On the Power of Color Refinement

V. Arvind[1], Johannes Köbler[2(✉)], Gaurav Rattan[1], and 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, 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. We determine the exact range of applicability of color refinement by showing that amenable graphs are 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.

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. Nevertheless, color refinement turns out to be a useful tool not only in isomorphism testing but also in a number of other areas; see [9,12,17] and references there.

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.

This work was supported by the Alexander von Humboldt Foundation in its research group linkage program. The second author and the fourth author were supported by DFG grants KO 1053/7-2 and VE 652/1-2, respectively.

A. Kosowski and I. Walukiewicz (Eds.): FCT 2015, LNCS 9210, pp. 339–350, 2015.
DOI: 10.1007/978-3-319-22177-9_26

Immerman and Lander [10], in their highly influential paper, established a close connection between color refinement and 2-variable first-order logic with counting quantifiers. They show that color refinement distinguishes G and H if and only if 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 a fractional isomorphism (introduced by Tinhofer [18] using the term doubly stochastic isomorphism), turns out to be equivalent to color refinement. Building on Tinhofer's work [18], it is shown by Ramana, Scheinerman and Ullman [16] (see also Godsil [8]) 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., [5,19].
2. Trees are amenable (Edmonds [6,20]).
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. Babai, Erdös, and Selkow [2] have shown that a random graph $G_{n,1/2}$ has this property with high probability. Moreover, the discrete partition of $G_{n,1/2}$ is reached within at most two refinement steps. This implies that graph isomorphism is solvable very efficiently in the average case (see also [3]).

Our Contribution. 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. This result is shown in Sect. 5, after we unravel the structure of amenable graphs in Sects. 3 and 4. 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 [10] 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.

Related Work. In an accompanying paper [1], we use our characterization of amenable graphs to prove that the polytope of fractional automorphisms of an amenable graph is integral. A characterization of amenable graphs similar to that in the present paper has been suggested independently by Kiefer, Schweitzer, and Selman [13]. Moreover, they obtain a generalization of this result to arbitrary relational structures (which includes, in particular, directed graphs).

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 *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' if and only if 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 Basic Definitions and Facts

Throughout the paper, we consider vertex-colored graphs. A *vertex-colored graph* is an undirected simple graph G endowed with a vertex coloring $c : V(G) \to \{1, \ldots, k\}$. Isomorphisms between vertex-colored graphs are required to preserve vertex colors. We get usual graphs when c is constant.

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 the vertex coloring of G, i.e., $C^0(u) = c(u)$. Then,

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

where $\{\!\!\{\ldots\}\!\!\}$ denotes a multiset.

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 called the *stable partition* of G and denoted by $\mathcal{P}_G$.

Given a partition $\mathcal{P}$ of the vertex set of a graph G, we call its elements *cells*. We call $\mathcal{P}$ *equitable* if:

(i) Each cell $X \in \mathcal{P}$ is monochromatic, i.e., all vertices $u, v \in X$ have the same color $c(u) = c(v)$.
(ii) For any cell $X \in \mathcal{P}$ the graph $G[X]$ induced by X is *regular*, that is, all vertices in $G[X]$ have equal degrees.
(iii) For any two cells $X, Y \in \mathcal{P}$ the bipartite graph $G[X, Y]$ induced by X and Y is *biregular*, that is, all vertices in X have equally many neighbors in Y and vice versa.

It is easy to see that the stable partition of G is equitable; our analysis in the next section will make use of this fact.

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 the stable partition of the disjoint union of G and H. In fact, Equality (2) holds true for all i if it is true for $i = n$; see, e.g., [15]. 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 for every graph H, procedure CR works correctly on the input pair G and H. That is, Equality (2) is false for $i = n$ whenever $H \not\cong G$.*

3 Local Structure of Amenable Graphs

Consider the stable partition $\mathcal{P}_G$ of an amenable graph G. The following lemma gives a list of all possible regular and biregular graphs that can occur, respectively, as $G[X]$ and $G[X, Y]$ for cells X, Y of $\mathcal{P}_G$.

