

# Graphs Identified by Logics with Counting

Sandra Kiefer<sup>(✉)</sup>, Pascal Schweitzer, and Erkal Selman

RWTH Aachen University, Aachen, Germany  
{kiefer,schweitzer,selman}@informatik.rwth-aachen.de

**Abstract.** We classify graphs and, more generally, finite relational structures that are identified by  $C^2$ , that is, two-variable first-order logic with counting. Using this classification, we show that it can be decided in almost linear time whether a structure is identified by  $C^2$ . Our classification implies that for every graph identified by this logic, all vertex-colored versions of it are also identified. A similar statement is true for finite relational structures.

We provide constructions that solve the inversion problem for finite structures in linear time. This problem has previously been shown to be polynomial time solvable by Martin Otto. For graphs, we conclude that every  $C^2$ -equivalence class contains a graph whose orbits are exactly the classes of the  $C^2$ -partition of its vertex set and which has a single automorphism witnessing this fact.

For general  $k$ , we show that such statements are not true by providing examples of graphs of size linear in  $k$  which are identified by  $C^3$  but for which the orbit partition is strictly finer than the  $C^k$ -partition. We also provide identified graphs which have vertex-colored versions that are not identified by  $C^k$ .

## 1 Introduction

The  $k$ -variable fragment of counting logic, denoted by  $C^k$ , is obtained from first-order logic by adding counting quantifiers but only allowing formulas that use at most  $k$  variables. These finite variable logics play a central role in the area of model-checking since for them the model-checking problem and the equivalence problem are solvable in polynomial time (see [11]). For a while, there was the hope that for some fixed  $k$  the logic  $C^k$  can distinguish every pair of non-isomorphic graphs. This would imply that the graph isomorphism problem is solvable in polynomial time. However, in 1992, it was shown by Cai, Fürer and Immerman [6] that  $\Omega(n)$  variables are required to identify all graphs on  $n$  vertices. Since the examples presented in that paper consist of graph isomorphism instances which are actually known to be solvable in polynomial time, this also shows that  $C^k$  does not capture polynomial time.

Concerning  $C^k$ , there are striking connections to other seemingly unrelated areas. For example, there exist several Ehrenfeucht-Fraïssé type games characterizing  $C^k$  [6, 8, 13]. Also strongly related is the  $(k - 1)$ -dimensional version of a well-known color refinement algorithm, named after Weisfeiler and Lehman by

Babai (see [6]). It turns out that the  $(k-1)$ -dimensional Weisfeiler-Lehman algorithm does nothing else but partition  $(k-1)$ -tuples of vertices according to their  $C^k$ -types. Another surprising connection exists to linear programming. The  $k$ -th level of the Sherali-Adams hierarchy of a natural linear integer programming formulation of graph isomorphism essentially corresponds to the expressive power of  $C^k$  [2, 12].

Among the finite variable logics, the fragment  $C^2$  has been of particular interest because its satisfiability problem is known to be decidable [10], and the complexity of the decision problem has been studied extensively [22]. Numerous results for this logic are known, we refer the reader to a survey by Grädel and Otto [9]. In practice, due to its strength and the fact that it can be evaluated in almost linear time, the logic  $C^2$  (more specifically, the corresponding 1-dimensional Weisfeiler-Lehman algorithm) is an essential subroutine in all competitive canonical labeling tools (see [20]). Very recent results concerning  $C^2$  include a paper by Krebs and Verbitsky studying the quantifier depth of  $C^2$ -formulas for  $C^2$ -equivalence classes of graphs [17]. Kopczynski and Tan show that for every fixed  $C^2$ -formula, the set of those  $n$  for which there is a structure with a universe of size  $n$  satisfying the formula is semilinear [16]. Moreover, they describe the characteristics of the spectrum of a  $C^2$ -formula.

