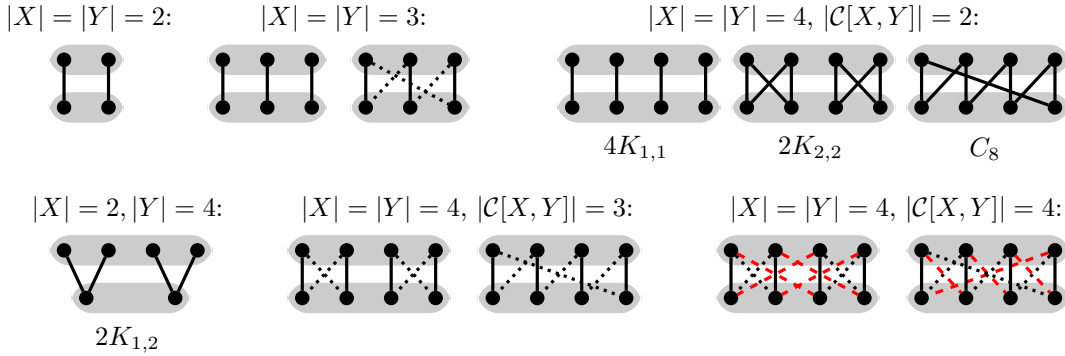**Figure 1** Cells $\mathcal{C}[X]$ on at most 4 points.

Figure 2 Non-uniform interspaces $\mathcal{C}[X,Y]$ for fibers X,Y with at most 4 points.

If $X \neq Y$, we call the partition $\mathcal{C}[X,Y]$ an *interspace* of $\mathcal{C}$. If $|\mathcal{C}[X,Y]| = 1$, that is, $X \times Y$ is a basis relation of $\mathcal{C}$, then the interspace $\mathcal{C}[X,Y]$ will be called *uniform*. If $R \in \mathcal{C}[X,Y]$, then the number of arrows in R from a point $x \in X$ is the same for each x in X. We call this number the *valency* of R and denote it by $d(R)$.

▶ **Lemma 2.2.** *Let $X,Y \in F(\mathcal{C})$. If $|X|$ and $|Y|$ are coprime, then $\mathcal{C}[X,Y]$ is uniform.*

Proof. Let R be a basis relation such that $R \subseteq \mathcal{C}[X,Y]$. Recall that the valency $d(R)$ is equal to the number of arrows in R from each point $x \in X$. Note also that the valency $d(R^*)$ of the transpose relation R^* is equal to the number of arrows in R to a point $y \in Y$; it does not depend on the choice of y. It follows that $d(R)|X| = |R| = d(R^*)|Y|$. Since $|X|$ and $|Y|$ are coprime, $d(R)$ is divisible by $|Y|$. Taking into account that $d(R) \leq |Y|$, we obtain the equality $d(R) = |Y|$. As a consequence, $R = X \times Y$. ◀

Thus, all interspaces $\mathcal{C}[X,Y]$ with $|X| = 1$ are uniform, and so are also interspaces with $|X| = 2$ and $|Y| = 3$. Figure 2 shows all non-uniform interspaces $\mathcal{C}[X,Y]$ with $|X|,|Y| \leq 4$. Here we depict pairs xy by undirected edges as they are implicitly ordered by the fibers. To facilitate visualization, one of the basis relations in each picture is missing as it is reconstructable from the others. We use the notation like $\mathcal{C}[X] \simeq C_4$, $\mathcal{C}[X,Y] \simeq 2K_{2,2}$ etc. to indicate the type of a cell or an interspace.

3   Cutting it down

Given a family of sets $\mathcal{P}$, we use $\mathcal{P}^{\cup}$ to denote the closure of $\mathcal{P}$ under unions. For any $U \in F(\mathcal{C})^{\cup}$ we let $\mathcal{C}[U]$ denote the set of all basis relations of a coherent configuration $\mathcal{C}$ contained in U^2. Note that $\mathcal{C}[U]$ is a coherent configuration on the point set U. Let $W = V \setminus U$. We say that $\mathcal{C}$ is the *direct sum* of coherent configurations $\mathcal{C}[U]$ and $\mathcal{C}[W]$ and write $\mathcal{C} = \mathcal{C}[U] \boxplus \mathcal{C}[W]$ if the interspace $\mathcal{C}[X,Y]$ is uniform for any two fibers $X,Y \in F(\mathcal{C})$ with $X \subseteq U$ and $Y \subseteq W$.
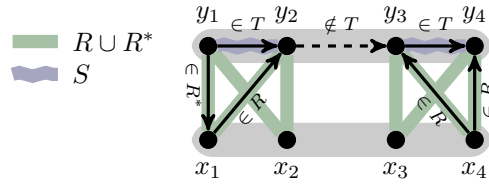
Figure 3 Proof of Lemma 3.4.

▶ **Lemma 3.1** (see [10, Corollary 3.2.8]). *Suppose that $\mathcal{C} = \mathcal{C}_1 \boxplus \mathcal{C}_2$. The coherent configuration $\mathcal{C}$ is separable if and only if both $\mathcal{C}_1$ and $\mathcal{C}_2$ are separable.*

Lemma 3.1 reduces the general separability problem to its restriction for *indecomposable* coherent configurations, that is, those configurations which cannot be split into a direct sum. Lemma 2.2 implies that an indecomposable coherent configuration of maximum fiber size at most 4 either has maximum fiber size at most 3 or has only fibers of size 4 or 2. We use the following known fact.

▶ **Lemma 3.2** (cf. [10, Exercise 3.7.20]). *Every coherent configuration $\mathcal{D}$ with maximum fiber size at most 3 is separable.*

Lemma 3.2, along with Lemmas 3.1 and 2.2, reduces the decision problem of whether a coherent configuration $\mathcal{C}$ with maximum fiber size 4 is separable to the case that $\mathcal{C}$ has fibers only of size 4 or 2. Next, we reduce the separability problem to instances having only fibers of size 4 and non-uniform interspaces of type $2K_{2,2}$.

Let M be a basis relation of a coherent configuration $\mathcal{C}$. Suppose that $M \in \mathcal{C}[X, Y]$ for distinct fibers X and Y, We call M a *matching* if M is irreflexive and both M and its transpose have valency 1, i.e., $d(M) = 1$ and $d(M^*) = 1$. This means that M determines a one-to-one correspondence between X and Y. If $M \in \mathcal{C}[X]$ for a fiber X, we additionally require that M is symmetric. In this case, M determines a partition of X into pairs of points.

The backward implication in Part 1 of the following lemma follows from [14, Lemma 9.4]. Part 2 applies, in particular, to the multipede graphs of color multiplicity at most 4. In this setting, Neuen and Schweitzer [30, Section 4.2] use exactly this shrinking operation in order to reduce the number of vertices in their construction of benchmark graphs challenging for practical isomorphism solvers.

