Electronic Colloquium on Computational Complexity, Report No. 32 (2015)

Graph Isomorphism, Color Refinement, and Compactness*

V. Arvind[1], Johannes Köbler[2], Gaurav Rattan[1], Oleg Verbitsky[2,3]

[1] The Institute of Mathematical Sciences, Chennai 600 113, India {arvind,grattan}@imsc.res.in
[2] Institut für Informatik, Humboldt Universität zu Berlin, Germany
{koebler,verbitsk}@informatik.hu-berlin.de
[3] On leave from the Institute for Applied Problems of Mechanics and Mathematics, Lviv, Ukraine.

Abstract. *Color refinement* is a classical technique used to show that two given graphs G and H are non-isomorphic; it is very efficient, although it does not succeed on all graphs. We call a graph G *amenable* to color refinement if the color-refinement procedure succeeds in distinguishing G from any non-isomorphic graph H. Babai, Erdős, and Selkow (1982) have shown that random graphs are amenable with high probability. Our main results are the following:

- We determine the exact range of applicability of color refinement by showing that the class of amenable graphs is recognizable in time $O((n+m)\log n)$, where n and m denote the number of vertices and the number of edges in the input graph.
- Furthermore, we prove that amenable graphs are compact in the sense of Tinhofer (1991). That is, their polytopes of fractional automorphisms are integral. The concept of compactness was introduced in order to identify the class of graphs G for which isomorphism $G \cong H$ can be decided by computing an extreme point of the polytope of fractional isomorphisms from G to H and checking if this point is integral. Our result implies that the applicability range for this linear programming approach to isomorphism testing is at least as large as for the combinatorial approach based on color refinement.

1 Introduction

The well-known *color refinement* (also known as *naive vertex classification*) procedure for Graph Isomorphism works as follows: it begins with a uniform coloring of the vertices of two graphs G and H and refines the vertex coloring step by step. In a refinement step, if two vertices have identical colors but differently colored neighborhoods (with the multiplicities of colors counted), then these vertices get new different colors. The procedure terminates when no further refinement of the vertex color classes is possible. Upon termination, if the multisets of vertex colors in G and H are different, we can correctly conclude that they are not isomorphic. However, color refinement sometimes fails to distinguish non-isomorphic graphs. The simplest example is given by any two non-isomorphic regular graphs of the same degree with the same number of vertices.

For which pairs of graphs G and H does the color refinement procedure succeed in solving Graph Isomorphism? Mainly this question has motivated the study of color refinement from different perspectives.

Immerman and Lander [13], in their highly influential paper, established a close connection between color refinement and 2-variable first-order logic with counting quantifiers. They

* This work was supported by the Alexander von Humboldt Foundation in its research group linkage program. The fourth author was supported by DFG grant VE 652/1.

ISSN 1433-8092

show that color refinement distinguishes G and H if and only these graphs are distinguishable by a sentence in this logic.

A well-known approach to tackling intractable optimization problems is to consider an appropriate linear programming relaxation. A similar approach to isomorphism testing, based on the notion of fractional isomorphisms (see Section 3), turns out to be equivalent to color refinement. Building on Tinhofer's work [21], it is shown by Ramana, Scheinerman and Ullman [18] (see also Godsil [10]) that two graphs are indistinguishable by color refinement if and only if they are fractionally isomorphic.

We say that color refinement *applies* to a graph G if it succeeds in distinguishing G from any non-isomorphic H. A graph to which color refinement applies is called *amenable*. There are interesting classes of amenable graphs:

1. An obvious class of graphs to which color refinement is applicable is the class of *unigraphs*. Unigraphs are graphs that are determined up to isomorphism by their degree sequences; see, e.g., [4, 25].
2. Trees are amenable (Edmonds [5, 26]).
3. It is easy to see that all graphs for which the color refinement procedure terminates with all singleton color classes (i.e. the color classes form the discrete partition) are amenable. We call this graph class $\mathsf{Discrete}$. Babai, Erdös, and Selkow [2] have shown that a random graph $G_{n,1/2}$ is in $\mathsf{Discrete}$ with high probability.

Our results

What is the class of graphs to which color refinement applies? The logical and linear programming based characterizations of color refinement do not provide any efficient criterion answering this question.

We aim at determining the exact range of applicability of color refinement. We find an efficient characterization of the entire class of amenable graphs, which allows for a quasilinear-time test whether or not color refinement applies to a given graph.

Corollary 12. *The class of amenable graphs is recognizable in time $O((n+m)\log n)$, where n and m denote the number of vertices and edges of the input graph.*

This result is shown in Section 2, where we unravel the structure of amenable graphs. We note that a weak *a priori* upper bound for the complexity of recognizing amenable graphs is $\mathsf{coNP}^{\mathsf{GI}[1]}$, where the superscript means the one-query access to an oracle solving the graph isomorphism problem. To the best of our knowledge, no better upper bound was known before.

Combined with the Immerman-Lander result [13] mentioned above, it follows that the class of graphs definable by first-order sentences with 2 variables and counting quantifiers is recognizable in polynomial time. This result naturally generalizes to structures over any binary relational vocabulary (see Corollary 13).

Compact graphs The following linear programming approach to isomorphism testing was suggested by Tinhofer in [23]. A graph is called *compact* if the polytope of all its fractional automorphisms is integral. That is, all extreme points of this polytope have integer coordinates. If G is compact then it has the following remarkable property: If $G \cong H$, then the polytope of all fractional *isomorphisms* from G to H is also integral while if $G \ncong H$, then this polytope has *no* integer extreme point (in particular, it can be empty). It follows that for compact graphs G, the relation $G \cong H$ can be checked in polynomial time by using a polynomial-time linear programming algorithm to compute some extreme point of the polytope and testing if it is integral. Our second main result in the paper shows that the class of compact graphs contains all amenable graphs.

Theorem 17. *All amenable graphs are compact.*

This result implies that Tinhofer's approach to Graph Isomorphism has at least as large an applicability range as color refinement. It remains an intriguing open problem to find an efficient characterization for the class of compact graphs (or to show that its decision problem is hard).

Finally, we consider the class of graphs for which color refinement succeeds in solving the *Graph Automorphism* problem: namely, the problem of checking if an input graph G has a nontrivial automorphism. Specifically, we call a graph G *refinable* if the color partition produced by color refinement coincides with the orbit partition of the automorphism group of G. The fact that all trees are refinable was observed independently by several authors; see a survey in [24]. Our results imply that all amenable graphs are refinable. In Section 5 we discuss structural and algorithmic graph properties that were introduced by Tinhofer [23] and Godsil [10]. We note that these properties, along with compactness, define a hierarchy of graph classes between the amenable and the refinable graphs.

Related work. Color refinement turns out to be a useful tool not only in isomorphism testing but also in a number of other areas; see [11, 15, 20] and references there. The concept of compactness is generalized to *weak compactness* in [8, 9]. The linear programming approach of [21, 18] to isomorphism testing is extended in [1, 12], where it is shown that this extension corresponds to the k-dimensional Weisfeiler-Lehman algorithm (which is just color refinement if $k = 1$).

Notation. The vertex set of a graph G is denoted by $V(G)$. The vertices adjacent to a vertex $u \in V(G)$ form its neighborhood $N(u)$. A set of vertices $X \subseteq V(G)$ induces a subgraph of G, that is denoted by $G[X]$. For two disjoint sets X and Y, $G[X, Y]$ is the bipartite graph with vertex classes X and Y formed by all edges of G connecting a vertex in X with a vertex in Y. The vertex-disjoint union of graphs G and H will be denoted by $G + H$. Furthermore, we write mG for the disjoint union of m copies of G. The complement of a graph G is denoted by $\overline{G}$. The *bipartite complement* of a bipartite graph G with vertex classes X and Y is the bipartite graph G' with the same vertex classes such that $\{x, y\}$ with $x \in X$ and $y \in Y$ is an edge in G' iff it is not an edge in G. We use the standard notation K_n for the complete graph on n vertices, $K_{s,t}$ for the complete bipartite graph whose vertex classes have s and t vertices, and C_n for the cycle on n vertices.

2 Amenable graphs

2.1 Basic definitions and facts

Given a graph G, the *color refinement* algorithm (to be abbreviated as *CR*) iteratively computes a sequence of colorings C^i of $V(G)$. The initial coloring C^0 is uniform. Then,

$$C^{i+1}(u) = \{\!\{ C^i(a) : a \in N(u) \}\!\}, \tag{1}$$

where $\{\!\{ \ldots \}\!\}$ denotes a multiset. Note that $C^1(u) = C^1(v)$ iff the two vertices have the same degree.

A simple inductive argument shows that $C^{i+1}(u) = C^{i+1}(v)$ implies $C^i(u) = C^i(v)$. Therefore, the partition $\mathcal{P}^{i+1}$ of $V(G)$ into the color classes of C^{i+1} is a refinement of the partition $\mathcal{P}^i$ corresponding to C^i. It follows that, eventually, $\mathcal{P}^{s+1} = \mathcal{P}^s$ for some s; hence, $\mathcal{P}^i = \mathcal{P}^s$ for all $i \geq s$. The partition $\mathcal{P}^s$ is the *stable* partition of G.

A partition $\mathcal{P}$ of $V(G)$ is called *equitable* if:

(i) For any $X \in \mathcal{P}$ the graph $G[X]$ induced by X is *regular*, that is, all vertices in $G[X]$ have equal degrees.
(ii) For any $X, Y \in \mathcal{P}$ the bipartite graph $G[X, Y]$ induced by X and Y is *biregular*, that is, all vertices in each vertex class have equal degrees: The vertices in X have equally many neighbors in Y and vice versa.

Given an equitable partition $\mathcal{P}$ of a graph G, we call its elements *cells*.

A trivial example of an equitable partition is the partition of $V(G)$ into singletons, which we call *discrete*. There is a unique equitable partition $\mathcal{P}_G$ that is the *coarsest* in the sense that any other equitable partition $\mathcal{P}$ of G is a subpartition of $\mathcal{P}_G$. It is easy to see that the stable partition of G is equitable, and an inductive argument shows that it is actually the coarsest [6, Lemma 1].

A straightforward inductive argument shows that the colorings C^i are preserved under isomorphisms.

Lemma 1. *If ϕ is an isomorphism from G to H, then $C^i(u) = C^i(\phi(u))$ for any vertex u of G.*

Lemma 1 readily implies that, if graphs G and H are isomorphic, then

$$\{\!\{ C^i(u) : u \in V(G) \}\!\} = \{\!\{ C^i(v) : v \in V(H) \}\!\} \tag{2}$$

for all $i \geq 0$. When used for isomorphism testing, the CR algorithm accepts two graphs G and H as isomorphic exactly when the above condition is met on input $G + H$. Note that this condition is actually finitary: If Equality (2) is false for some i, it must be false for some $i < 2n$, where n denotes the number of vertices in each of the graphs. This follows from the observation that the partition $\mathcal{P}^{2n-1}$ induced by the coloring C^{2n-1} must be a stable partition of the disjoint union of G and H. In fact, Equality (2) holds true for all i iff it is true for $i = n$; see [17]. Thus, it is enough that CR verifies (2) for $i = n$.

Note that computing the vertex colors literally according to (1) would lead to an exponential growth of the lengths of color names. This can be avoided by renaming the colors after each refinement step. Then CR never needs more than n color names (appearance of more than n colors is an indication that the graphs are non-isomorphic).

Definition 2. *We call a graph G* amenable *if CR works correctly on the input G, H for every H, that is, Equality (2) is false for $i = n$ whenever $H \not\cong G$. The class of all amenable graphs is denoted* $\mathsf{Amenable}$.

2.2 Local structure of amenable graphs

Consider the stable (i.e., the coarsest equitable) partition $\mathcal{P}_G$ of an amenable graph G. For different cells $X, Y \in \mathcal{P}$, we will analyze the possible regular graphs $G[X]$ and biregular graphs $G[X, Y]$ that can occur. The following lemma gives a list of all possible regular and biregular graphs that can occur in the stable (i.e., the coarsest equitable) partition of an amenable graph.

Lemma 3. *Let $\mathcal{P}_G$ be the stable partition of an amenable graph G.*

(A) *If $X \in \mathcal{P}_G$, then $G[X]$ is an empty graph, a complete graph, a matching graph mK_2, the complement of a matching graph, or the 5-cycle;*
(B) *If $X, Y \in \mathcal{P}_G$, then $G[X, Y]$ is an empty graph, a complete bipartite graph, the disjoint union of stars $sK_{1,t}$ where X and Y are the set of s central vertices and the set of st leaves, or the bipartite complement of the last graph.*

The proof of Lemma 3 is based on the following facts.