While most of the results further above deal with the problem of distinguishing two graphs from each other using finite variable counting logics, in this paper we are concerned with the concept of distinguishing a graph from every other non-isomorphic graph. We say that the graph is **identified** by the logic. More formally, a graph (or a finite relational structure)  $G$  is identified by a logic  $L$  if there is a sentence  $\varphi$  in  $L$  such that  $G$  satisfies  $\varphi$  and every graph (or finite relational structure) which satisfies  $\varphi$  is isomorphic to  $G$ .

Of course every graph is identified by some first-order sentence. However, by [6], as mentioned above, there is no  $k \in \mathbb{N}$  such that every graph is identified by some formula in  $C^k$ . Let us focus on the case  $k = 2$ . It is not difficult to see that all forests are identified by  $C^2$ . Moreover, a graph is asymptotically almost surely identified by  $C^2$ , that is, the fraction of graphs of size  $n$  which are not identified by  $C^2$  tends to 0 as  $n$  tends to infinity [3]. Even more strongly, it is known [4] that the fraction of graphs which are not identified is exponentially small in  $n$ . Similarly, a regular graph is asymptotically almost surely identified by  $C^3$  [18]. However, not all graphs are identified. (For  $C^2$ , consider a cycle of length at least 6, for example.) The following question arises.

*What is the structure of graphs that are identified by  $C^k$ ?*

*Our results.* We study graphs that are identified by  $C^2$  and provide a complete classification for them. This classification can be used to draw several conclusions about general properties of identified graphs. For example, one can derive that if an undirected graph is identified by  $C^2$ , then the  $C^2$ -partition classes of the vertices are exactly the orbits of the automorphism group of the graph. This corollary is neither true when considering finite (relational) structures nor when considering  $C^k$  with  $k > 2$ . For  $C^2$ , we also conclude that if an undirected graph is identified, every vertex-colored version of it is identified by  $C^2$  as well.

This statement holds for finite relational structures, too, but is again not true for  $C^k$  with  $k > 2$ . Using our classification, we show that in time  $O((n+m) \log n)$  it is possible to determine whether an undirected graph is identified by  $C^2$ .

The proof of the correctness of our classification hinges upon explicit constructions that solve the inversion problem and the canonization problem for  $C^2$ . The inversion problem asks whether to a certain invariant a graph (or more generally, a finite structure) can be constructed. In the case of  $C^2$ , such an invariant is the count of the  $C^2$ -types of pairs of vertices. A celebrated result by Otto [21] shows that the inversion problem and the canonization problem for  $C^2$  can be solved in polynomial time. As a side-product, our direct constructions provide an alternative proof for this. In fact, we show that the inversion problem for  $C^2$  can be solved in linear time. Our constructions make use of circulant graphs and doubly-circulant graphs. With these, we observe that every  $C^2$ -equivalence class contains a graph whose  $C^2$ -partition classes are the orbits. More strongly, there is a single automorphism of the graph witnessing this. (That is, there is an automorphism  $\varphi$  such that for all pairs of vertices  $v, v'$  that are in the same  $C^2$ -partition class there is an integer  $i$  such that  $\varphi^i(v) = v'$ .) To achieve inversion for finite structures, we use an old 1-factorization construction due to Walecki (see [19]) that decomposes the complete graph  $K_{2n}$  into  $2n - 1$  disjoint perfect matchings.

Building on the classification of graphs identified by  $C^2$ , we also classify finite structures that are identified by  $C^2$ . For graphs, there is only one special case that may appear within a  $C^2$ -partition class (namely the cycle of length 5). However, for finite structures there are 7 different special cases for a  $C^2$ -partition class, which are of sizes 3, 4, 5 and 6. Our classification theorem describes how these may be combined to form structures that are identified by  $C^2$ . Due to the nature of the different special cases, the classification is more involved (see Theorem 4). Nevertheless, we show that one can decide in almost linear time whether a structure is identified by  $C^2$ . Our characterization also provides a graph-theoretical classification result. From the class of (possibly edge-colored partially oriented) graphs, it explicitly lists all those (color-)regular graphs that are determined up to isomorphism by their (color) degrees, see Theorem 3.