▶ **Lemma 3.3** (Cut-Down Lemma). *Let $\mathcal{C}$ be a coherent configuration on $V = V(\mathcal{C})$.*
1. *Suppose that an interspace $\mathcal{C}[X, Y]$ contains a matching M. Then $\mathcal{C}$ is separable if and only if $\mathcal{C}[V \setminus X]$ is separable.*
2. *Suppose that $|V| > 2$, all fibers of $\mathcal{C}$ have size 4 or 2, and no interspace of $\mathcal{C}$ contains a matching. Let $X \in F(\mathcal{C})$ with $|X| = 2$. Under these conditions, $\mathcal{C}$ is separable if and only if $\mathcal{C}[V \setminus X]$ is separable.*
3. *Suppose that $|F(\mathcal{C})| > 3$, all fibers of $\mathcal{C}$ have size 4, and no interspace of $\mathcal{C}$ contains a matching. Let $\mathcal{C}[X, Y]$ be a C_8-interspace (see Figure 2). Under these conditions, $\mathcal{C}$ is separable if and only if $\mathcal{C}[V \setminus (X \cup Y)]$ is separable.*

We, therefore, focus on coherent configurations with non-uniform fibers only of type $2K_{2,2}$.

▶ **Lemma 3.4.** *Let $\mathcal{C}[X, Y] \simeq 2K_{2,2}$ and suppose that $\mathcal{C}[X, Y]$ contains a relation $R = \{x_1, x_2\} \times \{y_1, y_2\} \cup \{x_3, x_4\} \times \{y_3, y_4\}$. Then $\mathcal{C}[Y]$ contains the basis relation $S = \{y_1 y_2, y_2 y_1, y_3 y_4, y_4 y_3\}$.*

Proof. Let T be the basis relation of $\mathcal{C}[Y]$ containing the arrow $y_1 y_2$. We have $T \subseteq S$ because $p_{R^* R}^T > 0$. For example, $y_2 y_3 \notin T$ because $y_1 y_2$ extends to $y_1 x_1 y_2$ and $y_2 y_3$ cannot be extended to a triangle of this kind. On the other hand, $S \subseteq T$ because $p_{RT}^R > 0$. For example, $y_3 y_4 \in T$ because otherwise, while $x_1 y_2$ extends to $x_1 y_1 y_2$, the pair $x_4 y_4$ could not be extended to a triangle of this kind; see Figure 3.   ◀

In the context of Lemma 3.4, we say that $\mathcal{C}[X, Y]$ *determines* a matching in Y (namely $\{y_1 y_2, y_2 y_1, y_3 y_4, y_4 y_3\}$). Suppose that $\mathcal{C}[X, Y]$ determines a matching M in Y, and $\mathcal{C}[Z, Y]$ determines a matching M' in Y. We say that $\mathcal{C}[X, Y]$ and $\mathcal{C}[Z, Y]$ have a *direct connection at* Y if $M = M'$ (or are *directly connected at* Y). If $M \neq M'$, we say that $\mathcal{C}[X, Y]$ and $\mathcal{C}[Z, Y]$ have a *skewed connection at* Y (or are *askew connected at* Y).

▶ **Lemma 3.5** (Transitivity of direct $2K_{2,2}$-connections). *If $\mathcal{C}[X, Y] \simeq 2K_{2,2}$ and $\mathcal{C}[Z, Y] \simeq 2K_{2,2}$ are directly connected at Y, then either $\mathcal{C}[X, Z]$ contains a matching or $\mathcal{C}[X, Z] \simeq 2K_{2,2}$ and the connections between $\mathcal{C}[Z, X]$ and $\mathcal{C}[Y, X]$ at X and between $\mathcal{C}[X, Z]$ and $\mathcal{C}[Y, Z]$ at Z are direct.*

4   Irredundant configurations

The Cut-Down Lemma and the preceding analysis in Section 3 reduce our task to deciding separability of a coherent configuration $\mathcal{C}$ under the following three conditions:

(1) $\mathcal{C}$ is indecomposable,

(2) all fibers of $\mathcal{C}$ have size 4,

(3) every non-uniform interspace of $\mathcal{C}$ is of type $2K_{2,2}$.

A coherent configuration satisfying Conditions (1)–(3) will be called *irredundant*. Irredundant configurations are closely related to the *reduced Klein configurations* studied in [10, Section 4.1.2], but the two classes of coherent configurations are not identical. In particular, a reduced Klein configuration cannot contain C_4-cells.

We begin with noticing that, for irredundant configurations, every algebraic isomorphism f gives rise to a combinatorial isomorphism ϕ, even though ϕ does not need to induce f on the whole coherent configuration.

▶ **Lemma 4.1.** *Suppose that a coherent configuration $\mathcal{C}$ is irredundant. If f is an algebraic isomorphism from $\mathcal{C}$ to a coherent configuration $\mathcal{C}'$, then there exists a combinatorial isomorphism ϕ from $\mathcal{C}$ to $\mathcal{C}'$ such that ϕ induces f on each cell $\mathcal{C}[X]$ of $\mathcal{C}$.*

By Lemma 4.1, if a coherent configuration $\mathcal{C}$ is irredundant, then $\mathcal{C} \cong_{alg} \mathcal{C}'$ implies $\mathcal{C} \cong_{comb} \mathcal{C}'$. This has the following practical consequence: An irredundant configuration $\mathcal{C}$ is separable if and only if every algebraic *automorphism* of $\mathcal{C}$, i.e., an algebraic isomorphism from $\mathcal{C}$ to itself, is induced by a combinatorial automorphism of $\mathcal{C}$. Moreover, we call an algebraic automorphism f of $\mathcal{C}$ *strict* if f is the identity on each cell $\mathcal{C}[X]$ of $\mathcal{C}$.

▶ **Lemma 4.2.** *An irredundant coherent configuration $\mathcal{C}$ is separable if and only if every strict algebraic automorphism of $\mathcal{C}$ is induced by a combinatorial automorphism of $\mathcal{C}$.*

Let $\mathbb{A}(\mathcal{C})$ denote the set of strict algebraic automorphisms of $\mathcal{C}$. Our next task, which will be accomplished by Lemma 4.5 below, is to describe $\mathbb{A}(\mathcal{C})$ for a given irredundant coherent configuration $\mathcal{C}$. Call a permutation f on $\mathcal{C}$ *bound* if f is the identity on each cell, maps each interspace onto itself, and satisfies the condition $f(R^*) = f(R)^*$ for every basis relation R of $\mathcal{C}$. Since the last condition is obeyed by any algebraic isomorphism, every strict algebraic automorphism is bound. If $\mathcal{C}[X, Y] \simeq 2K_{2,2}$, then for a bound permutation f there are two

possibilities. Specifically, suppose that $\mathcal{C}[X, Y]$ partitions $X \times Y$ into two parts R_1 and R_2. We say that f *fixes* $\mathcal{C}[X, Y]$ if $f(R_i) = R_i$ and that f *switches* $\mathcal{C}[X, Y]$ if $f(R_i) = R_{3-i}$ for $i = 1, 2$. Note that, if f switches $\mathcal{C}[X, Y]$, then it switches also $\mathcal{C}[Y, X]$. Given a set S of pairs $\{X, Y\}$ such that $\mathcal{C}[X, Y]$ is non-uniform, let f_S denote the bijection from $\mathcal{C}$ onto itself which switches the interspace $\mathcal{C}[X, Y]$ as well as the interspace $\mathcal{C}[Y, X]$ for each $\{X, Y\} \in S$ and leaves the rest of $\mathcal{C}$ fixed. Thus, every bound permutation of $\mathcal{C}$ coincides with f_S for some S. Conversely, every f_S is a bound permutation, but not all f_S must be algebraic automorphisms.