Lemma 4 (Johnson [14]). *A regular graph of degree d with n vertices is a unigraph if and only if $d \in \{0, 1, n-2, n-1\}$ or $d = 2$ and $n = 5$.*[4]

Lemma 5 (Koren [16]). *A bipartite graph is determined up to isomorphism by the conditions that every of the m vertices in one part has degree c and every of the n vertices in the other part has degree d if and only if $c \in \{0, 1, n-1, n\}$ or $d \in \{0, 1, m-1, m\}$.*

Lemma 4 and 5 show that, if G contains a subgraph $G[X]$ or $G[X, Y]$ that is induced by some $X, Y \in \mathcal{P}_G$ but not listed in Lemma 3, then this subgraph can be replaced by a non-isomorphic regular or biregular graph with the same parameters. In order to prove Lemma 3, it now suffices to show that the resulting graph H is indistinguishable from G by color refinement. The graphs G and H in the following lemma have the same vertex set. Given a vertex u, we distinguish its neighborhoods $N_G(u)$ and $N_H(u)$ and its colors $C_G^i(u)$ and $C_H^i(u)$ in the two graphs.

Lemma 6. *Suppose that X and Y are cells of the stable partition of a graph G.*

[4] The last case, in which the graph is the 5-cycle, is missing from the statement of this result in [14, Theorem 2.12]. The proof in [14] tacitly considers only graphs with at least 6 vertices.

(i) Let H be obtained from G by changing the edges between the vertices in X so that $G[X]$ is replaced with a regular graph of the same degree. Then $C^i_G(u) = C^i_H(u)$ for any $u \in V(G)$ and any i.

(ii) Let H be obtained from G by changing the edges between X and Y so that $G[X, Y]$ is replaced with a biregular graph having the same vertex degrees. Then $C^i_G(u) = C^i_H(u)$ for any $u \in V(G)$ and any i.

Proof. We proceed by induction on i. In the base case of $i = 0$ the claim is trivially true. Assume that $C^i_G(a) = C^i_H(a)$ for all $a \in V(G)$. Consider an arbitrary vertex u and prove that

$$C^{i+1}_G(u) = C^{i+1}_H(u). \tag{3}$$

From now on we treat parts (i) and (ii) separately.

(i) Suppose first that $u \notin X$. Since the transformation of G into H does not affect the edges emanating from u, we have $N_G(u) = N_H(u)$. Looking at the definition (1), we immediately derive (3) from the induction assumption.
The case of $u \in X$ is a bit more complicated. Now we have only equality $N_G(u) \setminus X = N_H(u) \setminus X$, which implies

$$\{\!\!\{\, C^i_G(a) : a \in N_G(u) \setminus X \}\!\!\} = \{\!\!\{\, C^i_H(a) : a \in N_H(u) \setminus X \}\!\!\}. \tag{4}$$

The equality $N_G(u) \cap X = N_H(u) \cap X$ is not necessarily true. However, u has equally many neighbors from X in G and in H. Furthermore, for any two vertices a and a' in X we have $C^i_G(a) = C^i_G(a')$ because X is a cell of G, and $C^i_H(a) = C^i_G(a) = C^i_G(a') = C^i_H(a')$ by the induction assumption. That is, all vertices in X have the same C^i-color both in G and in H. It follows that

$$\{\!\!\{\, C^i_G(a) : a \in N_G(u) \cap X \}\!\!\} = \{\!\!\{\, C^i_H(a) : a \in N_H(u) \cap X \}\!\!\}. \tag{5}$$

Combining (4) and (5), we conclude that (3) holds in any case.

(ii) If $u \notin X \cup Y$, we have $N_G(u) = N_H(u)$ and Equality (3) readily follows from the induction assumption.
Suppose that $u \in Y$. In this case we still have (4) and, exactly as in part (i), we also derive (5). Equality (3) follows.
The case of $u \in X$ is symmetric. □

Proof of Lemma 3.

(A) If $G[X]$ is a graph not from the list, by Lemma 4, it is not a unigraph. We, therefore, can modify G locally on X by replacing $G[X]$ with a non-isomorphic regular graph with the same parameters. Part (i) of Lemma 6 implies that the resulting graph H satisfies Equality (2) for any i, that is, CR does not distinguishes between G and H. The graphs G and H are non-isomorphic because, by part (i) of Lemma 6 and by Lemma 1, an isomorphism from G to H would induce an isomorphism from $G[X]$ to $H[X]$. This shows that G is not amenable.

(B) This part follows, similarly to Condition **A**, from Lemma 5 and part (ii) of 6. □

2.3 Global structure of amenable graphs

Recall that $\mathcal{P}_G$ is the stable partition of the vertex set of a graph G, and that elements of $\mathcal{P}_G$ are called cells. We define the auxiliary *cell graph* $C(G)$ of an amenable graph G to be the complete graph on the vertex set $\mathcal{P}_G$ with the following labeling of vertices and edges. A vertex $X \in \mathcal{P}_G$ is labeled either *complete*, *empty*, *matching*, *co-matching*, or *pentagonal* depending on the type of $G[X]$, and an edge $\{X, Y\}$ is labeled either *complete*, *empty*, *constellation*, or *co-constellation*, depending on the type of $G[X, Y]$; see Conditions **A** and **B** in Lemma 3.

Each vertex and edge has exactly one label. This requires removing ambiguities in several cases. First, every singleton cell $X = \{u\}$ is labeled as *empty* (rather than *complete*). Note that the complete graph K_2 is as well a matching graph. This is resolved by labeling each two-element cell $X = \{u, v\}$ as *complete* or *empty* (rather than *matching* or *co-matching*) depending on whether or not u and v are adjacent in G. Thus, a matching or co-matching X always consists of at least 4 vertices. Furthermore, the edges $\{X, Y\}$ of the cell graph such that $G[X, Y] \cong K_{1,t}$ are labeled as *complete* (rather than *constellation*). If $G[X, Y]$ is the bipartite complement of $K_{1,t}$, then $\{X, Y\}$ is labeled as *empty* (rather than *co-constellation*). Another source of ambiguity is that $2K_{1,t}$ is isomorphic to its bipartite complement. The edges $\{X, Y\}$ corresponding to $G[X, Y] \cong 2K_{1,t}$ are labeled as *constellation* (rather than *co-constellation*).

A vertex of the cell graph is called *homogeneous* if it is labeled as *complete* or *empty* and *heterogeneous* in any of the other three cases. An edge of the cell graph is called *isotropic* if it is labeled as *complete* or *empty* and *anisotropic* if it is labeled as *constellation* or *co-constellation*.

A path $X_1X_2 \ldots X_l$ in $C(G)$ where every edge $\{X_i, X_{i+1}\}$ is anisotropic will be referred to as an *anisotropic path*. If also $\{X_l, X_1\}$ is an anisotropic edge, we speak of an *anisotropic cycle*. In the case that $|X_1| = |X_2| = \ldots = |X_l|$, such a path (or cycle) will be called *uniform*. Note that if an edge $\{X_i, X_{i+1}\}$ of a uniform path/cycle is *constellation* (resp. *co-constellation*), then $G[X_i, X_{i+1}]$ is a matching (resp. co-matching) graph.

Lemma 7. *If G is amenable, then*

- ***(C)*** *the cell graph $C(G)$ contains no uniform anisotropic path connecting two heterogeneous vertices;*
- ***(D)*** *the cell graph $C(G)$ contains no uniform anisotropic cycle;*
- ***(E)*** *the cell graph $C(G)$ contains neither an anisotropic path $XY_1 \ldots Y_lZ$ such that $|X| < |Y_1| = \ldots = |Y_l| > |Z|$ nor an anistropic cycle $XY_1 \ldots Y_lX$ such that $|X| < |Y_1| = \ldots = |Y_l|$;*
- ***(F)*** *the cell graph $C(G)$ contains no anisotropic path $XY_1 \ldots Y_l$ such that $|X| < |Y_1| = \ldots = |Y_l|$ and the vertex Y_l is heterogeneous.*

Proof.

- **(C)** Suppose that P is a uniform anisotropic path in $C(G)$ connecting heterogeneous vertices X and Y. Let $k = |X| = |Y|$. Complementing $G[A, B]$ for each co-constellation edge

$\{A, B\}$ of P, in G we obtain k vertex-disjoint paths connecting X and Y. These paths determine a one-to-one correspondence between X and Y. Given $v \in X$, denote its mate in Y by v^*. Call P *conducting* if this correspondence is an isomorphism between $G[X]$ and $G[Y]$, that is, two vertices u and v in X are adjacent exactly when their mates u^* and v^* are adjacent. In the case that one of X and Y is matching and the other is co-matching, we call P *conducting* also if the correspondence is an isomorphism between $G[X]$ and the complement of $G[Y]$.

Note that an isomorphism ϕ from G to another graph H preserves the mate relation, which is definable on pairs in $\phi(X) \times \phi(Y)$ in the same vein. In other terms, $\phi(v^*) = \phi(v)^*$ for any $v \in X$. It follows that, for any $u, v \in X$, $\phi(u)$ and $\phi(v)$ are adjacent exactly when $\phi(u)^*$ and $\phi(v)^*$ are adjacent, which means that ϕ preserves also the conducting property. More precisely, ϕ induces an isomorphism ϕ' from $C(G)$ to $C(H)$, which takes conducting paths in $C(G)$ to conducting paths in $C(H)$ and non-conducting ones to non-conducting ones.

If P is conducting, we can replace the subgraph $G[Y]$ with an isomorphic but different subgraph so that P becomes non-conducting in the cell graph $C(H)$ of the resulting graph H. Vice versa, if P is non-conducting, we can make such a replacement converting P to a conducting path.

By part (i) of Lemma 6, CR does not distinguish between G and H. As another consequence of part (i) of Lemma 6, $C(G) = C(H)$. Further, part (i) of Lemma 6 and Lemma 1 imply that if there is an isomorphism ϕ from G to H, the induced isomorphism ϕ' from $C(G)$ to $C(H)$ is the identity map on $\mathcal{P}_G$, the vertex set of $C(G)$. Therefore, ϕ' takes P onto itself, which contradicts preservation of the conducting property. We conclude that $G \ncong H$ and, hence, G is not amenable.

(D) Suppose that $C(G)$ contains a uniform anisotropic cycle Q of length m. All vertices of Q have the same cardinality as cells; denote it by k. Complementing $G[A, B]$ for each co-constellation edge $\{A, B\}$ of Q, in G we obtain the vertex-disjoint union of cycles whose lengths are multiples of m. As two extreme cases, we can have k cycles of length m each or we can have a single cycle of length km. Denote the isomorphism type of this union of cycles by $\tau(Q)$. Note that this type is isomorphism invariant: For an isomorphism ϕ from G to another graph H, $\tau(\phi'(Q)) = \tau(Q)$ for the induced isomorphism ϕ' from $C(G)$ to $C(H)$.

Let X and Y be two consecutive vertices in Q. We can replace the subgraph $G[X, Y]$ with an isomorphic but different bipartite graph so that, in the resulting graph H, $\tau(Q)$ becomes either kC_m or C_{km}, whatever we wish. We do replacement that changes $\tau(Q)$.

Similarly as for Condition **C**, we use part (ii) of Lemma 6 to argue that CR does not distinguish between G and H. Furthermore, $G \ncong H$ because the types $\tau(Q)$ in G and H are different. Therefore, G is not amenable.

(E) Suppose that $C(G)$ contains an anisotropic path $XY_1 \ldots Y_l Z$ such that $|X| < |Y_1| = \ldots = |Y_l| > |Z|$ (for the case of a cycle, where $Z = X$, the argument is virtually the same). Like in the proof of Condition **C**, the uniform anisotropic path $Y_1 \ldots Y_l$ determines a one-to-one correspondence between the sets Y_1 and Y_l. We make identification $Y_1 = Y_l = Y$ according to this correspondence.

Let $G[X, Y] = sK_{1,t}$ and $G[Z, Y] = aK_{1,b}$, where $s, a, t, b \geq 2$ (if any of these subgraphs is a co-constellation, we consider its complement). Thus, $|X| = s$, $|Z| = a$, and $|Y| = st = ab$. For each $x \in X$, let Y_x denote the set of vertices in Y adjacent to x. The set Y_z is defined similarly for each $z \in Z$. Note that

$$|Y_x| = t, \quad |Y_z| = b, \quad Y_x \cap Y_{x'} = \emptyset, \quad \text{and} \quad Y_z \cap Y_{z'} = \emptyset \tag{6}$$

for any $x \neq x'$ in X and $z \neq z'$ in Z (the disjointness conditions can be replaced with the covering conditions $\bigcup_{x \in X} Y_x = \bigcup_{z \in Z} Y_z = Y$). We regard $\mathcal{Y}_G = \{\, Y_x \,\}_{x \in X} \cup \{\, Y_z \,\}_{z \in Z}$ as a hypergraph on the vertex set Y.

For an isomorphism ϕ from G to another graph H, let $\mathcal{Y}_H$ be determined similarly by the path $\phi(X)\phi(Y_1)\ldots\phi(Y_l)\phi(Z)$ (note that $\mathcal{Y}_H$ does not depend on ϕ by Lemma 1). Then obviously $\mathcal{Y}_H \cong \mathcal{Y}_G$.

Now, let $\mathcal{H} = \{\, Y_x \,\}_{x \in X} \cup \{\, Y_z \,\}_{z \in Z}$ be an arbitrary hypergraph on the vertex set Y satisfying the conditions (6). Given $\mathcal{H}$, we can replace the subgraph $G[X, Y_1]$ with an isomorphic but different bipartite graph so that $\mathcal{Y}_H \cong \mathcal{H}$ for the resulting graph H. When we take $\mathcal{H} \ncong \mathcal{Y}_G$, this will ensure that $H \ncong G$. Similarly to Conditions **C** and **D**, part (ii) of Lemma 6 along with Lemma 1 implies that CR does not distinguish between G and H. Therefore, G cannot be amenable.