Lemma 3. *The stable partition $\mathcal{P}_G$ of an amenable graph G fulfills the following properties:*

- **(A)** *For any cell $X \in \mathcal{P}_G$, $G[X]$ is an empty graph, a complete graph, a matching graph mK_2, the complement of a matching graph, or the 5-cycle;*
- **(B)** *For any two cells $X, Y \in \mathcal{P}_G$, $G[X, Y]$ is an empty graph, a complete bipartite graph, a 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 [11]). *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 $n = 5$.*[1]

[1] The last case, in which the graph is the 5-cycle, is missing from the statement of this result in [11, Theorem 2.12]. The proof in [11] tacitly considers only graphs with at least 6 vertices.

Lemma 5 (Koren [14]). *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 $d \in \{0, 1, m-1, m\}$.*

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 Lemmas 4 and 5 imply that this subgraph can be replaced by a non-isomorphic regular or biregular graph with the same parameters. Hence, in order to prove Lemma 3 it 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 colors $C_G^i(u)$ and $C_H^i(u)$ in the two graphs.

Lemma 6. *Let X and Y be cells of the stable partition of a graph G.*

(i) If H is obtained from G by replacing the edges of the subgraph $G[X]$ with the edges of an arbitrary regular graph of the same degree on the same vertex set X, then $C_G^i(u) = C_H^i(u)$ for any $u \in V(G)$ and any i.
(ii) If H is obtained from G by replacing the edges of the subgraph $G[X, Y]$ with the edges of an arbitrary biregular graph with the same vertex partition such that the vertex degrees remain unchanged, then $C_G^i(u) = C_H^i(u)$ for any $u \in V(G)$ and any i.

Proof of Lemma 3. **(A)** If $G[X]$ is a graph not from the list, by Lemma 4, it is not a unigraph. Hence, we 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, implying that CR does not distinguish 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 condition follows, similarly to Condition **A**, from Lemma 5 and Part (ii) of Lemma 6. □

4 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 G to be the complete graph on the vertex set $\mathcal{P}_G$ with the following labeling of vertices and edges. A vertex X of $C(G)$ is called *homogeneous* if the graph $G[X]$ is complete or empty and *heterogeneous* otherwise. An edge $\{X, Y\}$ of $C(G)$ is called *isotropic* if the bipartite graph $G[X, Y]$ is either complete or empty and *anisotropic* otherwise. A path $X_1 X_2 \ldots X_l$ 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) is called *uniform*.

For graphs fulfilling Conditions **A** and **B** of Lemma 3 we refine the labeling of the vertices and edges of $C(G)$ as follows. A heterogeneous cell $X \in \mathcal{P}_G$ is called *matching*, *co-matching*, or *pentagonal* depending on the type of $G[X]$. Note that a matching or co-matching cell X always consists of at least 4 vertices. Further, an anisotropic edge $\{X, Y\}$ is called *constellation* if $G[X, Y]$ is a disjoint union of stars, and *co-constellation* otherwise (in the latter case, the bipartite complement of $G[X, Y]$ is a disjoint union of stars). Likewise, homogeneous cells X (and isotropic edges $\{X, Y\}$) are called *empty* if the graph $G[X]$ (resp. $G[X, Y]$) is empty, and *complete* otherwise.

Note that if an edge $\{X, Y\}$ of a uniform path or cycle is a constellation, then $G[X, Y]$ is a matching graph.

Lemma 7. *The cell graph $C(G)$ of an amenable graph G has the following properties:*

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

Proof. **(C)** Suppose that P is a uniform anisotropic path in $C(G)$ connecting two heterogeneous cells 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 $G[Y]$.