Thus, we have to describe the class of those S for which f_S is an algebraic automorphism. Note that deciding whether a bound permutation f is a strict algebraic automorphism of $\mathcal{C}$ reduces to locally verifying this on all 3-fiber subconfigurations $\mathcal{C}[X \cup Y \cup Z]$. Therefore, we first consider coherent configurations with three fibers.

We call an irredundant configuration $\mathcal{C}$ *skew-connected* if $\mathcal{C}$ contains no directly connected interspaces.

▶ **Lemma 4.3.** *Let $\mathcal{C}$ be an irredundant coherent configuration with $F(\mathcal{C}) = \{X, Y, Z\}$ and f be a bound permutation of the set of basis relations of $\mathcal{C}$.*
1. *If $\mathcal{C}$ is skew-connected, then f is an algebraic automorphism of $\mathcal{C}$.*
2. *Suppose that $\mathcal{C}$ is not skew-connected. Then f is an algebraic automorphism of $\mathcal{C}$ if and only if f makes exactly two switches of interspaces (switching an interspace and its transpose is counted as a single switch).*

Let $\mathcal{C}$ be an irredundant coherent configuration. Like in the case of reduced Klein configurations [15], we define the *fiber graph* of $\mathcal{C}$, denoted by $F_{\mathcal{C}}$, as follows:
- The vertices of $F_{\mathcal{C}}$ are the fibers of $\mathcal{C}$, i.e., $V(F_{\mathcal{C}}) = F(\mathcal{C})$;
- Two fibers X and Y are adjacent in $F_{\mathcal{C}}$ if the interspace $\mathcal{C}[X, Y]$ is non-uniform.

Suppose that $\mathcal{C}[X, Y]$ is a non-uniform interspace. We define $D(X, Y)$ to be the set of fibers consisting of X, Y, and all Z such that $\mathcal{C}[Z, X]$ is non-uniform and directly connected with $\mathcal{C}[Y, X]$. Let $D_{\mathcal{C}}$ denote the family of all sets $D(X, Y)$ over non-uniform interspaces $\mathcal{C}[X, Y]$. We regard $D_{\mathcal{C}}$ as a hypergraph on $F(\mathcal{C})$ and call it the *hypergraph of direct connections* of $\mathcal{C}$.

The following properties of irredundant configurations are known for reduced Klein configurations [29]; see also [10, Lemma 4.1.18].

▶ **Lemma 4.4.**
1. *Every hyperedge of $D_{\mathcal{C}}$ is a clique in $F_{\mathcal{C}}$, and all interspace connections within this clique are direct.*
2. *Any two hyperedges of $D_{\mathcal{C}}$ have at most one common vertex.*

Proof. Lemma 3.5 implies that, if A and B are two fibers in $D(X, Y)$, then the interspace $\mathcal{C}[A, B]$ is non-uniform and $D(A, B) = D(X, Y)$. This implies both Parts 1 and 2.   ◀

Given $C \in D_{\mathcal{C}}$ and a non-empty $U \subsetneq C$, let $S(U, C)$ be the set of all edges $\{X, Y\}$ in $F_{\mathcal{C}}$ such that $X \in U$ and $Y \in C \setminus U$. Using the notation f_S introduced above, we now define $f_{X,C} = f_{S(\{X\}, C)}$ for $X \in C$.

▶ **Lemma 4.5.** *Suppose that a coherent configuration $\mathcal{C}$ is irredundant.*
1. *$f_S \in \mathbb{A}(\mathcal{C})$ if and only if, for every $C \in D_{\mathcal{C}}$, either the intersection $S \cap \binom{C}{2}$ is empty or it forms a spanning bipartite subgraph of $\binom{C}{2}$, where $\binom{C}{2}$ is considered the complete graph on the vertex set C.*
2. *$\mathbb{A}(\mathcal{C})$ is generated by the set of $f_{X,C}$ for all $C \in D_{\mathcal{C}}$ and all $X \in C$.*

Proof. *1.* For $C \in D_\mathcal{C}$, denote $S[C] = S \cap \binom{C}{2}$. By Lemma 4.4, $\{S[C]\}_{C \in D_\mathcal{C}}$ is a partition of S. Therefore,

$$f_S = \prod_{C \in D_\mathcal{C}} f_{S[C]},$$

where the product is in the group of permutations of $\mathcal{C}$.

($\Longleftarrow$) It suffices to prove that each $f_{S[C]}$ is an algebraic automorphism of $\mathcal{C}$. It is enough to check that, for every triple of fibers X, Y, Z, the restriction of $f_{S[C]}$ to $\mathcal{C}[X \cup Y \cup Z]$ is an algebraic automorphism of $\mathcal{C}[X \cup Y \cup Z]$. If $|\{X, Y, Z\} \cap C| \leq 1$, then $f_{S[C]}$ is the identity on $\mathcal{C}[X \cup Y \cup Z]$. If $|\{X, Y, Z\} \cap C| = 2$, then Lemma 3.5 implies that $\mathcal{C}[X \cup Y \cup Z]$ is either decomposable or skew-connected. The former case is obvious, and in the latter case we are done by Part 1 of Lemma 4.3. If $\{X, Y, Z\} \subseteq C$, then $\mathcal{C}[X \cup Y \cup Z]$ cannot be skew-connected by the definition of $D_\mathcal{C}$ and the bipartiteness of $S[C]$ implies that $f_{S[C]}$ switches either two (up to transposing) or no interspaces between X, Y, Z. In this case we are done by Part 2 of Lemma 4.3.

($\Longrightarrow$) Let $C \in D_\mathcal{C}$ and suppose that $S[C]$ is non-empty. The claim is trivially true if $|C| = 2$, so we assume that $|C| \geq 3$. Let X, Y, and Z be three fibers in C. By assumption, the restriction of f_S to $\mathcal{C}[X \cup Y \cup Z]$ is an algebraic automorphism of $\mathcal{C}[X \cup Y \cup Z]$. By Part 2 of Lemma 4.3, f_S makes either none or exactly two switches in $\mathcal{C}[X \cup Y \cup Z]$. For $S[C]$, seen as a graph on the vertex set C, this implies that $S[C]$ does not contain any induced subgraph isomorphic to K_3 or to $K_2 + K_1$, where the latter is the graph with 3 vertices and 1 edge. A graph is $(K_2 + K_1)$-free if and only if it is complete multipartite. To see this, look at the complement and note that a graph is a vertex-disjoint union of cliques if and only if it does not contain an induced copy of a path on 3 vertices, the complement of $K_2 + K_1$. Thus, $S[C]$ is a complete multipartite graph. Since $S[C]$ is also triangle-free, it is bipartite.