It remains to show that an appropriate choice of a hypergraph $\mathcal{H}$ is always available, that is, there are at least two non-isomorphic hypergraphs $\mathcal{H}_1$ and $\mathcal{H}_2$ on the vertex set Y with hyperedges denoted by Y_x, $x \in X$, and Y_z, $z \in Z$, satisfying the conditions (6) (if $t = b$, multiple hyperedges $Y_x = Y_z$ are allowed). Without loss of generality, suppose that $t \leq b$. In order to construct such $\mathcal{H}_1$ and $\mathcal{H}_2$, identify Y with the segment of integers $\{1, 2, \ldots, st\}$. Set $\{\, Y_x \,\}$ to be the partition of Y into s blocks of consecutive integers of length t in both $\mathcal{H}_1$ and $\mathcal{H}_2$. In $\mathcal{H}_1$, we set $\{\, Y_z \,\}$ to be the partition of Y into a blocks of consecutive integers of length b. In $\mathcal{H}_2$, we set $\{\, Y_z \,\}_{z \in Z} = \{\, \{j, j + a, j + 2a, \ldots, j + (b - 1)a\} \,\}_{j=1}^{a}$. The two hypergraphs are non-isomorphic because in $\mathcal{H}_1$ we have $Y_x \subset Y_z$ for some x and z, while no hyperedge includes another in $\mathcal{H}_2$.

(F) Suppose that $C(G)$ contains an anisotropic path $XY_1 \ldots Y_l$ where $|X| < |Y_1| = \ldots = |Y_l|$ and Y_l is heterogeneous. The uniform anisotropic path $Y_1 \ldots Y_l$ determines a one-to-one correspondence between Y_1 and Y_l, and we make identification $Y_1 = Y_l = Y$ accordingly to it. Consider an auxiliary graph A_G on the vertex set $X \cup Y$ where $A_G[X]$ is empty, $A_G[Y] = G[Y_l]$, and $A_G[X, Y] = G[X, Y_1]$.

For an isomorphism ϕ from G to another graph H, let A_H be determined similarly by the path $\phi(X)\phi(Y_1)\ldots\phi(Y_l)$. Then obviously $A_H \cong A_G$.

Like in the proof of Condition **C**, we can replace $G[Y_l]$ (hence $A_G[Y]$) with an isomorphic but different graph so that $A_H \ncong A_G$ for the resulting graph H. This will imply that G and H are non-isomorphic while indistinguishable by CR and, therefore, that G is not amenable. All what we have to show is that at least two different isomorphism types of A_H can be obtained by such a replacement.

Let $G[X, Y_1] = sK_{1,t}$ (in the case of a co-constellation, we consider the complement). Since $s, t \geq 2$ and $|Y_1| = st$, the cell Y_l cannot be pentagonal. Considering the complement if needed, we can assume without loss of generality that Y_l is matching. Consider a hypergraph $\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2$ where $\mathcal{H}_1 = \{Y_x\}_{x \in X}$ has hyperedges $Y_x = N(x) \cap Y_1$ as in the proof of Condition **E**, while $\mathcal{H}_2$ consists now of 2-element hyperedges corresponding to the edges of $G[Y_l]$. Similarly to Condition **E**, we can change the isomorphism type of $\mathcal{H}$ by modifying the subgraph $G[X, Y_1]$. This results in different isomorphism types of A_H. □

It turns out that Conditions **A**–**F** are not only necessary for amenability but also sufficient.

Theorem 8. *A graph G is amenable if and only if it satisfies Conditions **A**–**F**.*

The necessity part of the theorem is given by Lemmas 3 and 7. The proof of the sufficiency consists of Lemmas 9 and 10 below, that reveal a tree-like structure of amenable graphs. By an *anisotropic component* of the cell graph $C(G)$ we mean a maximal connected subgraph of $C(G)$ whose edges are all anisotropic. Note that if a vertex of $C(G)$ has no incident anisotropic edges, it forms a single-vertex anisotropic component.

Lemma 9. *Suppose that a graph G satisfies Conditions **A**–**F**. Then for any anisotropic component A of $C(G)$, the following is true.*

(i) A is a tree with the following monotonicity property. Let R be a vertex of A of the minimum cardinality (as a cell). Let A_R be the rooted directed tree obtained from A by rooting it at R. Then $|X| \leq |Y|$ for any directed edge (X, Y) of A_R.

(ii) A contains at most one heterogeneous vertex. If R is such a vertex, it has minimum cardinality among the vertices of A.

Proof.

(i) A cannot contain any uniform cycle by Condition **D** and any other cycle by Condition **E**. The monotonicity property follows from Condition **E**.

(ii) Assume that A contains more than one heterogeneous vertex. Consider two such vertices S and T. Let $S = Z_1, Z_2, \ldots, Z_l = T$ be the path from S to T in A. The monotonicity property stated in part (i) implies that there is j (possibly $j = 1, l$) such that $|Z_1| \geq \ldots \geq |Z_j| \leq \ldots \leq |Z_l|$. Since the path cannot be uniform by Condition **C**, at least one of the inequalities is strict. However, this contradicts Condition **F**.
Suppose that S is a heterogeneous vertex in A. Consider now a path $S = Z_1, Z_2, \ldots, Z_l = R$ in A where R is a vertex with the smallest cardinality. By the monotonicity property and Condition **F**, this path must be uniform, proving that $|S| = |R|$. □

Lemma 10. *Suppose that a graph G satisfies Conditions **A** and **B**, which allows us to consider the cell graph $C(G)$. Assume that every anisotropic component A of $C(G)$*

(G) *is a tree and*

(H) *has at most one heterogeneous vertex.*

Then G is amenable.

Proof. Given a graph H indistinguishable from G by CR, we have to show that G and H are isomorphic.

Since G and H satisfy the condition (2) for any i, any coloring C^i stable on the disjoint union of G and H determines a one-to-one correspondence f between the cells of the stable partitions of G and H. As follows directly from (2), $|X| = |f(X)|$ for every cell X of G. The map f is an isomorphism from $C(G)$ to $C(H)$ and, moreover, for any cells X and Y of G

(a) $G[X] \cong H[f(X)]$ and
(b) $G[X, Y] \cong H[f(X), f(Y)]$.

To show (a), consider a coloring $C = C^i$ stable on the disjoint union of G and H. Since C is stable on both G and H and X and $f(X)$ are cells of the corresponding partitions, both $G[X]$ and $H[f(X)]$ are regular. By Condition **A**, $G[X]$ is a unigraph. Since G and H have equal number of vertices, their non-isomorphism would mean that they have different degrees. This is impossible because then $X \cup f(X)$, a cell of C, would split in the next refinement step. The relation (b) follows from Condition **B** by a similar argument.

We now construct an isomorphism ϕ from G to H. By Lemma 1, we should have $\phi(X) = f(X)$ for each cell X. Therefore, we have to define the map $\phi : X \to f(X)$ on each X.

By assumption, an anisotropic component A of the cell graph $C(G)$ contains at most one heterogeneous vertex. Denote it by R_A if it exists. Otherwise fix R_A to be an arbitrary vertex in A.

For each A, define ϕ on $R = R_A$ to be an arbitrary isomorphism from $G[R]$ to $H[f(R)]$, which exists according to (a). After this, propagate ϕ to any other cell in A as follows. By assumption, A is a tree. Let A_R be the directed rooted tree obtained from A by rooting it at R. Suppose that ϕ is already defined on X and (X, Y) is an edge in A. Then ϕ is extended to Y so that this is an isomorphism from $G[X, Y]$ to $H[f(X), f(Y)]$.

It remains to argue that the map ϕ obtained in this way is indeed an isomorphism from G to H. It suffices to show that ϕ is an isomorphism between $G[X]$ and $H[f(X)]$ for each cell X of G and between $G[X, Y]$ and $H[f(X), f(Y)]$ for each pair of cells X and Y.

If X is homogeneous, $f(X)$ is homogeneous of the same type, complete or empty, according to (a). In this case, any ϕ is an isomorphism from $G[X]$ to $H[f(X)]$. If X is heterogeneous, the assumption of the lemma says that it belongs to a unique anisotropic component A (and $X = R_A$). Then ϕ is an isomorphism from $G[X]$ to $H[f(X)]$ by construction.

If $\{X, Y\}$ is an isotropic edge of $C(G)$, then (b) implies that $\{f(X), f(Y)\}$ is an isotropic edge of $C(H)$ of the same type, complete or empty. In this case, ϕ is an isomorphism from $G[X, Y]$ to $H[f(X), f(Y)]$, no matter how it is defined. If $\{X, Y\}$ is anisotropic, it belongs to some anisotropic component A, and ϕ is an isomorphism from $G[X, Y]$ to $H[f(X), f(Y)]$ by construction. $\square$

Lemmas 9 and 10 immediately imply the sufficiency part of Theorem 8. The proof of this theorem is therewith complete.

2.4 Examples and applications

Lemma 10 is a convenient tool for verifying amenability. For example, amenability of discrete graphs is a well-known fact. Note that Conditions **A** and **B** as well as Conditions **G** and **H** in Lemma 10 are fulfilled for discrete graphs by trivial reasons.

Checking these four conditions, we can also reprove the amenability of trees. Moreover, we can extend this result to the class of forests.

Corollary 11. *All forests are amenable.*

Proof. A regular acyclic graph is either an empty or a matching graph. This implies Condition **A**. Condition **B** follows from the observation that biregular acyclic graphs are either empty or forests of stars.

Let $C^*(G)$ be the version of the cell graph $C(G)$ where all empty edges are removed. Note that, if $C^*(G)$ contains a cycle, G must contain a cycle as well. Therefore, if G is acyclic, then $C^*(G)$ is acyclic too, and any anisotropic component of $C(G)$ must be a tree (which is Condition **G**).

To prove Condition **H**, suppose that an anisotropic component of $C(G)$ contains a path $X_0, X_1, \ldots, X_l$ connecting two heterogeneous vertices X_0 and X_l. Consider the subgraph $G[X_0] \cup G[X_0, X_1] \cup \ldots \cup G[X_{l-1}, X_l] \cup G[X_l]$. Since this graph has no vertex of degree 0 or 1, it must contain a cycle, a contradiction. ◻

Our characterization of amenable graphs leads to an efficient test for amenability of a given graph, that has the same time complexity as CR. It is known (Cardon and Crochemore [6]; see also [3]) that the stable partition of a given graph G can be computed in time $O((n+m)\log n)$. It is supposed that G is presented by its adjacency list.

Corollary 12. *The class of amenable graphs is recognizable in time $O((n+m)\log n)$, where n and m denote the number of vertices and edges of the input graph.*

Proof. Using known algorithms, we first compute the stable partition $\mathcal{P}_G$ of the input graph G. Sorting the adjacency list of each vertex according to $\mathcal{P}_G$, we compute a list of entries $d_{X,Y}$ of the *degree refinement matrix* $(d_{X,Y})$, $X, Y \in \mathcal{P}_G$, where $d_{X,Y}$ is equal to the number of neighbors in Y of any vertex in X. Along with the numbers $|X|$ and $|Y|$, $d_{X,Y}$ allows us to determine whether or not each subgraph $G[X, Y]$ is one of the graphs listed in Condition **B** of Lemma 3. Similarly, $|X|$ and $d_{X,X}$ allows us to determine whether or not each subgraph $G[X]$ is one of the graphs listed in Condition **A** of this lemma. If Conditions **A** and **B** are fulfilled, therewith we also obtain the cell graph $C(G)$.

Using breadth-first search, we find all anisotropic components of $C(G)$ and, simulateneously, for each of them we check Conditions **G** and **H** of Lemma 10. As follows from Theorem 8 along with Lemmas 9 and 10, any graph satisfying Conditions **A** and **B** is amenable if and only if it satisfies also Conditions **G** and **H**. ◻

We conclude this section by considering logical aspects of our result. A *counting quantifier* $\exists^m$ opens a sentence saying that there are at least m elements satisfying some property.

Immerman and Lander [13] discovered an intimate connection between color refinement and 2-variable first-order logic with counting quantifiers. This connection implies that amenability of a graph is equivalent to its definability in this logic. Thus, Corollary 12 asserts that the class of graphs definable by a first-order sentence with counting quantifiers and occurrences of just 2 variables is recognizable in polynomial time. Standard reductions lead to the following extension of this fact.

Corollary 13. *Let σ be a vocabulary consisting of binary relation symbols. Then the class of structures over σ definable in 2-variable first-order logic with counting quantifiers is recognizable in polynomial time.*

Finally, note that CR admits a natural extension to structures over any binary vocabulary σ (which is most obvious for vertex-colored graphs), and the Immerman-Lander result is preserved under this extension. Thus, Corollary 13 implies that there is an efficient way to check whether or not the generalized CR applies to a given input structure.