We construct a non-isomorphic graph H such that CR does not distinguish between G and H. Since X and Y are heterogeneous, we can replace the edges of the subgraph $G[X]$ with the edges of an isomorphic but different graph on the same vertex set X such that P is a conducting path in the resulting graph H if and only if P is a non-conducting path in G. Now, Part (i) of Lemma 6 implies that CR computes the same coloring for G and H and does not distinguish between them. On the other hand, Lemma 1 implies that any isomorphism ϕ between G and H must map each cell to itself. Since $\phi(v^*) = \phi(v)^*$, ϕ must also preserve the conducting property along the path P. It follows that G and H are not isomorphic. Hence, G is not amenable.

(D) Suppose that $C(G)$ contains a uniform anisotropic cycle Q of length m. All cells in Q have the same cardinality; 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 cells 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. In particular, we can replace the subgraph $G[X, Y]$ in such a way that $\tau(Q)$ is changed.

Similarly as for Condition **C**, we use Part (ii) of Lemma 6 to argue that CR does not distinguish between G and H. Furthermore, $G \not\cong H$ 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 $P = XY_1 \dots Y_l Z$ such that $|X| < |Y_1| = \dots = |Y_l| > |Z|$ (for the case of a cycle, where $Z = X$, the argument is virtually the same). Let $G[X, Y_1] = sK_{1,t}$ and $G[Z, Y_l] = aK_{1,b}$, 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 $|Y_1| = |Y_l| = st = ab$.

Like in the proof of Condition**C**, the uniform anisotropic path $Y_1 \dots Y_l$ determines a one-to-one correspondence between the cells Y_1 and Y_l that can be used to make the identification $Y_1 = Y_l = \{1, 2, \dots, st\} = Y$. 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 for any $x \neq x'$ in X and $z \neq z'$ in Z,

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

We regard $\mathcal{Y}_G = \{\, Y_x \,\}_{x \in X} \cup \{\, Y_z \,\}_{z \in Z}$ as a hypergraph on the vertex set Y. Note that $\mathcal{Y}_G$ has multiple hyperedges if $Y_x = Y_z$ for some x and z. Without loss of generality, we can assume that the hyperedges Y_z, $z \in Z$, form consecutive intervals in Y. We call the anisotropic path P *flat*, if there exists no pair $(x, z) \in X \times Z$ such that one of the two hyperedges Y_x and Y_z is contained in the other.

We construct a non-isomorphic graph H such that CR does not distinguish between G and H. If P is flat in G, we replace the edges of the subgraph $G[X, Y_1]$ by the edges of an isomorphic but different biregular graph such that P becomes non-flat in the resulting graph H. More precisely, we replace the edges in such a way that all hyperedges of $\mathcal{Y}_H$ form consecutive intervals in Y by letting $\mathcal{Y}_H = \{\, Y_i \,\}_{i \in [s]} \cup \{\, Y_z \,\}_{z \in Z}$, where $Y_i = \{(i-1)t + 1, \dots, it\}$. Likewise, if P is non-flat in G, we replace the edges of $G[X, Y_1]$ such that P becomes flat in H by letting $Y_i = \{i, i+s, \dots, i + (t-1)s\}$.

Now, Part (i) of Lemma 6 implies that CR computes the same coloring for G and H and does not distinguish between them. On the other hand, Lemma 1 implies that any isomorphism ϕ between G and H must map each cell to itself. As ϕ must also preserve the flatness property of the path P, it follows that G and H are not isomorphic. Hence, G is not amenable.