2. Part 1 implies that $\mathbb{A}(\mathcal{C})$ is generated by the set of $f_{S(U,C)}$ for all $C \in D_\mathcal{C}$ and $\emptyset \neq U \subsetneq C$. Note that, if U is split into two non-empty parts U_1 and U_2, then $f_{S(U,C)} = f_{S(U_1,C)} \circ f_{S(U_2,C)}$ (as each interspace between U_1 and U_2 is switched twice). It follows that

$$f_{S(U,C)} = \prod_{X \in U} f_{X,C},$$

which implies the lemma. ◀

Separability test for irredundant coherent configurations

By Lemma 4.2, it suffices to check whether every strict algebraic automorphism $f \in \mathbb{A}(\mathcal{C})$ is induced by a combinatorial automorphism of $\mathcal{C}$. Note that $\mathbb{A}(\mathcal{C})$ forms a group of permutations of the set of basis relations of $\mathcal{C}$. Therefore, it is enough to choose an arbitrary generating set of $\mathbb{A}(\mathcal{C})$ and to check whether every f in this set is induced by a combinatorial automorphism. We use the generating set provided by Part 2 of Lemma 4.5. For deciding whether $f = f_{X,C}$ is induced by a combinatorial automorphism of $\mathcal{C}$, we construct a vertex-colored graph $G = G(\mathcal{C})$ whose automorphism group $\mathrm{Aut}(G)$ consists of all those combinatorial automorphisms of $\mathcal{C}$ which map every basis relation of $\mathcal{C}$ onto itself:

- $V(G) = V(\mathcal{C})$.
- The vertex color classes of G are exactly the fibers of $\mathcal{C}$.
- For two disjoint sets X and Y of vertices of G, let $G[X, Y]$ denote the subgraph of G on the vertex set $X \cup Y$ formed by the edges between a vertex in X and a vertex in Y. For each non-uniform interspace $\mathcal{C}[X, Y]$, we set $G[X, Y]$ to be one of the two $2K_{2,2}$ graphs underlying the basis relations of $\mathcal{C}[X, Y]$.

For each $X \in F(\mathcal{C})$, the subgraph $G[X]$ induced by G on X is defined as follows:

- If there are interspaces $\mathcal{C}[Y, X]$ and $\mathcal{C}[Z, X]$ with askew connection at X, then $G[X]$ is empty (in this case $\mathcal{C}[X] \simeq F_4$ by Lemma 3.4, and each matching relation on X will be anyway preserved by any automorphism of $G(\mathcal{C})$);
- Otherwise, $G[X]$ depends on $\mathcal{C}[X]$. We define $G[X]$ so that $\mathrm{Aut}(G[X])$ consists exactly of the combinatorial automorphisms of $\mathcal{C}[X]$ mapping each basis relation onto itself. Specifically,
  - if $\mathcal{C}[X] \simeq F_4$, then we put a matching $2K_2$ in $G[X]$ different from the one determined by some interspace $\mathcal{C}[Y, X]$ (note that at least one such interspace must exist);
  - if $\mathcal{C}[X] \simeq C_4$, we leave $G[X]$ empty (a matching on X is implicitly determined anyway);
  - if $\mathcal{C}[X] \simeq \vec{C}_4$, we have to put a directed 4-cycle in $G[X]$ coherently with the matching implicitly determined on X. To avoid making $G(\mathcal{C})$ a directed graph, we subdivide each edge of this cycle with two differently colored vertices in the direction given by $\vec{C}_4$. This costs us two new colors and four new vertices of each of these colors (which we put in $V(G)$ in addition to the vertices of $\mathcal{C}$).

For each pair (X, C) where $X \in C \in D_{\mathcal{C}}$, we now have to check whether the algebraic automorphism $f_{X,C}$ is induced by a combinatorial automorphism. A crucial fact is that the number of such pairs is polynomially bounded. Fix $G = G(\mathcal{C})$ as above and obtain a graph $G_{X,C}$ from G by complementing each bipartite subgraph $G[X, Y]$ spanned by the fiber X and a fiber Y in $C \setminus \{X\}$ (i.e., by connecting a vertex in X and a vertex in Y by an edge if and only if they are not adjacent in $G[X, Y]$). By construction, a combinatorial automorphism ϕ of $\mathcal{C}$ induces $f_{X,C}$ exactly when ϕ is an isomorphism of the graphs G and $G_{X,C}$. Thus, $f_{X,C}$ is induced by a combinatorial automorphism if and only if $G \cong G_{X,C}$. The last condition is efficiently verifiable [1] as the graphs G and $G_{X,C}$ are of color multiplicity 4.

5   Putting it together

5.1   Proof of Theorem 1.1

The Cut-Down Lemma (Lemma 3.3) and the preceding analysis of the irredundant case in Section 4 yield the following algorithm for recognizing whether or not a given coherent configuration $\mathcal{C}$ with fibers of size at most 4 is separable.

- Decompose $\mathcal{C}$ in the direct sum of indecomposable subconfigurations $\mathcal{C}_1, \ldots, \mathcal{C}_m$ and treat each $\mathcal{C}_i$ separately. By Lemma 3.1, $\mathcal{C}$ is separable if and only if every $\mathcal{C}_i$ is separable.
- Assume, therefore, that the input configuration $\mathcal{C}$ is indecomposable. If all fibers of $\mathcal{C}$ are of size at most 3, immediately decide that $\mathcal{C}$ is separable (see Lemma 3.2). Otherwise:
  - Remove all fibers of size 2 from $\mathcal{C}$.
  - Remove all pairs of fibers X and Y with $\mathcal{C}[X, Y] \simeq C_8$.
  - As long as $\mathcal{C}$ contains an interspace $\mathcal{C}[X, Y]$ with a matching, remove the fiber X from $\mathcal{C}$.
- If $\mathcal{C}$ becomes decomposable, split it into indecomposable components and handle each of them separately once again in the same way.
- If $\mathcal{C}$ becomes empty, decide that $\mathcal{C}$ is separable.
- Otherwise, we arrive at the case that $\mathcal{C}$ is irredundant and proceed as in Section 4.
- If all computational paths terminate with a positive decision, output "$\mathcal{C}$ *is separable*"; otherwise, output "$\mathcal{C}$ *is non-separable*".

Due to [1], each isomorphism test performed by the algorithm in Section 4 for an irredundant coherent configuration is implementable in $\oplus$L. A list of all subconfigurations to which this step is applied can be generated in logarithmic space [33]. Since $L^{\oplus L} = \oplus L$ (see [8]), the whole algorithm can be implemented in $\oplus$L. Theorem 1.1 is proved.

5.2   Proof of Theorem 1.2

Suppose that the color multiplicity of G is bounded by 4. By Theorem 2.1, G is amenable to 2-WL if and only if its coherent closure $\mathcal{C}(G)$ is separable. Given G with n vertices, the coherent closure $\mathcal{C}(G)$ is computable in time $O(n^3 \log n)$ using the algorithm in [24]. Since G has color multiplicity at most 4, the coherent configuration $\mathcal{C}(G)$ has only fibers with at most 4 points. Therefore, we can decide separability of $\mathcal{C}(G)$ using the algorithm presented in Section 5.1. This algorithm reduces deciding separability for $\mathcal{C}(G)$ to deciding separability for a number of irredundant subconfigurations $\mathcal{C}_1, \ldots, \mathcal{C}_t$ such that

$$\bigcup_{i=1}^{t} F(\mathcal{C}_i) \subseteq F(\mathcal{C}(G)). \tag{4}$$

Producing the list of coherent configurations $\mathcal{C}_1, \ldots, \mathcal{C}_t$ has low time complexity. For each $i \leq t$, we decide separability of $\mathcal{C}_i$ as described in Section 4. Specifically, $\mathcal{C}_i$ is separable if and only if the associated vertex-colored graph G^i is isomorphic to its modified version $H^i = G^i_{X,C}$ for every $X \in F(\mathcal{C}_i)$, where C is the hyperedge of $D_{\mathcal{C}_i}$ containing X. Denote the number of vertices in G^i by n_i. The isomorphism algorithm for graphs of color multiplicity 4 in [1] performs a low-cost conversion of the pair (G^i, H^i) into a system of $M_i < (n_i)^2$ linear equations with $N_i < n_i$ unknowns over the field $\mathbb{Z}_2$ such that $G^i \cong H^i$ if and only if the system is consistent.