3 Amenable graphs are compact

For a permutation π of the set $\{1,\ldots,n\}$, the corresponding *permutation matrix* $P_\pi = (p_{ij})$ is defined by $p_{ij} = 1$ if $\pi(i) = j$ and $p_{ij} = 0$ otherwise. An $n \times n$ real matrix $X = (x_{ij})$ is *doubly stochastic* if

$$\sum_{i=1}^{n} x_{ij} = 1 \text{ for every } j, \quad \sum_{j=1}^{n} x_{ij} = 1 \text{ for every } i, \quad \text{and } x_{ij} \geq 0 \text{ for all } i,j.$$

Doubly stochastic matrices are closed under products and convex combinations. The set of all $n \times n$ doubly stochastic matrices defined by the above linear inequalities is the *Birkhoff polytope* $B_n \subset \mathbb{R}^{n^2}$. By Birkhoff's Theorem, the $n!$ permutation matrices form precisely the set of all extreme points of B_n. Equivalently, every doubly stochastic matrix is a convex combination of permutation matrices.

Fractional isomorphisms. Consider graphs with vertex set $\{1,\ldots,n\}$. Let $\mathrm{Iso}(G,H)$ denote the set of all permutation matrices P_π such that π is an isomorphism from G to H and let $\mathrm{Aut}(G) = \mathrm{Iso}(G,G)$. If A and B are the adjacency matrices of graphs G and H respectively, then G and H are isomorphic iff there is a permutation matrix X such that

$$AX = XB. \tag{7}$$

In fact, (7) is true for a permutation matrix X exactly when $X \in \mathrm{Iso}(G,H)$.

G and H are called *fractionally isomorphic* if (7) is satisfied by a doubly stochastic matrix X. Such an X is a *fractional isomorphism* from G to H. It is easy to verify that being fractionally isomorphic is an equivalence relation on graphs.

Analyzing the connection between fractional isomorphism and the color refinement procedure, Ramana, Scheinerman, and Ullman [18] show the following result that can be considered an analog of Lemma 1 for fractional isomorphisms. For a partition $V_1, \ldots, V_m$ of $\{1, \ldots, n\}$ let $X_1, \ldots, X_m$ be matrices, where the rows and columns of X_i are indexed by elements of V_i. Then we denote the block-diagonal matrix with blocks $X_1, \ldots, X_m$ by $X_1 \oplus \cdots \oplus X_m$.

Lemma 14 (Ramana et al. [18]). *Let G be a graph on vertex set $\{1, \ldots, n\}$ and assume that the elements $V_1, \ldots, V_m$ of the coarsest equitable partition $\mathcal{P}_G$ of G are intervals of consecutive integers. Let X be a fractional automorphism of G, i.e., a doubly stochastic matrix commuting with the adjacency matrix of G. Then X has the form $X = X_1 \oplus \cdots \oplus X_m$, that is, X is a block diagonal matrix where the blocks $X_1, \ldots, X_m$ correspond to $V_1, V_2, \ldots, V_m$.*

Note that the assumption of the lemma can be ensured for any graph by appropriately renaming its vertices. For the reader's convenience we include a self-contained proof of Lemma 14 in Appendix A.

Compact graphs. Denote the polytope of all fractional isomorphisms from G to H by $S(G, H) \subset \mathbb{R}^{n^2}$. The set of isomorphisms $\mathrm{Iso}(G, H)$ is contained in $\mathrm{Ext}(S(G, H))$ (where $\mathrm{Ext}(X)$ denotes the set of all extreme points of a set X). Indeed, $\mathrm{Iso}(G, H)$ is the set of integral extreme points of $S(G, H)$. The set $S(G) = S(G, G)$ is the polytope of *fractional automorphisms* of G.

A graph G is called *compact* [21] if $S(G)$ has no other extreme points than $\mathrm{Aut}(G)$, i.e., $\mathrm{Ext}(S(G)) = \mathrm{Aut}(G)$. Compactness of G can equivalently be defined by any of the following two conditions:

- The polytope $S(G)$ is integral;
- Every fractional automorphism of G is a convex combination of standard automorphisms of G, i.e., $S(G) = \langle \mathrm{Aut}(G) \rangle$, where $\langle X \rangle$ denotes the convex hull of a set X.

Example 15. *Complete graphs are compact as a consequence of Birkhoff's theorem. Trees and cycles are compact [21]. Matching graphs mK_2 are compact. This is a particular instance of a much more general result by Tinhofer [23]: If G is compact, then mG is compact for any m. Tinhofer [23] also observes that compact graphs are closed under complement.*

For a negative example, note that the graph $C_3 + C_4$ is not compact. This follows from a general result in [23]: All regular compact graphs must be vertex-transitive (and $C_3 + C_4$ is not).

The concept of a compact graph is motivated by a linear programming approach to Graph Isomorphism that is based on the following fact.

Proposition 16 (Tinhofer [23]). *Let G be a compact graph. Then for any graph H, either all or none of the extreme points of the polytope $S(G, H)$ are integral.*

As mentioned in the introduction, the above proposition yields a linear-programming based polynomial-time algorithm to test if a compact graph G is isomorphic to any other graph H.

However, no polynomial time algorithm is known to check if a graph is compact. The only known complexity upper bound, noted in [23], is coNP, because testing if every vertex of a given polytope is integral is in coNP. Nevertheless, the following result shows that Tinhofer's approach works for all amenable graphs.

Theorem 17. *All amenable graphs are compact.*

We defer the proof to the next section. Theorem 17 unifies and extends several earlier results providing examples of compact graphs. In particular, it gives another proof of the fact that almost all graphs are compact, which also follows from a result of Godsil [10, Corollary 1.6]. Indeed, while Babai, Erdös, and Selkow [2] proved that almost all graphs are discrete (and, moreover, the discrete partition is reachable in 2 refinement rounds), we already mentioned in Section 2.4 that all discrete graphs are amenable.

Furthermore, Theorem 17 reproves Tinhofer's result that trees are compact.[5] Using Corollary 11, we can extend this result to forests. This extension is not straightforward as compact graphs are not closed under disjoint union; see Example 15. In [22], Tinhofer proves compactness for the class of *strong tree-cographs*, which includes forests only with pairwise non-isomorphic connected components.

Compactness of unigraphs, which also follows from Theorem 17, seems to be earlier never observed. Summarizing, we note the following result.

Corollary 18. *Discrete graphs, forests, and unigraphs are compact.*

4 Proof of Theorem 17

Before we proceed to the proof, it will be helpful to recall useful facts about convex optimization and doubly stochastic matrices.

Convex sets and extreme points. For $x, y \in \mathbb{R}^n$, a *convex combination* of x and y is any vector of the form $\alpha x + (1-\alpha)y$ where $0 \le \alpha \le 1$. More generally, $\sum_{i=1}^{k} \alpha_i x_i$ is a convex combination of k points $x_1, \ldots, x_k \in \mathbb{R}^n$ if $\sum_{i=1}^{k} \alpha_i = 1$ and $\alpha_i \ge 0$ for all i. A set $S \subseteq \mathbb{R}^n$ is *convex* if, for every two points $x, y \in S$, S contains also any convex combination of these points, that is, the segment with endpoints x and y. The *convex hull* of a set $S \subseteq \mathbb{R}^n$, denoted by $\langle S \rangle$, is the inclusion-minimal convex set containing S. Equivalently, $\langle S \rangle$ is the set of all convex combinations of any finite number of points in S. A point $z \in S$ is called an *extreme point* of S if it cannot be represented as a convex combination of other points of S, that is, $z = \alpha x + (1-\alpha)y$ with $0 \le \alpha \le 1$ implies $z = x = y$. We will denote the set of all extreme points of S by $\mathrm{Ext}(S)$. The Minkowski theorem says that, if a convex set S is bounded and closed, then $S = \langle \mathrm{Ext}(S) \rangle$.

Polytopes. Speaking of polytopes, we always mean *convex polytopes.* Such a polytope P can be defined as the intersection of a set of half-spaces of $\mathbb{R}^n$ or, algebraically, $P = \{ x \in \mathbb{R}^n : Ax \le b \}$ where A is a real $m \times n$ matrix, x and b are supposed to be n-dimensional column vectors, and the inequality is understood row-wise. A matrix or a vector is *integral* if all its entries are integers. Given integral A and b, deciding if P is non-empty is exactly the linear programming problem, that can be solved in polynomial time, for example, by the famous ellipsoid method. A point $x \in P$ is called a *basic feasible solution* to the system

[5] The proof of Theorem 17 uses only compactness of complete graphs, matching graphs, and the 5-cycle.

$Ax \le b$ if $\operatorname{rank} A_x = n$, where the matrix A_x is obtained from A by removing those rows where the inequality is strict. It is known that $\operatorname{Ext}(P)$ consists exactly of basic feasible solutions to the underlying system of inequalities. This implies that $\operatorname{Ext}(P)$ is a finite set. If A and b are integral, this also implies that the bit length of any extreme point is bounded by a polynomial in the bit length of A and b. Though the set $\operatorname{Ext}(P)$ can be exponentially large in n, computing a single point in $\operatorname{Ext}(P)$ reduces in polynomial time to the linear programming problem.

A polytope is called *integral* if all its extreme points are integral.

Doubly stochastic matrices and the Birkhoff polytope. Let π be a permutation of the set $\{1, \ldots, n\}$. The corresponding *permutation matrix* $P_\pi = (p_{ij})$ is defined by $p_{ij} = 1$ if $\pi(i) = j$ and $p_{ij} = 0$ otherwise. An $n \times n$ real matrix $X = (x_{ij})$ is *doubly stochastic* if $\sum_{i=1}^n x_{ij} = 1$ for every j, $\sum_{j=1}^n x_{ij} = 1$ for every i, and $x_{ij} \ge 0$ for all i, j. The product and any convex combination of doubly stochastic matrices are themselves doubly stochastic matrices. Considered as a subset of $\mathbb{R}^{n^2}$, the set of all $n \times n$ doubly stochastic matrices is known as the Birkhoff polytope B_n. Note that every permutation matrix is an extreme point of B_n. The Birkhoff theorem says that B_n has no other extreme points. Equivalently, every doubly stochastic matrix is a convex combination of permutation matrices.

We now proceed to the proof of the theorem. Given an amenable graph G and its fractional automorphism X, we have to express X as a convex combination of permutation matrices in $\operatorname{Aut}(G)$. Our proof strategy consists in exploiting the structure of amenable graphs as described by Theorem 8 and Lemma 9. The last lemma refers to anisotropic components of the cell graph $C(G)$. Given an anisotropic component A of $C(G)$, we define the *anisotropic component G_A of G* as the subgraph of G induced by the union of all cells belonging to A. Our overall idea is to prove the claim separately for each anisotropic component G_A, applying an inductive argument on the number of cells in A. A key role will be played by the facts that, according to Lemma 9, A is a tree with at most one heterogeneous vertex.

This scenario cannot be implemented directly by a simple reason: In order to run induction, we need that a subgraph of an amenable graph induced by a number of cells is also amenable, which is not always the case. In order to remove this complication, we introduce a definition generalizing amenable graphs for the purpose of applying induction.

Definition 19. Let G be a vertex-colored graph with the partition $V(G) = V_1 \cup \cdots \cup V_m$ of its vertex set into color classes. We say that G is *pseudo-amenable* if there is an amenable graph G' such that

1. $V(G) \subset V(G')$ and $V_1, \ldots, V_m$ are cells of G';
2. G is an induced subgraph of G' obtained by deleting the remaining cells of G'.

The next definition is needed in order to extend the notion of compactness to pseudo-amenable graphs. From now on, without loss of generality we suppose that the vertices $1, \ldots, n$ of a vertex-colored graph G are named so that every color class is an interval of consecutive integers.

Definition 20. For a vertex-colored graph G, *color-preserving automorphisms* are automorphisms that map each color class to itself. A *color-preserving fractional automorphism* of G is a fractional automorphism

$$X = X_1 \oplus \cdots \oplus X_m$$

such that the blocks $X_1, \ldots, X_m$ of the block-diagonal doubly stochastic matrix X correspond to the color classes $V_1, \ldots, V_m$ of G. More precisely, the rows and columns of X_i are indexed by the vertices in the set V_i, for each i.

Claim 21. *Every pseudo-amenable graph G is compact in the sense that every color-preserving fractional automorphism of G is a convex combination of color-preserving automorphisms of G.*

This claim implies the theorem because we can consider every amenable graph G as pseudo-amenable with the coarsest equitable partition $\mathcal{P}_G$ defining the color classes. By Lemma 14, all fractional automorphisms of G (which includes all automorphisms of G; see also Lemma 1) will be color-preserving.

In the sequel we prove the claim, essentially by induction on the number of color classes. For a pseudo-amenable graph G we define its cell graph $C(G)$ on the set of color classes exactly as $C(G)$ is defined on the coarsest equitable partitions of amenable graphs. Now, it makes sense to talk of anisotropic components of $C(G)$ and, hence, of anisotropic components of the pseudo-amenable graph G.

We first consider the case when G consists of a single anisotropic component. By Lemma 9, the corresponding cell graph $C(G)$ has at most one heterogeneous vertex and the anisotropic edges form a spanning tree of $C(G)$. Without loss of generality, we can number the cells $V_1, \ldots, V_m$ of G so that V_1 is the unique heterogeneous cell if such exists; otherwise V_1 is chosen among the cells of minimum cardinality. Moreover, we can suppose that, for each $i \le m$, the cells $V_1, \ldots, V_i$ induce a connected subgraph in the tree of anisotropic edges of $C(G)$.