(F) Suppose that $C(G)$ contains an anisotropic path $XY_1 \dots Y_l$ where $|X| < |Y_1| = \dots = |Y_l|$ and Y_l is heterogeneous. 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. Like in the proof of Condition **E**, the uniform anisotropic path $Y_1 \dots Y_l$ determines a one-to-one correspondence between the cells Y_1 and Y_l that can be used to make the identification $Y_1 = Y_l = \{1, 2, \dots, st\} = Y$. Consider the hypergraph $\mathcal{Y}_G = \{Y_x\}_{x \in X} \cup \mathcal{E}$, where $Y_x = N(x) \cap Y_1$ and $\mathcal{E}$ consists of the pairs of adjacent vertices in $G[Y_l]$. Now, exactly as in the proof of Condition **E**, we can change the isomorphism type of $\mathcal{Y}_G$ by replacing the edges of the subgraph $G[X, Y_1]$ by the edges of an isomorphic biregular graph. This yields a non-isomorphic graph H that is indistinguishable from G by CR. □

It turns out that Conditions **A**–**F** are not only necessary for amenability (as shown in Lemmas 3 and 7) but also sufficient. As a preparation we first prove the following Lemma 8 that reveals 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 8. *Suppose that a graph G satisfies Conditions* **A**–**F**. *Then for any anisotropic component A of $C(G)$, the following is true.*

- **(G)** *A is a tree with the following monotonicity property. Let R be a cell in A of minimum cardinality and let A_R be the rooted directed tree obtained from A by rooting A at R. Then $|X| \leq |Y|$ for any directed edge (X, Y) of A_R.*
- **(H)** *A contains at most one heterogeneous vertex. If R is such a vertex, it has minimum cardinality among the cells of A.*

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

(H) Assume that A contains more than one heterogeneous cell. Consider two such cells S and T. Let $S = Z_1, Z_2, \dots, Z_l = T$ be the path from S to T in A. The monotonicity property stated in Condition **G** implies that there is j (possibly $j = 1, l$) such that $|Z_1| \geq \dots \geq |Z_j| \leq \dots \leq |Z_l|$. Since the path cannot be uniform by Condition **C**, at least one of the inequalities is strict. However, this contradicts Condition **F**.

Suppose that R is a heterogeneous cell in A. Consider now a path $R = Z_1, Z_2, \dots, Z_l = S$ in A where S is a cell with the smallest cardinality. By the monotonicity property and Condition **F**, this path must be uniform, proving that $|R| = |S|$. □

In combination with Conditions **A** and **B**, Conditions **G** and **H** on anisotropic components give a very stringent characterization of amenability.

Theorem 9. *For a graph G the following conditions are equivalent:*

(i) G is amenable.
(ii) G satisfies Conditions **A**–**F**.

(iii) G satisfies Conditions **A, B, G** *and* **H**.

Proof. It only remains to show that any graph G fulfilling the Conditions **A, B, G** and **H** is amenable. Let H be a graph indistinguishable from G by CR. Then we have to show that G and H are isomorphic.

Consider the coloring C^s corresponding to the stable partition $\mathcal{P}^s$ of the disjoint union $G + H$. Since G and H satisfy Equality (2) for $i = s$, there is a bijection $f : \mathcal{P}_G \to \mathcal{P}_H$ matching each cell X of the stable partition of G to the cell $f(X) \in \mathcal{P}_H$ such that the vertices in X and $f(X)$ have the same C^s-color. Moreover, Equality (2) implies that $|X| = |f(X)|$. We claim that for any cells X and Y of G,

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

implying that f is an isomorphism from $C(G)$ to $C(H)$.

Indeed, since X and $f(X)$ are cells of the stable partitions $\mathcal{P}_G$ and $\mathcal{P}_H$, both $G[X]$ and $H[f(X)]$ are regular. Since $X \cup f(X)$ is a cell of the stable partition $\mathcal{P}^s$ of $G + H$, the graphs $G[X]$ and $H[f(X)]$ have the same degree. By Condition **A**, $G[X]$ is a unigraph, implying Property (a). Property (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 Condition **H**, an anisotropic component A of the cell graph $C(G)$ contains at most one heterogeneous cell. Denote it by R_A if it exists. Otherwise fix R_A to be an arbitrary cell of the minimum cardinality 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 Condition **G**, 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. By the monotonicity property in Condition **G** and our choice of R, we can assume that $|X| \leq |Y|$. Then ϕ can be extended to Y so that this is an isomorphism from $G[X,Y]$ to $H[f(X), f(Y)]$. This is possible by (b) due to the fact that all vertices in Y have degree 1 in $G[X,Y]$ or its bipartite complement (and the same holds for all vertices in $f(Y)$ in the graph $H[f(X), f(Y)]$).