Specifically, we here describe a simplified version of this general reduction suitable for any pair (G^i, H^i) arising from $\mathcal{C}_i$. Recall that $V(G^i) = V(H^i) = V(\mathcal{C}_i)$, and the vertex color classes of both G^i and H^i are exactly the fibers $X_1, \ldots, X_s$ of $\mathcal{C}_i$, where each X_j has the same color both in G^i and H^i. For every X_j, we have $G^i[X_j] = H^i[X_j]$. Every non-empty bipartite subgraph $G^i[X_j, X_k]$ is isomorphic to $2K_{2,2}$. Moreover, $H^i[X_j, X_k]$ is equal either to $G^i[X_j, X_k]$ or to its bipartite complement.

Any isomorphism from G^i and H^i maps each vertex color class X_j onto itself. Moreover, if G^i and H^i are isomorphic, then there is an isomorphism ϕ preserving each of the three matchings on X_j for every j (recall that any isomorphism ϕ induces a strict algebraic automorphism of $\mathcal{C}_i$ and, hence, preserves the matchings in each X_j such that $\mathcal{C}_i[X_j] \simeq F_4$ and can be modified to obey this condition for each X_j such that $\mathcal{C}_i[X_j] \simeq C_4$). Denote the restriction of ϕ to X_j by ϕ_j. Thus, ϕ_j is one of the four elements of the *Klein group* $K(X_j)$, where

$$K(X) = \{\mathrm{id}_X, (x_1 x_2)(x_3 x_4), (x_1 x_3)(x_2 x_4), (x_1 x_4)(x_2 x_3)\}$$

for a 4-element set $X = \{x_1, x_2, x_3, x_4\}$. We say that a matching on X is *fixed* by a permutation from $K(X)$ if each of the two matched pairs is mapped onto itself, and we say that it is *flipped* if the matched pairs are mapped onto each other. Denote the matchings on X_j by A_j, B_j, C_j. An element of $K(X_j)$ is uniquely determined by a triple (a_j, b_j, c_j), where $a_j = 1$ if A_j is flipped and $a_j = 0$ if A_j is fixed, and similarly for b_j and c_j. Since a non-identity element of $K(X)$ fixes one matching and flips the other two, we have

$$a_j \oplus b_j \oplus c_j = 0. \tag{E_j}$$
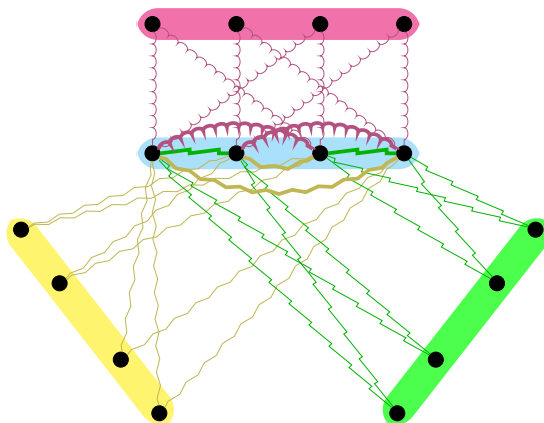
Figure 4 Three pairwise skew-connected interspaces and the matchings they induce.

Another constraint on ϕ_j is imposed by each pair X_j, X_k such that $G^i[X_j, X_k]$ is non-empty. To be specific, suppose that $G^i[X_j, X_k]$ determines the matching A_j in X_j and the matching B_k in X_k. Then

$$a_j \oplus b_k = d_{j,k}, \tag{$E_{j,k}$}$$

where $d_{j,k} = 0$ if $H^i[X_j, X_k]$ is equal to $G^i[X_j, X_k]$ and $d_{j,k} = 1$ if $H^i[X_j, X_k]$ is the bipartite complement of $G^i[X_j, X_k]$. It remains to notice that a set of permutations $\{\phi_j\}_{j=1}^s$ composing an isomorphism from G^i to H^i exists if and only if the system of equations consisting of (E_j) for all $j \leq s$ and $(E_{j,k})$ for all non-empty $G^i[X_j, X_k]$ has a solution.

The rank of an $M \times N$ matrix over a finite field is computable in time $O(MN^{\omega-1})$, where $N \leq M$ (see [7, 23]), or in randomized time $O(N^3 \log N)$ (see [11]). Since $|F(\mathcal{C}_i)| = n_i/4$, we can test separability of $\mathcal{C}_i$ in time $O((n_i)^{2+\omega})$ deterministically or in time $O((n_i)^4 \log n_i)$ using randomization. Taking into account the inequality $\sum_{i=1}^t n_i \leq n$, which follows from (4), and the general inequality $\sum_{i=1}^t (n_i)^\alpha \leq \left(\sum_{i=1}^t n_i\right)^\alpha$ for any real $\alpha \geq 1$, we conclude that separability of $\mathcal{C}(G)$ is decidable in deterministic time $O(n^{2+\omega})$ or in randomized time $O(n^4 \log^2 n)$, where an extra logarithmic factor corresponds to the number of repetitions needed to make the failure probability an arbitrarily small constant.

6 Examples of graphs hard for 2-WL

In Section 4 we constructed a vertex-colored graph $G(\mathcal{C})$ underlying the structure of an irredundant coherent configuration $\mathcal{C}$. As easily seen, if $\mathcal{C}$ contains no $\vec{C}_4$-cells, then $\mathcal{C}$ is the coherent closure of $G(\mathcal{C})$. Theorem 2.1, therefore, implies that every non-separable $\mathcal{C}$ of this kind yields a graph not identifiable by 2-WL.

6.1 The Cai-Fürer-Immerman construction

Skew-connected irredundant coherent configurations correspond to the seminal CFI construction. Indeed, Part 3 of the following theorem is reminiscent of [9, Lemma 6.2]. It shows that, if all connections between non-uniform interspaces are skewed and exactly 3 non-uniform interspaces emanate from each fiber (which is a direct analog of the famous CFI gadget; see Fig. 4), then the coherent configuration is non-separable.

As usually, $\delta(G)$ denotes the minimum degree of a vertex in the graph G. We also use the notation introduced in Section 4.