We will prove this case by induction on the number m of cells. In the base case of $m = 1$, our graph $G = G[V_1]$ is one of the graphs listed in Condition **A** of Lemma 3. All of them are known to be compact; see Example 15.1–4. As induction hypothesis, assume that the graph $H = G[V_1 \cup \cdots \cup V_{m-1}]$ is compact. For the induction step, we have to show compactness of $G = G[V_1 \cup \cdots \cup V_m]$.

Denote $D = V_m$. Since G has no more than one heterogeneous cell, $G[D]$ is complete or empty. It will be instructive to think of D as a "leaf" cell having a unique anisotripic link to the remaning part H of G. Let $C \in \{V_1, \ldots, V_{m-1}\}$ be the unique cell such that $\{C, D\}$ is an anisotropic edge of $C(G)$. To be specific, suppose that $G[C, D] = sK_{1,t}$. If $G[C, D]$ is a co-constallation, we can consider the complement of G and use the facts that the class of amenable graphs is closed under complementation and that complementation does not change color-preserving fractional isomorphisms of the graph (cf. Example 15.4). By the monotonicity property in part (i) of Lemma 9, $|C| = s$ and $|D| = st$. Let $C = \{c_1, c_2, \ldots, c_s\}$ and, for each i, $N(c_i) \subset D$ be the neighborhood of c_i in $G[C, D]$. Thus, $D = \bigcup_{i=1}^{s} N(c_i)$.

Let X be a color-preserving fractional automorphism of G. It is convenient to break it up into three blocks:

$$X = X' \oplus Y \oplus Z,$$

where Y and Z correspond to C and D respectively, and X' is the rest. By induction hypothesis we have the convex combination

$$X' \oplus Y = \sum_{P' \oplus P \in \mathrm{Aut}(H)} \alpha_{P',P} \, P' \oplus P, \tag{8}$$

where $P' \oplus P$ are permutation matrices corresponding to automorphisms of the graph H, such that the permutation matrix block P denotes automorphism's action on color class C and P' the action on the remaining color classes of H.

We need to show that X is a convex combination of automorphisms of G. Let A denote the adjacency matrix of G and $A_{C,D}$ denote the $s \times st$ submatrix corresponding to the block row-indexed by C and column-indexed by D. Likewise, $A_{D,C}$ denotes the $st \times s$ submatrix with rows indexed by D and columns by C. Since X is a fractional automorphism of G, we have

$$XA = AX. \tag{9}$$

Recall that Y and Z are blocks of X corresponding to color classes C and D. Looking at the corner fragments of the matrices in the left and the right hand sides of (9), we get

$$\begin{pmatrix} Y & 0 \\ 0 & Z \end{pmatrix} \begin{pmatrix} A_{C,C} & A_{C,D} \\ A_{D,C} & A_{D,D} \end{pmatrix} = \begin{pmatrix} A_{C,C} & A_{C,D} \\ A_{D,C} & A_{D,D} \end{pmatrix} \begin{pmatrix} Y & 0 \\ 0 & Z \end{pmatrix},$$

which implies

$$Y A_{C,D} = A_{C,D} \, Z, \tag{10}$$

$$A_{D,C} \, Y = Z \, A_{D,C}. \tag{11}$$

Consider Z as an $st \times st$ matrix whose rows and columns are indexed by the elements of sets $N(c_1), N(c_2), \ldots, N(c_r)$ in that order. We can thus think of Z as an $s \times s$ block matrix of $t \times t$ matrix blocks $Z^{(k,\ell)}, 1 \le k, \ell \le s$. The next claim is a consequence of Equations (10) and (11).

Claim 22. *Each block $Z^{(k,\ell)}$ in Z is of the form*

$$Z^{(k,\ell)} = y_{k,\ell} W^{(k,\ell)}, \tag{12}$$

where $y_{k,\ell}$ is the $(k,\ell)^{th}$ entry of Y, and $W^{(k,\ell)}$ is a doubly stochastic matrix.

Proof. We first note from Equation (10) that the $(k,j)^{th}$ entry of the $s \times st$ matrix $YA_{C,D} = A_{C,D}Z$ can be computed in two different ways. In the left hand side matrix, it is $y_{k,\ell}$ for each $j \in N(c_\ell)$. On the other hand, the right hand side matrix implies that the same $(k,j)^{th}$ entry is also the sum of the j^{th} column of the $N(c_k) \times N(c_\ell)$ block $Z^{(k,\ell)}$ of the matrix Z.

We conclude, for $1 \leq k, \ell \leq s$, that each column in $Z^{(k,\ell)}$ adds up to $y_{k,\ell}$. By a similar argument, applied to Equation (11) this time, it follows, for each $1 \leq k, \ell \leq s$, that each *row* of any block $Z^{(k,\ell)}$ of Z adds up to $y_{k,\ell}$.

We conclude that, if $y_{k,\ell} \neq 0$, then the matrix $W^{(k,\ell)} = \frac{1}{y_{k,\ell}} Z^{(k,\ell)}$ is doubly stochastic. If $y_{k,\ell} = 0$, then (12) is true for any choice of $W^{(k,\ell)}$. $\square$

For every $P = (p_{k\ell})$ appearing in an automorphism $P' \oplus P$ of H (see Equation (8)), we define the $st \times st$ doubly stochastic matrix W_P by its $t \times t$ blocks indexed by $1 \leq k, \ell \leq s$ as follows:

$$W_P^{(k,\ell)} = \begin{cases} W^{(k,\ell)} & \text{if } p_{k\ell} = 1, \\ 0 & \text{if } p_{k\ell} = 0. \end{cases} \tag{13}$$

Equations (8) and (12) imply that

$$X = X' \oplus Y \oplus Z = \sum_{P' \oplus P \in \mathrm{Aut}(H)} \alpha_{P',P} \, P' \oplus P \oplus W_P. \tag{14}$$

In order to see this, on the left hand side consider the $(k, \ell)^{th}$ block $Z^{(k,\ell)}$ of Z. On the right hand side, note that the corresponding block in each $P' \oplus P \oplus W_P$ is the matrix $W^{(k,\ell)}$. Clearly, the overall coefficient for this block equals the sum of $\alpha_{P',P}$ over all P' and P such that $p_{k,\ell} = 1$, which is precisely $y_{k,\ell}$ by Equation (8).

Since each $W^{(k,\ell)}$ is a doubly stochastic matrix, by Birkhoff's theorem we can write it as a convex combination of $t \times t$ permutation matrices $Q_{j,k,\ell}$, whose rows are indexed by elements of $N(c_k)$ and columns by elements of $N(c_\ell)$:

$$W^{(k,\ell)} = \sum_{j=1}^{t!} \beta_{j,k,\ell} \, Q_{j,k,\ell}.$$

Substituting the above expression in Equation (13), that defines the doubly stochastic matrix W_P, we express W_P as a convex combination of permutation matrices:

$$W_P = \sum_{Q} \delta_{Q,P} \, Q$$

where Q runs over all $st \times st$ permutation matrices indexed by the vertices in color class D. Notice here that $\delta_{Q,P}$ is nonzero only for those permutation matrices Q that have structure similar to that described in Equation (13): The block $Q^{(k,\ell)}$ is a null matrix if $p_{k\ell} = 0$ and it is some $t \times t$ permutation matrix if $p_{k\ell} = 1$. For each such Q, the $(s + st) \times (s + st)$ permutation matrix $P \oplus Q$ is an automorphism of the subgraph $G[C, D] = sK_{1,t}$ (because Q maps $N(c_i)$ to $N(c_j)$ whenever P maps c_i to c_j). Since $P \in \mathrm{Aut}(G[C])$ and D is a homogeneous set in G, we conclude that, moreover, $P \oplus Q$ is an automorphism of the subgraph $G[C \cup D]$.

Now, if we plug the above expression for each W_P in Equation (14), we will finally obtain the desired convex combination

$$X = \sum_{P',P,Q} \gamma_{P',P,Q} \, P' \oplus P \oplus Q.$$

It remains to argue that every $P' \oplus P \oplus Q$ occurring in this sum is an automorphism of G. Recall that a pair P', P can appear here only if $P' \oplus P \in \mathrm{Aut}(H)$. Moreover, if such a pair is extended to a matrix $P' \oplus P \oplus Q$, then $P \oplus Q \in \mathrm{Aut}(G[C \cup D])$. Since $G[B, D]$ is isotropic for every color class $B \neq D$ of G, we conclude that $P' \oplus P \oplus Q \in \mathrm{Aut}(G)$. This completes the induction step and finishes the case when G has one anisotropic component.

Next, we consider the case when $C(G)$ has several anisotropic components $T_1, \ldots, T_k$, $k \geq 2$. Let $G_1, \ldots, G_k$, where $G_i = G[\bigcup_{U \in V(T_i)} U]$, be the corresponding anisotropic components of G. By the proof of the previous case we already know that G_i is compact for each i.

Claim 23. *The automorphism group* $\mathrm{Aut}(G)$ *of* G *is the product of the automorphism groups* $\mathrm{Aut}(G_i)$, $1 \leq i \leq k$.

Proof. Recall that any automorphism of G must map each color class of G, which is a cell of the underlying amenable graph G', onto itself. Thus, any automorphism π of G is of the form $(\pi_1, \ldots, \pi_k)$, where π_i is an automorphism of the subgraph G_i. Now, for any two subgraphs G_i and G_j, we examine the edges between $V(G_i)$ and $V(G_j)$. For any color classes $U \subseteq V(G_i)$ and $U' \subseteq V(G_j)$, the edge $\{U, U'\}$ is isotropic because it is not contained in any anisotropic component of $C(G)$. Therefore, the bipartite graph $G[U, U']$ is either complete or empty. It follows that for any automorphisms π_i of G_i, $1 \leq i \leq k$, the permutation $\pi = (\pi_1, \ldots, \pi_k)$ is a color-preserving automorphism of the graph G. ☐

As follows from Lemma 14, any fractional automorphism X of G is of the form

$$X = X_1 \oplus \cdots \oplus X_k,$$

where X_i is a fractional automorphism of G_i for each i. As each G_i is compact we can write each X_i as a convex combination

$$X_i = \sum_{\pi \in \mathrm{Aut}(G_i)} \alpha_{i,\pi} P_\pi.$$

This implies

$$I \oplus \cdots \oplus I \oplus X_i \oplus I \oplus \cdots \oplus I = \sum_{\pi \in \mathrm{Aut}(G_i)} \alpha_{i,\pi} I \oplus \cdots \oplus I \oplus P_\pi \oplus I \oplus \cdots \oplus I, \tag{15}$$

where block diagonal matrices in the above expression have X_i and P_π respectively in the i^{th} block (indexed by elements of $V(G_i)$) and identity matrices as the remaining blocks.

We now decompose the fractional automorphism X as a matrix product of fractional automorphisms of G

$$\begin{aligned} X &= X_1 \oplus \cdots \oplus X_k \\ &= (X_1 \oplus I \oplus \cdots \oplus I) \cdot (I \oplus X_2 \oplus \cdots \oplus I) \cdot \cdots \cdot (I \oplus \cdots \oplus I \oplus X_k). \end{aligned}$$

Substituting for $I \oplus \cdots \oplus I \oplus X_i \oplus I \oplus \cdots \oplus I$ from Equation (15) in the above expression and writing the product of sums as a sum of products, we see that X is a convex combination of permutation matrices of the form $P_{\pi_1} \oplus \cdots \oplus P_{\pi_k}$ where $\pi_i \in \mathrm{Aut}(G_i)$ for each i. By Claim 23, all the terms $P_{\pi_1} \oplus \cdots \oplus P_{\pi_k}$ correspond to automorphisms of G. Therefore, G is compact.

The proof of Claim 21 and, hence, of Theorem 17 is complete.

5 A color-refinement based hierarchy of graphs

Let $u \in V(G)$ and $v \in V(H)$ be vertices of two graphs G and H. By *individualization* of u and v we mean assigning the same *new color* to u and v, which makes them distinguished from the remaining vertices of G and H. Tinhofer [23] proved that, if G is compact, then the following polynomial-time algorithm correctly decides if G and H are isomorphic.

1. Run Color Refinement on G and H until the coloring of $V(G) \cup V(H)$ stabilizes.
2. If the multisets of colors in G and H are different, then output "non-isomorphic" and stop. Otherwise,
   (a) if all color classes are singletons in G and H, then if the mapping $u \mapsto v$ (where $u \in V(G)$ and $v \in V(H)$ have the same color) is an isomorphism, output "isomorphic" and stop. Else output "non-isomorphic" and stop.
   (b) pick any color class with at least two vertices in both G and H, select an arbitrary $u \in V(G)$ and $v \in V(H)$ in this color class and individualize them. Goto Step 1.

If G and H are any two non-isomorphic graphs then Tinhofer's algorithm will always output "non-isomorphic". However, it can fail for isomorphic input graphs, in general. We call G a *Tinhofer graph* if the algorithm works correctly on G and every H for all choices of vertices to be individualized. If G is a Tinhofer graph, then the algorithm can be used to even find a canonical labeling of G. In particular, this applies to all compact graphs, and Theorem 17 gives us also the following fact.