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. □

5 Examples and Applications

Theorem 9 is a convenient tool for verifying amenability. For example, amenability of discrete graphs is a well-known fact. Recall that those are graphs whose stable partitions consist of singletons. Since the cell graph has no anisotropic edge in this case, any anisotropic component of a discrete graph consists of a single cell. Hence, Conditions **A** and **B** as well as Conditions **G** and **H** on anisotropic components are fulfilled by trivial reasons.

Checking these four conditions, we can also reprove the amenability of trees. Moreover, our argument extends to the class of forests. Note in this respect that the class of amenable graphs is not closed under disjoint unions. For example, $C_3 + C_4$ is indistinguishable by CR from C_7 and, hence, is not amenable.

Corollary 10. *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 disjoint unions of stars.

Let $C^*(G)$ be the version of the cell graph $C(G)$ where all empty edges are removed. 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. To prove the monotonicity property in Condition **G**, it suffices to show that $C(G)$ cannot contain an anisotropic path $XY_1 \ldots Y_l Z$ with $|X| < |Y_1| = \cdots = |Y_l| > |Z|$. But this easily follows since in this case each vertex of the induced subgraph $G[X \cup Y_1 \cup \ldots \cup Y_l \cup Z]$ has degree at least 2 in G, contradicting the acyclicity of G.

To prove Condition **H**, suppose that $C(G)$ contains an anisotropic path $X_0, X_1, \ldots, X_l$ connecting two heterogeneous cells X_0 and X_l. Then each vertex of the induced subgraph $G[X_0 \cup X_1 \cup \ldots \cup X_{l-1} \cup X_l]$ has degree at least 2 in G, a contradiction. The same contradiction arises if such a path connects a heterogeneous cell X_0 with an arbitrary cell X_l, where $|X_l| < |X_{l-1}|$. Hence, X_0 must have minimum cardinality among all cells belonging to the same anisotropic component. □

Our characterization of amenable graphs via Conditions **A**, **B**, **G** and **H** 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 [7]; see also [4]) 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 11. *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 $\mathcal{P}_G = \{X_1, \ldots, X_k\}$ of the input graph G. Let $C^*(G)$ be the version of the cell graph $C(G)$ where all empty edges are removed. We can compute the adjacency list of each vertex X_i of $C^*(G)$ by traversing the adjacency list of an arbitrary vertex $u \in X_i$ and listing all cells X_j that contain a vertex v adjacent to u. Simultaneously, we compute for each pair (i, j) such that $i = j$ or $\{X_i, X_j\}$ is an edge of $C^*(G)$ the number d_{ij} of neighbors in X_j of any vertex in X_i. Knowing the numbers $|X_i|$, $|X_j|$ and d_{ij} allows us to determine whether all the subgraphs $G[X_i]$ and $G[X_i, X_j]$ fulfill Conditions **A** and **B** of Lemma 3.

To check Conditions **G** and **H** we use breadth-first search in the graph $C^*(G)$ to find all anisotropic components A of $C(G)$ and, simultaneously, to check that each component A is a tree containing at most one heterogeneous cell. If we restart the search from an arbitrary cell in A having minimum cardinality, we can also check for each forward edge of the resulting search tree whether the monotonicity property of Condition **G** is fulfilled. $\square$

We conclude 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 [10] 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 11 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.