▶ **Theorem 6.1.** *If $\mathcal{C}$ is skew-connected, then the following is true.*

1. $\mathbb{A}(\mathcal{C}) = \{ f_S : S \subseteq E(F_\mathcal{C}) \}$.
2. *If $\delta(F_\mathcal{C}) \leq 2$, then every f_S is induced by a combinatorial automorphism of $\mathcal{C}$ and, hence, $\mathcal{C}$ is separable.*
3. *If $\delta(F_\mathcal{C}) = 3$, i.e., $F_\mathcal{C}$ is a regular graph of degree 3, then f_S is induced by a combinatorial automorphism of $\mathcal{C}$ exactly when $|S|$ is even. Hence, $\mathcal{C}$ is non-separable in this case.*

6.2   Examples coming from incidence geometry

Lemma 4.4 says exactly that, if $\mathcal{C}$ is irredundant, then the hypergraph of direct connections $D_\mathcal{C}$ is a configuration known in incidence geometry [13, 28] as a partial linear space. Here vertices of the hypergraph are interpreted as *points* and hyperedges as *lines*, even though not every partial linear space admits a geometric realization. More precisely, a hypergraph is called *linear* if every two hyperedges have at most one common vertex. A *partial linear space* is a linear hypergraph with each hyperedge of size at least 2.

A relationship between partial linear spaces and reduced Klein configurations was noticed in [10, Corollary 4.1.19]. The following lemma implies that, if the hypergraph $D_\mathcal{C}$ is 3-regular, that is, every fiber of $\mathcal{C}$ belongs to exactly three cliques in $D_\mathcal{C}$, then the coherent configuration $\mathcal{C}$ is uniquely determined by $D_\mathcal{C}$. Moreover, every partial linear space D where each point belongs to at most 3 lines is the hypergraph of direct connections for some coherent configuration. As a consequence, partial linear spaces are a rich source of templates for constructing coherent configurations.

Specifically, a hypergraph is called *connected* if its Gaifman graph is connected. The degree of a vertex v in a hypergraph H is the number of hyperedges of H containing v. Similarly to graphs, $\Delta(H)$ (resp., $\delta(H)$) denotes the maximum (resp., minimum) degree of a vertex in the hypergraph H. Note that $1 \leq \delta(D_\mathcal{C}) \leq \Delta(D_\mathcal{C}) \leq 3$.

▶ **Lemma 6.2.**

1. *Let $\mathcal{C}$ be an irredundant configuration. If $\mathcal{C} \cong_{alg} \mathcal{C}'$, then $D_\mathcal{C} \cong D_{\mathcal{C}'}$, where $\cong$ denotes isomorphism of hypergraphs.*
2. *Under the condition $\delta(D_\mathcal{C}) \geq 2$, $D_\mathcal{C} \cong D_{\mathcal{C}'}$ implies that $\mathcal{C} \cong_{comb} \mathcal{C}'$.*
3. *For any connected partial linear space D with $\Delta(D) \leq 3$ there is an irredundant configuration $\mathcal{C}$ such that $D_\mathcal{C} \cong D$.*

Proof.

1. This part follows from the fact that an algebraic isomorphism respects fibers, non-uniformity of interspaces, and direct connections of interspaces.
2. Let $h : F(\mathcal{C}) \to F(\mathcal{C}')$ be an isomorphism from the hypergraph $D_\mathcal{C}$ to the hypergraph $D_{\mathcal{C}'}$. Based on h, we define a bijection $\bar{h}$ from the set of all matchings of $\mathcal{C}$ to the set of all matchings of $\mathcal{C}'$. Consider a fiber $X \in F(\mathcal{C})$. Let C_1 and C_2 be two hyperedges of $D_\mathcal{C}$ containing X. All interspaces $\mathcal{C}[Y, X]$ for $Y \in C_1$ determine the same matching in the cell $\mathcal{C}[X]$, which we denote by M_1. All interspaces $\mathcal{C}[Y, X]$ for $Y \in C_2$ determine a matching M_2, different from M_1. Denote the third matching in $\mathcal{C}[X]$ by M_3. Similarly, the interspaces $\mathcal{C}'[Y', h(X)]$ for $Y' \in h(C_1)$ determine a matching M_1', and the interspaces $\mathcal{C}'[Y', h(X)]$ for $Y' \in h(C_2)$ determine a matching $M_2' \neq M_1'$ in $\mathcal{C}'[h(X)]$. Denote the third matching in $\mathcal{C}'[h(X)]$ by M_3' and set $\bar{h}(M_i) = M_i'$ for $i = 1, 2, 3$. Let ψ_X be a bijection from X onto $h(X)$ such that $\phi_X(M) = \bar{h}(M)$ for each matching M in $\mathcal{C}[X]$. Combining all ψ_X over $X \in F(\mathcal{C})$, we obtain a bijection from $V(\mathcal{C})$ onto $V(\mathcal{C}')$ which is a combinatorial isomorphism from $\mathcal{C}$ to $\mathcal{C}'$.
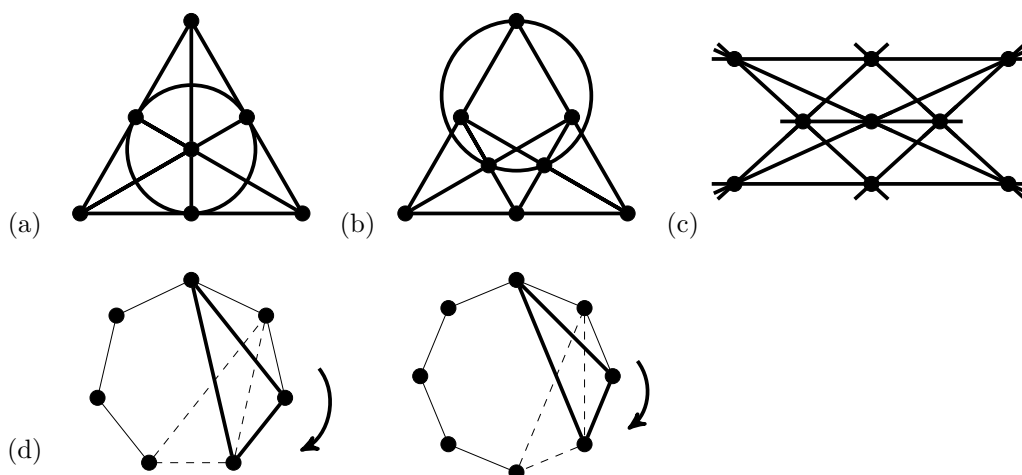
Figure 5 (a) The Fano plane. (b) The Möbius-Kantor configuration. One 3-point "line" in (a) and in (b) is drawn as a circle. (c) The Pappus configuration. (d) Construction of the cyclic versions D_7 and D_8 of the Fano and the Möbius-Kantor configurations.