Corollary 24. *The class of amenable graphs admits a polynomial-time canonical labeling algorithm.*

Let $A \subseteq Aut(G)$ be a subgroup of automorphisms of a graph G. Then the partition of $V(G)$ into A-orbits is called an *orbit partition*. Any orbit partition of G is equitable, but the converse is not true, in general. However, Godsil [10, Corollary 1.3] has shown that the converse holds for compact graphs. We call the set of all graphs with this property *Godsil graphs*, i.e., these are graphs for which the two notions of an equitable and an orbit partition coincide. The aforementioned result by Tinhofer can be easily strengthened as follows.

Lemma 25. *Any Godsil graph is a Tinhofer graph.*

Proof. Assume that G is a Godsil graph. It suffices to show that Tinhofer's algorithm is correct whenever G and H are isomorphic. Let ϕ be an isomorphism from G to H. We will prove that, after the i-th refinement step made by the algorithm, there exists an isomorphism ϕ_i from G to H that preserves colors of the vertices. If this is true for each i, the algorithm terminates only if the discrete partition (i.e., the finest partition into singletons) is reached. Suppose that this happens in the k-th step. Then ϕ_k ensures that the algorithm decides isomorphism.

We prove the claim by induction on i. At the beginning, $\phi_i = \phi$. Assume that an isomorphism ϕ_i exists and the partition is still not discrete. Suppose that now the algorithm individualizes $u \in V(G)$ and $v \in V(H)$. If $v = \phi_i(u)$, then $\phi_{i+1} = \phi_i$. Otherwise, consider the vertices u and $\phi_i^{-1}(v)$, which are in the same monochromatic class of G. Note that the partition of G produced in each refinement step is equitable. Since G is Godsil, there is an automorphism α preserving the partition such that $\alpha(u) = \phi_i^{-1}(v)$. We can, therefore, take $\phi_{i+1} = \phi_i \circ \alpha$. □

The orbit partition of G with respect to $\mathrm{Aut}(G)$ is always a refinement of the coarsest equitable partition of G. We call G *refinable* if the coarsest equitable partition is the orbit partition of $Aut(G)$. It is easy to show the following.

Lemma 26. *Any Tinhofer graph is refinable.*

Proof. Suppose that G is not refinable. Then G has vertices u and v that are in the same lex-least element of the coarsest equitable partition but in different orbits. The latter means that individualization of u and v in isomorphic copies G' and G'' of G gives non-isomorphic results. Therefore, if Tinhofer's algorithm is run on G' and G'' and individualizes u and v, it eventually decides non-isomorphism. □

Summarizing Theorem 17, Lemmas 25 and 26, and [10, Corollary 1.3], we obtain the chain of inclusions:

$$\mathsf{Discrete} \subset \mathsf{Amenable} \subset \mathsf{Compact} \subset \mathsf{Godsil} \subset \mathsf{Tinhofer} \subseteq \mathsf{Refinable} \tag{16}$$

All but the last inclusion are provably strict. The details can be found in Appendix B.

It remains open to establish a separation between Tinhofer and Refinable. Note that, while Amenable is the class of graphs for which the color-refinement procedure suffices to solve Graph Isomorphism, Refinable consists of graphs for which color refinement correctly solves Graph Automorphism (checking if there is a nontrivial automorphism).

Finally, we mention that the hierarchy (16) collapses to Discrete if we restrict ourselves to only rigid (i.e. asymmetric) graphs.

References

1. A. Atserias and E. N. Maneva. Sherali-adams relaxations and indistinguishability in counting logics. *SIAM J. Comput.*, 42(1):112–137, 2013.
2. L. Babai, P. Erdös, and S. M. Selkow. Random graph isomorphism. *SIAM J. Comput.*, 9(3):628–635, 1980.
3. C. Berkholz, P. Bonsma, and M. Grohe. Tight lower and upper bounds for the complexity of canonical colour refinement. In H. L. Bodlaender and G. F. Italiano, editors, *Algorithms - ESA 2013 - 21st Annual European Symposium, Sophia Antipolis, France, September 2-4, 2013. Proceedings*, volume 8125 of *Lecture Notes in Computer Science*, pages 145–156. Springer, 2013.
4. A. Borri, T. Calamoneri, and R. Petreschi. Recognition of unigraphs through superposition of graphs. *J. Graph Algorithms Appl.*, 15(3):323–343, 2011.
5. R. Busacker and T. Saaty. *Finite graphs and networks: an introduction with applications.* International Series in Pure and Applied Mathematics. McGraw-Hill Book Company, New York etc., 1965.
6. A. Cardon and M. Crochemore. Partitioning a graph in $O(|A| \log_2 |V|)$. *Theor. Comput. Sci.*, 19:85–98, 1982.
7. A. Chan and C. Godsil. Symmetry and eigenvectors. In *Graph Symmetry*, 497:75–106, 1997.
8. S. Evdokimov, M. Karpinski, and I. N. Ponomarenko. Compact cellular algebras and permutation groups. *Discrete Mathematics*, 197-198:247–267, 1999.
9. S. Evdokimov, I. N. Ponomarenko, and G. Tinhofer. Forestal algebras and algebraic forests (on a new class of weakly compact graphs). *Discrete Mathematics*, 225(1-3):149–172, 2000.

10. C. Godsil. Compact graphs and equitable partitions. *Linear Algebra and its Applications*, 255(1–3):259 – 266, 1997.
11. M. Grohe, K. Kersting, M. Mladenov, and E. Selman. Dimension reduction via colour refinement. In A. S. Schulz and D. Wagner, editors, *Algorithms - ESA 2014 - 22th Annual European Symposium, Wroclaw, Poland, September 8-10, 2014. Proceedings*, volume 8737 of *Lecture Notes in Computer Science*, pages 505–516. Springer, 2014.
12. M. Grohe and M. Otto. Pebble games and linear equations. In P. Cégielski and A. Durand, editors, *Computer Science Logic (CSL'12) - 26th International Workshop/21st Annual Conference of the EACSL*, volume 16 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 289–304, Dagstuhl, Germany, 2012. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
13. N. Immerman and E. Lander. Describing graphs: A first-order approach to graph canonization. In A. Selman, editor, *Complexity Theory Retrospective*, pages 59–81. Springer, 1990.
14. R. Johnson. Simple separable graphs. *Pac. J. Math.*, 56:143–158, 1975.
15. K. Kersting, M. Mladenov, R. Garnett, and M. Grohe. Power iterated color refinement. In C. E. Brodley and P. Stone, editors, *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada.*, pages 1904–1910. AAAI Press, 2014.
16. M. Koren. Pairs of sequences with a unique realization by bipartite graphs. *Journal of Combinatorial Theory, Series B*, 21(3):224 – 234, 1976.
17. A. Krebs and O. Verbitsky. Universal covers, color refinement, and two-variable logic with counting quantifiers: Lower bounds for the depth. *E-print*, *arxiv.org/abs/1407.3175*, 2014.
18. M. V. Ramana, E. R. Scheinerman, and D. Ullman. Fractional isomorphism of graphs. *Discrete Mathematics*, 132(1-3):247–265, 1994.
19. E. R. Scheinerman and D. H. Ullman. *Fractional graph theory. A rational approach to the theory of graphs.* Wiley: John Wiley & Sons, 1997.
20. N. Shervashidze, P. Schweitzer, E. J. van Leeuwen, K. Mehlhorn, and K. M. Borgwardt. Weisfeiler-Lehman graph kernels. *Journal of Machine Learning Research*, 12:2539–2561, 2011.
21. G. Tinhofer. Graph isomorphism and theorems of birkhoff type. *Computing*, 36:285–300, 1986.
22. G. Tinhofer. Strong tree-cographs are birkhoff graphs. *Discrete Applied Mathematics*, 22(3):275–288, 1989.
23. G. Tinhofer. A note on compact graphs. *Discrete Applied Mathematics*, 30(2-3):253–264, 1991.
24. G. Tinhofer and M. Klin. Algebraic combinatorics in mathematical chemistry. Methods and algorithms III. Graph invariants and stabilization methods, 1999.
25. R. Tyshkevich. Decomposition of graphical sequences and unigraphs. *Discrete Mathematics*, 220(1-3):201–238, 2000.
26. G. Valiente. *Algorithms on Trees and Graphs.* Springer, 2002.
27. H. Whitney. *Congruent graphs and connectivity of graphs.* Amer. J. Math. 54:150-168, 1932.
28. M. Ziv-Av. Results of computer algebra calculations for triangle free strongly regular graphs. E-print: *www.math.bgu.ac.il/ zivav/math/eqpart.pdf*, 2013.

A Proof of Lemma 14

Various proofs of this fact are available in [10, 11, 18]. Our exposition is close to [19].

Let $X = (x_{ij})$ be any fractional automorphism of G. We define a directed graph D_X with vertex set $V(G) = \{1, \ldots, n\}$ and edge set

$$E = \{ (i,j) : X_{ij} \neq 0 \} .$$

Let $V = S_1 \cup S_2 \cup \cdots \cup S_t$ be the partition of V into the strongly connected components S_i of the digraph D_X. We first show that there can be no directed edges between the strongly connected components. We can assume, without loss of generality, that the components are topologically sorted: if (u, v) is a directed edge in D_X with $u \in S_i$ and $v \in S_j$ then $i \leq j$. Hence, for $S = S_i \cup S_{i+1} \cup \ldots S_t$ there are no directed edges (u, v) with $u \in S$ and $v \notin S$. It implies that for each $i \in S$ the row sum $\sum_{j \in S} X_{ij} = 1$. Hence all the row sums of the submatrix indexed by S on both rows and columns are 1. Since X is doubly stochastic this forces each column sum of this submatrix is also 1. More precisely, $\sum_{i \in S} X_{ij} = 1$ for $j \in S$. Hence there are *no* edges between S and $\overline{S}$ for each such S. Consequently, there are no edges between the strongly connected components S_i of D_X.

Therefore, relabeling $V(G)$ so that $S_1, S_2, \ldots, S_t$ become intervals of consecutive integers, we can bring X to a block diagonal form

$$X = X_1 \oplus X_2 \oplus \cdots \oplus X_t \tag{17}$$

where each X_i is a doubly stochastic matrix. Note that the underlying directed graph D_{X_i} induced by S_i is strongly connected, which means that each matrix X_i is *irreducible*.

Now, we claim that the strongly connected components S_i form an equitable partition of the graph G. Since X is a fractional automorphism we have $AX = XA$ where A is the adjacency matrix of the relabeled graph G. Then for any pair of components S_i and S_j we have $A_{ij} X_j = X_i A_{ij}$, where A_{ij} denotes the $(i,j)^{th}$ block indexed by S_i on the rows and S_j on columns. Let u denote the all 1's vector of dimension $|S_j|$. Multiplying by u to the right of both sides we obtain

$$A_{ij} u = A_{ij} X_j u = X_i A_{ij} u,$$

since $X_j u = u$. Hence $A_{ij} u$ is an eigenvector of X_j for eigenvalue 1. However, the matrix X_j is is nonnegative and irreducible, hence by the Perron-Frobenius theorem the maximum eigenvalue (which is 1 for a stochastic matrix) has a 1 dimensional eigenspace. Since u is an eigenvector, it follows that $A_{ij} u$ is its scalar multiple. This means that $A_{ij} u = d_{ij} u$, where d_{ij} is the degree of every vertex in S_i in the bipartite graph $G[S_i, S_j]$ if $i \neq j$ and in the subgraph $G[S_i]$ if $i = j$. We conclude that each $G[S_i, S_j]$ is biregular and each $G[S_i]$ is regular.

Thus, the S_i's do yield an equitable partition. Since this partition is a refinement of the coarsest equitable partition $\mathcal{P}_G$, (17) is a block diagonal form of X also with respect to $\mathcal{P}_G$.

B Separations in the Color-Refinement Hierarchy

We recall the chain of inclusions mentioned in Section 5.

$$\mathsf{Discrete} \subset \mathsf{Amenable} \subset \mathsf{Compact} \subset \mathsf{Godsil} \subset \mathsf{Tinhofer} \subseteq \mathsf{Refinable}.$$

The strict inclusions are proved by the following separating examples. The classes Discrete and Amenable are separated, for example, by K_n for $n \geq 2$. The other separations are not so simple.

- The classes Amenable and Compact are separated, for example, by C_n for $n \geq 6$. These graphs are not amenable because they are indistinguishable from a pair of disjoint cycles on same number of vertices. On the other hand, cycles are known to be compact graphs [21, Theorem 2].
- The classes Compact and Godsil are separated by the well-known Petersen graph. Evdokimov, Karpinski, and Ponomarenko [8, Corollary 5.4] prove that the Petersen graph is not compact. They explicitly give a fractional automorphism of the Petersen graph which cannot be written as a convex combination of the automorphisms. It remains to show that the Petersen graph belongs to the class Godsil. The proof is given in Subsection B.1 below.
- The classes Godsil and Tinhofer are separated by the Johnson graphs $J(n, 2)$ for $n \geq 7$. The *Johnson graph* $J(n, k)$ has the k-element subsets of $[n] = \{1, \ldots, n\}$ as vertices; any two of them are adjacent if their intersection consists of $k - 1$ elements. Note that $J(n, 1) = K_n$. Furthermore, the graph $J(n, 2)$ is the line graph of K_n: it has all 2-element subsets of $[n]$ as vertices; any two of them are adjacent if the intersection of the 2-subsets is non-empty. It is noticed in [7] that $J(n, 2)$ is not Godsil for $n \geq 7$. For establishing the separation, we show that $J(n, 2)$ is indeed Tinhofer. The proof is given in Subsection B.2 below.