For the logics  $C^k$  with  $k > 2$  we collect several negative results. One can first observe that the triangular graphs form an infinite non-trivial class of strongly regular graphs which are identified by  $C^3$ , implying that any classification result would have to include non-trivial infinite families [7, 14].

Contrasting our results for  $C^2$ , we provide examples of graphs that are identified by  $C^3$  but for which conclusions analogous to the ones mentioned above do not hold. More specifically, we present graphs identified by  $C^3$  for which even the logic  $C^k$  with  $k$  linear in the size of the graph does not correctly determine the orbit partition. This yields graphs which the logic  $C^k$  identifies, but for which not all vertex-colored versions are identified by  $C^k$ . These ideas are based on the construction by Cai, Fürer and Immerman [6].

The existence of these graphs highlights an important fact. Even if a graph is identified by the logic  $C^k$ , it is not clear that it is possible to take advantage

of that in order to canonize the graph. This stands in contrast to a remark in [6] claiming that a graph  $G$  identified by  $C^k$  can be canonized in polynomial time. In fact, the crucial property required for the approach hinted at there to be successful is that all vertex-colored versions of  $G$  need to be identified by  $C^k$ . Indeed, if this property holds then a standard recursive individualization approach canonizes the graph  $G$ . It would suffice to show that for all vertex-colored versions of  $G$  the orbits are determined by the logic. However, with a slight alteration of our construction, we obtain a graph  $G$  that is identified by  $C^k$  and whose orbits are correctly determined, but for which there exist vertex-colored versions whose orbits are not correctly determined.

Independently of our work, Arvind, Köbler, Rattan and Verbitsky [1] have investigated the structure of undirected graphs identified by  $C^2$  obtaining results similar to the ones we provide in Sect. 4. Throughout the document, proofs are omitted. For these, we refer to the full version [15].

## 2 Preliminaries

Unless specified otherwise, a **graph**  $G$  is a finite undirected graph without loops, its vertex set is  $V(G)$  and its edge set  $E(G)$ . For  $P \subseteq V(G)$ , the subgraph induced by  $P$  is  $G[P]$ . If all vertices have degree  $k$  then  $G$  is  **$k$ -regular**. A  **$(k, \ell)$ -biregular graph  $G$  on bipartition  $(P, Q)$**  is a graph on vertex set  $P \dot{\cup} Q$  such that  $P$  and  $Q$  are independent sets, every vertex in  $P$  has exactly  $k$  neighbors in  $Q$  and every vertex in  $Q$  has exactly  $\ell$  neighbors in  $P$ . Two vertices  $v, v' \in V(G)$  are in the same **orbit of  $G$**  if  $G$  has an automorphism  $\varphi$  such that  $\varphi(v) = v'$ . The **orbit partition** of  $G$  is the partition of  $V(G)$  into the orbits of  $G$ .

A graph  $G$  is **identified** by a logic  $L$  if there is a sentence  $\varphi$  in  $L$  such that  $G$  satisfies  $\varphi$  and every graph which satisfies  $\varphi$  is isomorphic to  $G$ . A logic  $L$  **distinguishes** two graphs  $G$  and  $G'$  if there is a sentence  $\varphi$  in  $L$  such that  $G \models \varphi$  and  $G' \not\models \varphi$ . We say that  $G$  and  $G'$  are  **$L$ -equivalent** if they are not distinguished by  $L$ .