3. Given D, we construct $\mathcal{C}$ as follows. Each point p of D gives rise to a 4-point fiber X_p in $\mathcal{C}$, with the cell $\mathcal{C}[X_p]$ being of type F_4. With each hyperedge C of D containing p, we associate a matching relation $M_{p,C}$ in $\mathcal{C}[X_p]$ such that $M_{p,C} \neq M_{p,C'}$ if $C \neq C'$. For each pair of points p and q in the same hyperedge C, we make the interspace $\mathcal{C}[X_p, X_q]$ non-uniform so that it determines the matching $M_{p,C}$ in $\mathcal{C}[X_p]$ and the matching $M_{q,C}$ in $\mathcal{C}[X_q]$.   ◄

We now give examples of non-amenable graphs (or, equivalently, examples of non-separable irredundant coherent configurations) arising from classical incidence geometries. Partial linear spaces with n points where every line contains exactly 3 points and every point is incident to exactly 3 lines are known as (n_3)-*configurations*; see [20, 32]. There is no (n_3)-configuration for $n \leq 6$. There are a unique (7_3)-configuration, namely the *Fano plane*, and a unique (8_3)-configuration, namely the *Möbius-Kantor configuration*; see Figure 5. We denote the coherent configurations whose hypergraphs of direct connections are isomorphic to these two line-point configurations by $\mathcal{C}_{\mathrm{Fano}}$ and $\mathcal{C}_{\mathrm{MK}}$ respectively. These configurations exist by Part 3 of Lemma 6.2 and are unique by Part 2 of this lemma.

▶ **Theorem 6.3.** $\mathcal{C}_{\mathrm{Fano}}$ *is non-separable, and* $\mathcal{C}_{\mathrm{MK}}$ *is separable.*

Thus, the 7-point Fano plane gives rise to a graph of color multiplicity 4 with 28 vertices which is not identifiable by 2-WL and is not a CFI graph. Theorem 6.3 is actually a particular instance of a more general statement.

Let $n \geq 7$. The *cyclic (n_3)-configurations D_n* is constructed as follows [20, Section 2.1]. Let F_n be the Cayley graph of $\mathbb{Z}_n$ with the difference set $\{\pm 1, \pm 2, \pm 3\}$ and D_n be the hypergraph formed by 3-cliques $\{i, i+2, i+3\}$ in F_n, where $i \in \mathbb{Z}_n$. It is straightforward to see that D_n is really an (n_3)-configuration. By the uniqueness of (n_3)-configurations for $n = 7, 8$ (see, e.g., [32, Theorem 5.13]), the Fano plane is isomorphic, as a hypergraph, to D_7, and the Möbius-Kantor configuration is isomorphic to D_8. Let $\mathcal{C}_n$ be the coherent configuration constructed from D_n as in the proof of Part 3 of Lemma 6.2. This lemma implies that $\mathcal{C}_{\mathrm{Fano}} \cong_{comb} \mathcal{C}_7$ and $\mathcal{C}_{\mathrm{MK}} \cong_{comb} \mathcal{C}_8$. Thus, Theorem 6.3 is equivalent to the statement that $\mathcal{C}_n$ is non-separable if $n = 7$ and separable if $n = 8$.

▶ **Theorem 6.4.** *Let $n \geq 7$. The coherent configuration $\mathcal{C}_n$ is non-separable if and only if n is a multiple of 7.*

▶ Remark 6.5. There are exactly three (9_3)-configurations [20, 32]. The most famous of them is the *Pappus configuration* shown in Fig. 5(c). Computer-assisted verification shows that the corresponding 36-point coherent configuration is non-separable. Of the other two (9_3)-configurations, one is the cyclic (9_3)-configuration defined above, and the other is obtained similarly by rotating the triangle $\{0, 3, 4\}$ (instead of $\{0, 2, 3\}$) in $\mathbb{Z}_9$. These two produce separable coherent configurations.

▶ Remark 6.6. Curiously, Lemma 6.2 reveals a connection between irredundant coherent configurations and the multipede graphs introduced by Neuen and Schweitzer in [30]. Let $\mathcal{C}$ be an irredundant configuration and assume for the hypergraph of direct connections of $\mathcal{C}$ that $\delta(D_{\mathcal{C}}) = 3$. Consistently with the notation in [30], denote the incidence graph of the hypergraph $D_{\mathcal{C}}$ by $G = G(V, W)$, where $V = F(\mathcal{C})$ is the vertex set of $D_{\mathcal{C}}$, i.e., the set of all fibers of $\mathcal{C}$, and W is the set of the hyperedges of $D_{\mathcal{C}}$, i.e., the cliques of directly connected fibers. Two vertices $v \in V$ and $w \in W$ are adjacent in G if v belongs to w. Thus, every vertex in V has degree 3 in G. Any such bipartite graph G determines a multipede graph denoted in [30] by $R(G)$. This is a vertex-colored graph with vertex classes of size 4 and 2. Since we started from an irredundant configuration $\mathcal{C}$, the coloring of $R(G)$ is not refinable by 2-WL, and each color class of $R(G)$ stays as a fiber in the coherent closure $\mathcal{C}(R(G))$. Let $\mathcal{C}'$ be the coherent configuration obtained from $\mathcal{C}(R(G))$ by cutting down all fibers of size 2 (cf. Part 2 of Lemma 3.3). Lemma 6.2 implies that $\mathcal{C}'$ is combinatorially isomorphic to $\mathcal{C}$.

6.3   Small graphs

Our amenability criteria behind Theorem 1.2 are rather practical, which is illustrated by the following result.

▶ **Theorem 6.7.**
1. *All graphs of color multiplicity 4 with at most 15 vertices are amenable.*
2. *Up to isomorphism and color renaming, there are 434 non-amenable graphs of color multiplicity 4 with 16 vertices. More precisely, the number of non-trivial $\equiv_{2\text{-WL}}$-equivalence classes is 217, each consisting of exactly two non-isomorphic graphs.*

7   Conclusion and further questions

Our results raise questions about the parameterized complexity of recognizing the amenability of a given graph with the largest color multiplicity m taken as the parameter. The problem is trivial for $m = 3$ due to [24]. We show that it is solvable in polynomial time for $m = 4$. Our analysis surely generalizes to a few subsequent values of m. For any fixed m, the problem is in coNP, and it is open whether it is in P if m is large.

Another open question, that naturally arises in light of Theorem 1.2, concerns the next dimension of the Weisfeiler-Leman algorithm: Can the amenability to 3-WL be decided in polynomial time on input graphs with the largest color multiplicity 4?

The *WL dimension* of a graph G is defined as the minimum k such that G is amenable to k-WL. The graphs with large WL dimension are of significant interest in the study of the graph isomorphism problem. When we seek such graphs among graphs with color multiplicity 4, note that they must be at least non-amenable to 2-WL. Cai, Fürer, and Immerman [9] give conditions ensuring linear WL dimension for graphs whose coherent closure is, in our

terminology, skew-connected. Further such conditions are identified by the line of research [12, 21, 30, 31]. Can we achieve high WL dimension in other cases, say, for graphs whose coherent closure corresponds to a line-point (n_3)-configuration (see Section 6)?

---- **References** ----

1   Vikraman Arvind and Johannes Köbler. On hypergraph and graph isomorphism with bounded color classes. In *23rd Annual Symposium on Theoretical Aspects of Computer Science (STACS'06)*, volume 3884 of *Lecture Notes in Computer Science*, pages 384–395. Springer, 2006. `doi:10.1007/11672142_31`.

2   Vikraman Arvind, Johannes Köbler, Gaurav Rattan, and Oleg Verbitsky. Graph isomorphism, color refinement, and compactness. *Computational Complexity*, 26(3):627–685, 2017. `doi:10.1007/s00037-016-0147-6`.