B.1 The Petersen Graph is Godsil

It is well-known that the Petersen graph, denoted by P, is isomorphic to the Kneser graph $K(5, 2)$. Formally, it means the following. Fix a set $\Omega = \{a, b, c, d, e\}$. Then, the set of all 2-subsets of Ω, i.e. $\{ab, ac, ad, ae, bc, bd, be, cd, ce, de\}$, is the vertex set of P. Two vertices are connected by an edge iff the corresponding 2-subsets are disjoint. An important fact about the Petersen graph is that its automorphism group is isomorphic to the symmetric group acting on the set Ω, denoted by S_Ω (which is S_5). In fact, any automorphism of the Petersen graph can be realized by extending the action of a permutation $\pi \in S_\Omega$ to the graph P and vice-versa [27].

First, we state some useful facts about the Petersen graph.

Lemma 27. *The following statements are true about the Petersen graph.*

(i) There are no cycles of length 3,4 and 7.
(ii) There are no independent sets of size greater than 4.
(iii) Any two adjacent vertices have no common neighbors and any two non-adjacent vertices have a unique common neighbor.

We will need some definitions regarding equitable partitions in graphs. Given a partition $\Sigma = \{S_1, \ldots, S_k\}$, we call the sets $S_1, S_2, \ldots, S_k$ as *cells* of Σ. If the size of a cell is k, we call it a k-*cell* of Σ. Two cells S and S' are said to be *compatible* if the induced bipartite graph $P[S, S']$ is biregular (it can be empty). Otherwise, we call them to be *incompatible*. Recall that in an equitable partition, any cell S induces a regular graph $P[S]$. Moreover, any two cells S, S' are compatible. In that case, the number of edges in the graph $P[S, S']$ is a common multiple of $|S|$ and $|S'|$.

Now we are ready to prove the following theorem.

Theorem 28. *The Petersen graph is* Godsil*.*

Proof. For proof of the theorem, we will enumerate all equitable partitions of P. For each such partition, we describe a subgroup of $Aut(P)$, the orbits of which are the cells in this partition. We will use the Kneser graph definition to describe such subgroups of automorphisms by subgroups of S_Ω.

In is known [28] that, up to automorphisms, P has 10 equitable partitions. In [28], they are obtained with the aid of computer computation. We here perform a case analysis that allows us to find 10 equitable partitions of P, to prove that this list is complete, and to see that each of the partitions is an orbit partition.

The trivial partition of the entire vertex set is clearly an orbit partition, since the Petersen graph is vertex-transitive. For easy enumeration, we classify the non-trivial equitable partitions of P by size of the smallest cell in the partition. Given an equitable partition Σ, let δ be the minimum cardinality of a cell in Σ. Clearly, $\delta \leq 5$. Lemma 29 handles the case $\delta \in \{3, 4, 5\}$. Lemma 31 handles the case $\delta = 2$. Lemma 33 handles the case $\delta = 1$. $\square$

We first handle the case $\delta \in \{3, 4, 5\}$.

Lemma 29. *Let Σ be an equitable partition of P such that $\delta \in \{3, 4, 5\}$. Then, Σ is an orbit partition of P.*

We will require the following claim.

Claim 30. *The following holds for an equitable partition Σ of the Petersen graph.*

(i) $\delta \neq 3$.
(ii) If $\delta = 4$, the partition has one 4-cell S and one 6-cell T. Moreover, $P[S]$ is empty and $P[T]$ is a 3-matching (a matching of size 3).
(iii) If $\delta = 5$, the partition has two 5-cells S and T. Moreover, $P[S]$ and $P[T]$ are 5-cycles.

Proof. (i) Suppose $\delta = 3$. Let S be a 3-cell. Then, any equitable partition can be of two kinds: either $\{S, T\}$ where $|T| = 7$, or $\{S, U, V\}$ where $|U| = 3$, $|V| = 4$. The first case is ruled out since $P[T]$ can never be regular (P has neither independent sets of size 7 nor cycles of size 7). Suppose the second case is possible. Then $P[S]$ and $P[U]$ must be empty (since P has no three cycles). Furthermore, the bipartite graphs $P(S, V]$ and $P[U, V]$ must be both biregular. The graph $P[S, V]$ (likewise, $P[U, V]$) is empty or it has

12 edges. It is not possible that $P[S, V]$ has 12 edges because then $P[V]$ has only 3 edges and cannot be regular. If both $P[S, V]$ and $P[U, V]$ are empty then V is disconnected from the rest of the graph, which is a contradiction. Therefore, $\delta = 3$ is not possible.

(ii) Suppose $\delta = 4$. Then, any equitable partition must be of the kind $\{S, T\}$ where $|S| = 4$ and $|T| = 6$. Moreover, $P[S]$ must be empty (0-regular) or 2-matching (1-regular) since it cannot be a 4-cycle (2-regular). In fact, the case of 2-matching can also be ruled out by counting the number of edges as follows. For S and T to be compatible, there must be 12 edges in the graph $P[S, T]$. Then, there is exactly one edge left in the induced graph $P[T]$ which is impossible. Therefore, $P[S]$ must be empty. Also, this implies that the graph $P[S, T]$ has $4 \times 3 = 12$ edges. Hence, $P[T]$ must be a 3-matching.

(iii) Suppose $\delta = 5$. Then, any equitable partition must be of the kind $\{S, T\}$ where $|T| = 5$. Since P does not have independent sets of size 5, $P[S]$ and $P[T]$ must be 5-cycles. □

Proof of Lemma 29.

(a) Suppose $\delta = 4$. We know by Claim 30 that any equitable partition is of kind $\Sigma = \{S, T\}$ where $|S| = 4$ and $|T| = 6$. Moreover, $P[S]$ is empty and $P[T]$ is a 3-matching. Let us characterize all such partitions. By Kneser graph definition, any 4-independent-set must be of the kind $\{ab, ac, ad, ae\}$. Therefore, all such partitions must be of the kind $S = \{ab, ac, ad, ae\}$ and $T = \{bc, bd, be, cd, ce, de\}$. The partition $\{ab, ac, ad, ae\}$, $\{bc, bd, be, cd, ce, de\}$ can be easily verified to be equitable. Moreover, it is easy to check that it is the orbit partition for the subgroup $S_{\{b,c,d,e\}}$.

(b) Suppose $\delta = 5$. We know by Claim 30 that any equitable partition must be of kind $\Sigma = \{S, T\}$ where $|S| = |T| = 5$. Moreover, $P[S]$ and $P[T]$ are 5-cycles. Clearly, such partitions exist, and any such partition has a matching between sets S and T. It remains to show that this is indeed an orbit partition for some subgroup of $Aut(P)$. Denote the 5-cycle in S by 1-2-3-4-5. Let $1'$ be the matching partner of 1 in T and so on. Now, $1'$ and $2'$ cannot be adjacent, else there is a 4-cycle in P. The unique common neighbor of $1'$ and $2'$ must be $4'$, otherwise it is easy to verify that we will have a 4-cycle in P. The partners $3'$ and $5'$ can also be uniquely determined in T. The permutation $\pi = (12345)(1'2'3'4'5')$ can be verified to be an automorphism of P and the orbits of the subgroup generated by π are precisely $\{S, T\}$. □

Next, we handle the case $\delta = 2$.

Lemma 31. *Let Σ be an equitable partition of P such that $\delta = 2$. Then, Σ is an orbit partition of P.*

We first require the following claim. Let $N(S)$ denote the neighbourhood of a set $S \subset V(P)$. I.e., $N(S) = \{v \notin S : v \in N(u) \text{ for some } u \in S\}$.

Claim 32. *Let S be a cell of cardinality $\delta = 2$ in the partition Σ. Then, $N(S)$ is a cell of Σ.*

Proof. Let $S = \{u, v\}$. We first claim that uv must be an edge. This holds because any two non-adjacent vertices have a unique common neighbor x. The cell containing x can only be a singleton set, which contradicts $\delta = 2$. Since uv is an edge, there are no common neighbors of u and v. Therefore, $|N(S)| = 4$. Moreover, $N(S)$ is an independent set since any edge among vertices in $N(S)$ can be used to construct a 3 or 4 cycle passing through the edge uv (see Figure B.1). This is not possible by Lemma 27.

Next, we show that $N(S)$ is a cell. Let $R = V(P)\backslash(S \cup N(S))$ be the set of remaining vertices as shown in Figure 1. Clearly, $|R| = 4$. Observe a cell cannot contain vertices from both $N(S)$ and R, since it would be incompatible with S. Since $N(S)$ is independent set, there cannot be a 2-cell inside $N(S)$. Clearly, there cannot be 1-cells and hence 3-cells inside $N(S)$. Therefore, $N(S)$ must be a cell. □

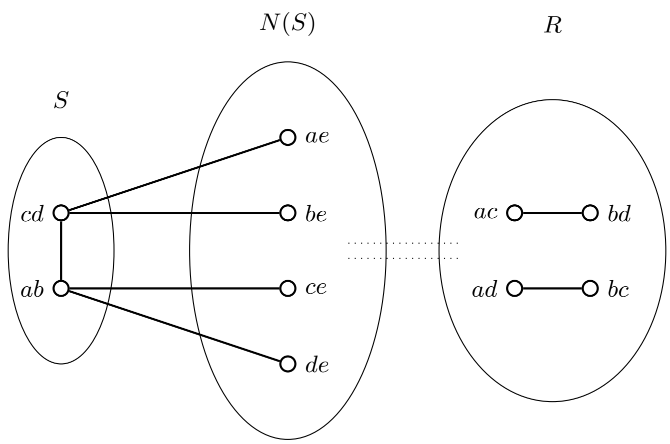

Fig. 1. Case $\delta = 2$.

Proof of Lemma 31. We will classify the equitable partitions by the partition induced on the set R, since S and $N(S)$ are already seen to be cells. By accounting for edges of S and $N(S)$, it is easy to verify that R has exactly two edges, and hence $P[R]$ must be a 2-matching. Since $\delta = 2$, R does not contain any 1-cells and hence, any 3-cells. This leaves us with only two cases.

Case 1: R is a cell. We characterize all such partitions by naming a typical case. W.l.o.g, let $S = \{ab, cd\}$ since S is an edge. Then, $N(S)$ must be $\{ae, be, ce, de\}$ and R must be $\{ac, ad, bc, bd\}$. The partition $\{ab, cd\}, \{ae, be, ce, de\}, \{ac, ad, bc, bd\}$ can be easily verified to be equitable. Moreover, it is easy to check that it is the orbit partition for the subgroup of all permutations in S_Ω which preserve the Ω-partition $\{ab\}, \{cd\}, \{e\}$. This is also the subgroup generated by the automorphisms $(ab), (cd), (ac)(bd)$.

Case 2: The induced partition on R is of the form $\{A, B\}$ where $|A| = |B| = 2$. Each 2-cell has to be an edge. Therefore, the sets A and B must be $\{ac, bd\}$ and $\{bc, ad\}$. The partition $\{ab, cd\}, \{ae, be, ce, de\}, \{ac, bd\}, \{ad, bc\}$ can be easily verified to be equitable. Moreover, it is easy to check that it is the orbit partition for the subgroup of all permutations in S_Ω which preserve the Ω-partition $\{ab\}, \{cd\}, \{e\}$ and additionally, stabilize the sets $\{ac, bd\}$ and $\{ad, bc\}$. This is also the subgroup generated by the automorphisms $(ac)(bd), (ad)(bc), (ab)(cd)$. □

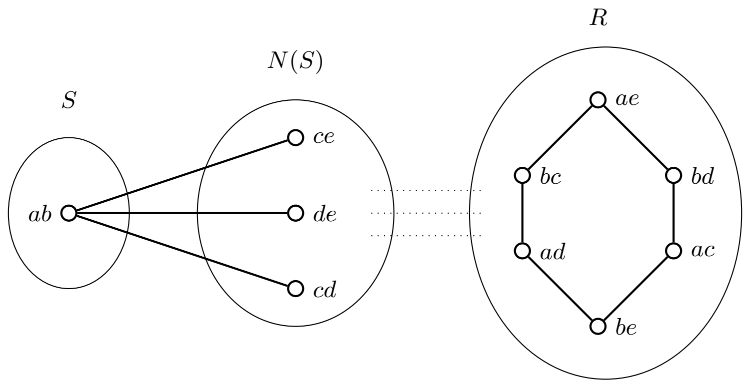

Fig. 2. Case $\delta = 1$.

Finally, we handle the case $\delta = 1$.