References

1. Arvind, V., Köbler, J., Rattan, G., Verbitsky, O.: On Tinhofer's linear programming approach to isomorphism testing. In: In: Proceedings of the 40th International Symposium on Mathematical Foundations of Computer Science (MFCS), Lecture Notes in Computer Science. Springer (2015) (to appear)
2. Babai, L., Erdös, P., Selkow, S.M.: Random graph isomorphism. SIAM J. Comput. **9**(3), 628–635 (1980)
3. Babai, L., Kucera, L.: Canonical labelling of graphs in linear average time. In: Proceedings of the 20th Annual Symposium on Foundations of Computer Science, pp. 39–46, (1979)
4. Berkholz, C., Bonsma, P., Grohe, M.: Tight lower and upper bounds for the complexity of canonical colour refinement. In: Bodlaender, H.L., Italiano, G.F. (eds.) ESA 2013. LNCS, vol. 8125, pp. 145–156. Springer, Heidelberg (2013)
5. Borri, A., Calamoneri, T., Petreschi, R.: Recognition of unigraphs through superposition of graphs. J. Graph Algorithms Appl. **15**(3), 323–343 (2011)
6. Busacker, R., Saaty, T.: Finite Graphs and Networks: An Introduction with Applications. International Series in Pure and Applied Mathematics. McGraw-Hill Book Company, New York (1965)

7. Cardon, A., Crochemore, M.: Partitioning a graph in $O(|A| \log_2 |V|)$. Theor. Comput. Sci. **19**, 85–98 (1982)
8. Godsil, C.: Compact graphs and equitable partitions. Linear Algebra Appl. **255**(13), 259–266 (1997)
9. Grohe, M., Kersting, K., Mladenov, M., Selman, E.: Dimension reduction via colour refinement. In: Schulz, A.S., Wagner, D. (eds.) ESA 2014. LNCS, vol. 8737, pp. 505–516. Springer, Heidelberg (2014)
10. Immerman, N., Lander, E.: Describing graphs: a first-order approach to graph canonization. In: Selman, A.L. (ed.) Complexity Theory Retrospective, pp. 59–81. Springer, Heidelberg (1990)
11. Johnson, R.: Simple separable graphs. Pac. J. Math. **56**, 143–158 (1975)
12. Kersting, K., Mladenov, M., Garnett, R., Grohe, M.: Power iterated color refinement. In: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, pp. 1904–1910. AAAI Press (2014)
13. Kiefer, S., Schweitzer, P., Selman, E.: Graphs identified by logics with counting. In:Graphs identified by logics with counting. In: Proceedings of the 40th International Symposium on Mathematical Foundations of Computer Science (MFCS), Lecture Notes in Computer Science. Springer (2015) (to appear)
14. Koren, M.: Pairs of sequences with a unique realization by bipartite graphs. J. Comb. Theor. Series B **21**(3), 224–234 (1976)
15. Krebs, A., Verbitsky, O.: Universal covers, color refinement, and two-variable logic with counting quantifiers: lower bounds for the depth. In: Proceedings of the 30-th ACM/IEEE Annual Symposium on Logic in Computer Science (LICS), IEEE Computer Society (2015) (to appear)
16. Ramana, M.V., Scheinerman, E.R., Ullman, D.: Fractional isomorphism of graphs. Discrete Math. **132**(1–3), 247–265 (1994)
17. Shervashidze, N., Schweitzer, P., van Leeuwen, E.J., Mehlhorn, K., Borgwardt, K.M.: Weisfeiler-Lehman graph kernels. J. Mach. Learn. Res. **12**, 2539–2561 (2011)
18. Tinhofer, G.: Graph isomorphism and theorems of Birkhoff type. Computing **36**, 285–300 (1986)
19. Tyshkevich, R.: Decomposition of graphical sequences and unigraphs. Discrete Math. **220**(1–3), 201–238 (2000)
20. Valiente, G.: Algorithms on Trees and Graphs. Springer, Heidelberg (2002)