3   Vikraman Arvind, Piyush P. Kurur, and T. C. Vijayaraghavan. Bounded color multiplicity graph isomorphism is in the #L hierarchy. In *20th Annual IEEE Conference on Computational Complexity (CCC'05)*, pages 13–27. IEEE Computer Society, 2005. `doi:10.1109/CCC.2005.7`.

4   László Babai. Graph isomorphism in quasipolynomial time. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC'16)*, pages 684–697, 2016. `doi:10.1145/2897518.2897542`.

5   Béla Bollobás. Distinguishing vertices of random graphs. *Annals of Discrete Mathematics*, 13:33–49, 1982. `doi:10.1016/S0304-0208(08)73545-X`.

6   R. C. Bose and Dale M. Mesner. On linear associative algebras corresponding to association schemes of partially balanced designs. *Ann. Math. Statist.*, 30:21–38, 1959. `doi:10.1214/aoms/1177706356`.

7   James R. Bunch and John E. Hopcroft. Triangular factorization and inversion by fast matrix multiplication. *Math. Comp.*, 28:231–236, 1974. `doi:10.2307/2005828`.

8   Gerhard Buntrock, Carsten Damm, Ulrich Hertrampf, and Christoph Meinel. Structure and importance of logspace-mod classes. *Mathematical systems theory*, 25(3):223–237, 1992. `doi:10.1007/BF01374526`.

9   Jin-yi Cai, Martin Fürer, and Neil Immerman. An optimal lower bound on the number of variables for graph identifications. *Combinatorica*, 12(4):389–410, 1992. `doi:10.1007/BF01305232`.

10 Gang Chen and Ilia Ponomarenko. *Coherent configurations*. Central China Normal University Press, 2019.

11 Ho Yee Cheung, Tsz Chiu Kwok, and Lap Chi Lau. Fast matrix rank algorithms and applications. *J. ACM*, 60(5):31:1–31:25, 2013. `doi:10.1145/2528404`.

12 Anuj Dawar and Kashif Khan. Constructing hard examples for graph isomorphism. *J. Graph Algorithms Appl.*, 23(2):293–316, 2019. `doi:10.7155/jgaa.00492`.

13 Bart de Bruyn. *An introduction to incidence geometry*. Front. Math. Basel: Birkhäuser/Springer, 2016.

14 S. Evdokimov and I. Ponomarenko. Characterization of cyclotomic schemes and normal schur rings over a cyclic group. *St. Petersburg Math. J.*, 14(2):189–221, 2003.

15 Sergei Evdokimov and Ilia Ponomarenko. On highly closed cellular algebras and highly closed isomorphisms. *Electr. J. Comb.*, 6, 1999. URL: `http://www.combinatorics.org/Volume_6/Abstracts/v6i1r18.html`.

16 Sergei Evdokimov, Ilia Ponomarenko, and Gottfried Tinhofer. Forestal algebras and algebraic forests (on a new class of weakly compact graphs). *Discrete Mathematics*, 225(1-3):149–172, 2000. `doi:10.1016/S0012-365X(00)00152-7`.

17 Frank Fuhlbrück, Johannes Köbler, and Oleg Verbitsky. Identifiability of graphs with small color classes by the Weisfeiler-Leman algorithm. Technical report, `https://arxiv.org/abs/1907.02892`, 2019.

18 Martin Fürer. On the combinatorial power of the Weisfeiler-Lehman algorithm. In *Algorithms and Complexity — 10th International Conference (CIAC'17) Proceedings*, volume 10236 of *Lecture Notes in Computer Science*, pages 260–271, 2017. `doi:10.1007/978-3-319-57586-5_22`.

19 François Le Gall. Powers of tensors and fast matrix multiplication. In *Proc. of the Int. Symposium on Symbolic and Algebraic Computation (ISSAC'14)*, pages 296–303. ACM, 2014. `doi:10.1145/2608628.2608664`.

20 Branko Grünbaum. *Configurations of points and lines*, volume 103 of *Grad. Stud. Math.* Providence, RI: American Mathematical Society (AMS), 2009.

21 Yuri Gurevich and Saharon Shelah. On finite rigid structures. *J. Symb. Log.*, 61(2):549–562, 1996. `doi:10.2307/2275675`.

22 D.G. Higman. Finite permutation groups of rank 3. *Math. Z.*, 86:145–156, 1964. `doi:10.1007/BF01111335`.

23 Oscar H. Ibarra, Shlomo Moran, and Roger Hui. A generalization of the fast LUP matrix decomposition algorithm and applications. *J. Algorithms*, 3(1):45–56, 1982. `doi:10.1016/0196-6774(82)90007-4`.

24 N. Immerman and E. Lander. Describing graphs: A first-order approach to graph canonization. In *Complexity Theory Retrospective*, pages 59–81. Springer, 1990.

25 Neil Immerman. *Descriptive complexity*. Graduate texts in computer science. Springer, 1999. `doi:10.1007/978-1-4612-0539-5`.

26 Sandra Kiefer, Pascal Schweitzer, and Erkal Selman. Graphs identified by logics with counting. In *Proceedings of the 40th International Symposium on Mathematical Foundations of Computer Science (MFCS'15)*, volume 9234 of *Lecture Notes in Computer Science*, pages 319–330. Springer, 2015. `doi:10.1007/978-3-662-48057-1_25`.

27 Eugene M. Luks. Parallel algorithms for permutation groups and graph isomorphism. In *27th Annual Symposium on Foundations of Computer Science (FOCS'86)*, pages 292–302. IEEE Computer Society, 1986. `doi:10.1109/SFCS.1986.39`.

28 Klaus Metsch. *Linear spaces with few lines*, volume 1490 of *Lect. Notes Math.* Berlin etc.: Springer-Verlag, 1991.

29 Mikhail Muzychuk and Ilya Ponomarenko. On quasi-thin association schemes. *J. Algebra*, 351(1):467–489, 2012.

30 Daniel Neuen and Pascal Schweitzer. Benchmark graphs for practical graph isomorphism. In *25th Annual European Symposium on Algorithms (ESA'17)*, volume 87 of *LIPIcs*, pages 60:1–60:14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2017. `doi:10.4230/LIPIcs.ESA.2017.60`.

31 Daniel Neuen and Pascal Schweitzer. An exponential lower bound for individualization-refinement algorithms for graph isomorphism. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing (STOC'18)*, pages 138–150. ACM, 2018. `doi:10.1145/3188745.3188900`.

32 Tomaž Pisanski and Brigitte Servatius. *Configurations from a graphical viewpoint.* Birkhäuser Adv. Texts, Basler Lehrbüch. New York, NY: Birkhäuser, 2013.

33 Omer Reingold. Undirected connectivity in log-space. *J. ACM*, 55(4):17:1–17:24, 2008. `doi:10.1145/1391289.1391291`.

34 B.Yu. Weisfeiler and A.A. Leman. The reduction of a graph to canonical form and the algebra which appears therein. *NTI, Ser. 2*, 9:12–16, 1968. English translation is available at `https://www.iti.zcu.cz/wl2018/pdf/wl_paper_translation.pdf`.

35 Carme Álvarez and Birgit Jenner. A very hard log-space counting class. *Theoretical Computer Science*, 107(1):3–30, 1993. `doi:10.1016/0304-3975(93)90252-O`.