Lemma 33. *Let Σ be an equitable partition of P such that $\delta = 1$. Then, Σ is an orbit partition of P.*

Proof. Let S be a singleton set in such an equitable partition. Similar to a previous argument, a cell cannot have vertices from both $N(S)$ and $V \backslash N(S)$. Therefore, any equitable partition refines the partition $S, N(S), R$ (see Figure B.1). Observe that $N(S)$ must be independent set (otherwise there is a 3-cycle). Moreover, if $S = \{ab\}$ in Kneser definition, $N(S)$ must be $\{ce, de, cd\}$ and therefore, $R = \{ae, be, ac, bc, ad, bd\}$ forms a 6-cycle, as shown in the figure. We proceed by further classifying equitable partitions on the basis of the partition induced by them inside $N(S)$. Since $|N(S)| = 3$, we have three possible cases. Either $N(S)$ is a cell, or it contains three singleton cells, or it contains one singleton and one 2-cell.

Case 1: $N(S)$ is a cell. We further classify the equitable partitions in this case on the basis of the partition induced on the set R. First, we examine the possible cells X in R which are compatible with $N(S)$. X cannot be of size 1 or 2, otherwise $P[N(S), X]$ has at most two edges. Also, X cannot be of size $4, 5$ since this would imply a 1 or 2 cell in R. Therefore, either R is a cell, or there are two 3-cells in R.

(a) R is a cell. The partition $\{ab\}$, $\{de, cd, ce\}$, $\{ac, ad, ae, bc, bd, be\}$ can be verified to be an equitable partition. Moreover, it is easy to check that it is the orbit partition for the subgroup $S_{\{c,d,e\}} \times S_{\{a,b\}}$.

(b) The partition induced on R is of form $\{A, B\}$, where $|A| = |B| = 3$. Because of regularity, the only possible 3-cells in R are the independent sets $\{ad, ac, ae\}$ and $\{bc, bd, be\}$. The partition $\{ab\}$, $\{de, cd, ce\}$, $\{ad, ac, ae\}$, $\{bc, bd, be\}$ is clearly equitable. Moreover, it is easy to check that this partition is the orbit partition for the subgroup $S_{\{c,d,e\}}$.

Case 2: $N(S)$ contains 3 singleton cells. Again, we classify the equitable partitions on the basis of the partition induced on the set R. We can check that a cell of size more than two in R will have at least one edge to some singleton in $N(S)$, and will be incompatible with that singleton. Therefore, cells in R must have size at most 2. Moreover, any 2-cell must be of the form $\{ax, bx\}$ for some $x \in \{d, c, e\}$ since all other 2-cells can be seen to be incompatible with some singleton cell in $N(S)$. Finally, it can be seen that every possible 1-cell is incompatible with these three 2-cells. Hence, we can have only two possible cases.

(a) R consists of six singleton cells. The partition is trivially equitable. Moreover, it is easy to check that it is the orbit partition for the subgroup $\{id\}$.

(b) R consists of three cells of size 2, namely $\{ad, bd\}$, $\{ac, bc\}$, $\{ae, be\}$. The partition $\{ab\}$, $\{cd\}$, $\{ce\}$, $\{de\}$, $\{ad, bd\}$, $\{ac, bc\}$, $\{ae, be\}$ can be easily seen to be equitable. Moreover, it is easy to check that it is the orbit partition for the subgroup $S_{\{a,b\}}$.

Case 3: $N(S)$ contains a 2-cell $U = \{ce, de\}$ and a 1-cell $V = \{cd\}$. Again, we need to classify the equitable partitions on the basis of the partition induced on the set R. First, we examine the possible cells X in R which are compatible with U and V. Clearly, X cannot be a 5-cell since $P[X]$ cannot be regular. It cannot be a 3-cell as well since the two candidate 3-cells are the independent sets $\{ad, ac, ae\}$ and $\{bc, bd, be\}$. Neither of them can be compatible with the singleton set V. Also, R cannot be a cell since it is incompatible with the singleton set V. Moreover, the only possible 4-cell is the neighborhood of the set U, i.e. $\{ac, bd, ad, bc\}$. Any other 4-cell is incompatible with U. Overall, we have no 3,5,6 cells in R. Therefore, we have only the following four remaining subcases.

(a) One 4-cell and two 1-cells. This case is not possible since a 1-cell cannot be compatible with a 4-cell.

(b) One 4-cell and one 2-cell. The cells are $\{ac, bd, ad, bc\}$ and $\{ae, be\}$. The partition $\{ab\}$, $\{cd\}$, $\{ce, de\}$, $\{ae, be\}$, $\{ac, bd, ad, bc\}$ can be verified to be an equitable partition. Moreover, it is easy to check that it is the orbit partition for the subgroup $S_{\{a,b\}} \times S_{\{c,d\}}$

(c) Three 2-cells. First, ae and be must be in the same 2-cell, otherwise the cell containing any of them would be incompatible with V. For the remaining vertices ac, ad, bc, bd, we can pair them up in three ways: (i) ac, ad and bc, bd, (ii) ac, bc and ad, bd, or (iii) ac, bd and ad, bc The first case is not possible since $\{ae, be\}$ and $\{ac, ad\}$ are not compatible. The second case is not possible because $\{ac, bc\}$ and $U = \{ce, de\}$ are not compatible. The third case gives an equitable partition $\{ab\}$, $\{cd\}$, $\{ce, de\}$, $\{ae, be\}$, $\{ac, bd\}$, $\{ad, bc\}$. Moreover, it is easy to check that it is the orbit partition for the subgroup generated by $(ab)(cd)$.

(d) A bunch of 1-cells and 2-cells. Clearly, the vertices ac, ad, bc, bd cannot form a singleton cell, since such a 1-cell will not be compatible with U. Therefore, $\{ae\}$ and $\{be\}$ are the only possible singleton cells. Neither of them can pair up with one of ac, ad, bc, bd since that cell would be incompatible with V. Therefore, they are forced to be singleton cells. It remains to partition ac, ad, bc, bd into two 2-cells. The vertex ac cannot be paired up with bd or bc since it will be incompatible with be. Therefore, the only possible case is to have 2-cells $\{ac, ad\}$ and $\{bc, bd\}$. The partition $\{ab\}$, $\{cd\}$, $\{ce, de\}$, $\{ae\}$, $\{be\}$, $\{ac, ad\}$, $\{bc, bd\}$ can be verified to be equitable. Moreover, it is easy to check that it is the orbit partition for the subgroup $S_{\{c,d\}}$. (This case is identical to Case 2(b)). $\square$

B.2 The Johnson Graphs $J(n, 2)$ are Tinhofer

In this section, we show that the Johnson graphs $J(n, 2)$ are Tinhofer. We begin with some necessary definitions. Let G be a graph and $F \subset V(G)$ be the set of fixed vertices of G. Denote the automorphism group of G by A. For $v \in V(G)$, A_v will denote the stabilizer subgroup of v. Furthermore, $A_F = \bigcap_{v \in F} A_v$. Let $\mathcal{P}_F$ denote the coarsest equitable subpartition of the partition of $V(G)$ individualizing each vertex in F. G is Tinhofer iff, for every F, the orbit partition of G with respect to A_F (which is a subpartition of $\mathcal{P}_F$) coincides with $\mathcal{P}_F$.

One way to prove that the two partitions coincide is to show that each orbit O of A_F is definable in terms of F in two-variable first-order logic. "In terms of F" means that a defining formula $\Phi_O(x)$ can use constant symbols (names) for each vertex in F. $\Phi_O(x)$ contains occurrences of only two variables, x and y. At least one occurrence of x is free. $\Phi_O(x)$ uses two binary relation symbols $\sim$ and $=$ for adjacency and equality of vertices. $\Phi_O(x)$ is true on G for $x = v$ iff $v \in O$.

Once $\Phi_O(x)$ is found for each O, the equality of the partitions follows by Immerman-Lander argument [13]. Or directly from the definitions of orbits, as those will imply that any two orbits are separated by color refinement starting from the individualization of F. The number of refinement steps sufficient to separate O from any other orbit can be only one greater than the quantifier depth of $\Phi_O(x)$.

In order to implement this scenario for $G = J(n, 2)$, it will be convenient to assume that $V(G) = \binom{[n]}{2}$ (though the formulas $\Phi_O(x)$ will not involve variables over $[n]$). Given $\alpha \in S_n$, by $\ell(\alpha)$ we denote the corresponding permutation of $\binom{[n]}{2}$. Obviously, every $\ell(\alpha)$ is an automorphism of G. By the Whitney theorem [27], A contains nothing else.

Before designing the definitions $\Phi_O(x)$, we will need to make two preliminary steps:

- describe A_F,
- describe the orbits of A_F (first irrespectively of any logical formalism; expressing these descriptions in two-variable first-order logic will be the next task).

We now proceed to the proof of the theorem.

Theorem 34. *$J(n, 2)$ is a Tinhofer graph for all n.*

Proof. Note that $J(2, 2) = K_1$, $J(3, 2) = K_3$, and $J(4, 2)$ is the octahedral graph, whose complement is $K(4, 2) = 3K_2$. Thus, these three graphs are amenable and, hence, Tinhofer. We can, therefore, assume that $n \geq 5$.

Call a fixed vertex $p \in F$ *isolated* if F contains no vertex adjacent to p. Let $F = F_1 \cup F_2$ be the partition of F into non-isolated and isolated vertices. Furthermore, we define the partition

$$[n] = W_1 \cup W_2 \cup W_3$$

as follows: W_1 is the union of all non-isolated pairs p (i.e., all p in F_1), and W_2 is the union of all isolated pairs p (i.e., all p in F_2).

Note now that $\ell(\alpha) \in A_F$ iff

- $\alpha(w) = w$ for every $w \in W_1$ and

– $\alpha(p) = p$ for every $p \in F_2$.

Given a vertex $u = \{a, b\}$ of G, let $O(u)$ denote its orbit with respect to A_F. There are six kinds of the orbits. Below we describe all of them along with providing suitable formal definitions $\Phi_{O(u)}(x)$.

Case 1: $\{a, b\} \subseteq W_1$. Then $O(u) = \{u\}$. *Formal definition:* $x = u$.

Case 2: $\{a, b\} \subseteq W_2$. Here we have two cases. If $u \in F_2$, then $O(u) = \{u\}$ again. Otherwise, F_2 contains two pairs $p_1 = \{a, a'\}$ and $p_2 = \{b, b'\}$. In this case,

$$O(u) = \{\{a, b\}, \{a', b\}, \{a, b'\}, \{a', b'\}\},$$

which is exactly the common neighborhood of p_1 and p_2. *Formal definition:* $x \sim p_1 \wedge x \sim p_2$.

Case 3: $\{a, b\} \subseteq W_3$. Now $O(u) = \binom{W_3}{2}$, which are exactly the non-fixed vertices with no neighbor in F. *Formal definition:* $\bigwedge_{p \in F}(x \neq p \wedge x \not\sim p)$.

Case 4: $a \in W_1$, $b \in W_2$. In this case, there is $p = \{b, b'\}$ in F_2 and there are $q_1 = \{a, a_1\}$ and $q_2 = \{a, a_2\}$ in F_1. Then

$$O(u) = \{\{a, b\}, \{a, b'\}\}.$$

Formal definition: $x \sim p \wedge x \sim q_1 \wedge x \sim q_2$. Indeed, the condition $x \sim p$ forces x to contain either b or b'. This excludes the possibility that $x = \{a_1, a_2\}$ and, therefore, x is forced to contain a by the adjacency to q_1 and q_2.

Case 5: $a \in W_1, b \in W_3$. Then $O(u) = \{\, \{a, b'\} : b' \in W_3\}$. *Formal definition.* We know that there are $q_1 = \{a, a_1\}$ and $q_2 = \{a, a_2\}$ in F_1. First of all, we say that $x \sim q_1 \wedge x \sim q_2$. It remains to exclude the possibility that $x \subseteq W_1 \cup W_2$ (in particular, this will exclude $x = \{a_1, a_2\}$ and force x to contain a). We do this by adding the following expression

$$\bigwedge_{p \in F} x \neq p \wedge \bigwedge_{p,q \in F, p \not\sim q} \neg(x \sim p \wedge x \sim q) \wedge \bigwedge_{p,q \in F_1, p \sim q} (x \sim p \wedge x \sim q \rightarrow \exists y\, (y \sim x \wedge y \sim p \wedge y \sim q)).$$

The first conjunctive term prevents x to be one of the pairs in F. The second term excludes the case that x is covered by two disjoint pairs p and q in F. The third term excludes the case that x is covered by two intersecting pairs p and q in F or, equivalently, the case where x, p, and q form a triangle. It would be not enough just to forbid x, p, and q from forming a clique because this could also exclude a permissible case where x, p, and q form a star (which is captured by the subformula beginning with $\exists y$). Note, that we need $n \geq 5$ in this place.

Case 6: $a \in W_2$, $b \in W_3$. In this case, F_2 contains a pair $p = \{a, a'\}$ and

$$O(u) = \{\, \{a, b'\} : b' \in W_3\} \cup \{\, \{a', b'\} : b' \in W_3\}\,.$$

Formal definition: $x \sim p \wedge x \not\subseteq W_1 \cup W_2$, the latter being expressed as in the preceding case.

The proof is complete. $\square$

ECCC ISSN 1433-8092

http://eccc.hpi-web.de