The  **$k$ -variable counting logic**, denoted by  $C^k$ , is the  $k$ -variable fragment of first-order logic enriched by counting quantifiers. For every  $t \in \mathbb{N}$  we have the counting quantifier  $\exists^{\geq t}$ . For a formula  $\varphi(x)$  with the free variable  $x$  and for a graph  $G$  we have  $G \models \exists^{\geq t} x \varphi(x)$  if and only if there are at least  $t$  vertices  $v \in V(G)$  such that  $G \models \varphi[v]$ . The variables in a  $C^k$ -formula are all from a fixed  $k$ -element set, say  $\{x_1, \dots, x_k\}$ , but they can be reused. The  **$C^k$ -type** of a vertex  $v$  in  $G$  is the set of all  $C^k$ -formulas  $\varphi(x)$  such that  $G \models \varphi[v]$ . The  **$C^k$ -coloring** of  $G$  is the coloring of each vertex with its  $C^k$ -type. The  **$C^k$ -partition** of a graph  $G$  is the partition of its vertex set induced by their  $C^k$ -types. Similarly, the  **$C^k$ -type** of a tuple  $(v, w)$  is the set of  $C^k$ -formulas  $\varphi(x, y)$  such that  $G \models \varphi[v, w]$ . For a vertex  $v$  or a pair of vertices  $v, w$  to obtain the **atomic  $C^k$ -type** we consider only quantifier-free  $C^k$ -formulas.

Let  $G$  be a graph and  $\Pi$  be a partition of  $V(G)$ . We say that  $\Pi$  is **equitable** if for all  $P, Q \in \Pi$  and  $v, v' \in P$ , the vertices  $v$  and  $v'$  have the same number of neighbors in  $Q$ . A vertex coloring  $\chi$  is called equitable if the partition induced

by  $\chi$  is equitable. A partition  $\Pi$  is **coarser** than a partition  $\Pi'$  if every partition class of  $\Pi$  is contained in some class of  $\Pi'$ . It is a well-known fact that the  $C^2$ -partition of a graph is its **coarsest equitable partition**. The  $C^2$ -partition of a graph with  $n$  vertices and  $m$  edges can be calculated in time  $O((m+n)\log n)$  by the color refinement procedure (see [5]).

### 2.1 Relational Structures and Partially Oriented Graphs

An **edge-colored partially oriented graph** (an ec-POG) is an edge-colored directed graph  $(G, c)$  (with  $c$  an edge-coloring function) without loops such that for every  $(v, w) \in E(G)$ , it holds that if  $(w, v) \in E(G)$  then  $c((v, w)) = c((w, v))$ . Slightly abusing terminology, we say that an edge  $(v, w) \in E(G)$  is **undirected** if  $(w, v) \in E(G)$  and **directed** otherwise. We accordingly draw  $(v, w)$  and  $(w, v)$  as one undirected edge between  $v$  and  $w$  and denote it by  $\{v, w\}$ . An ec-POG  $(G, c)$  is **complete** if for all  $v, w \in V(G)$  with  $v \neq w$  we have  $(v, w) \in E(G)$  or  $(w, v) \in E(G)$ .

In the following, we consider finite relational structures over a fixed signature  $\sigma = (R_1, \dots, R_\ell)$  where  $R_i$  has arity  $r_i$ . The various definitions given for graphs are analogously defined for structures. Let  $\mathfrak{A}$  be a finite relational structure with universe  $A$ . For every  $i \in \{1, \dots, \ell\}$  we define a function  $c_i : A^2 \rightarrow \mathcal{P}(\{1, 2\}^{r_i})$  via  $c_i(v_1, v_2) := \{(j_1, \dots, j_{r_i}) \in \{1, 2\}^{r_i} \mid \mathfrak{A} \models R_i(v_{j_1}, \dots, v_{j_{r_i}})\}$  where, as usual,  $\mathcal{P}$  denotes the power set and  $\{1, 2\}^{r_i}$  denotes the set of all  $r_i$ -tuples over  $\{1, 2\}$ . For all  $v, w \in A$  with  $v \neq w$  we let  $c(v, w) := (c_1(v, w), \dots, c_\ell(v, w))$ . Since for each  $i$  the possible images of  $c_i$  come from a set of bounded size, by using the order of the relations  $R_i$  in  $\sigma$ , one can easily define a canonical linear ordering  $\leq$  on the image of  $c$  (for example, by using the lexicographic order). With the help of this ordering, we define  $\text{ec-POG}(\mathfrak{A}) := ((A, E_{\mathfrak{A}}), c_{\mathfrak{A}})$  as the complete ec-POG with vertex set  $A$ , edge set  $E_{\mathfrak{A}} := \{(v, w) \mid v, w \in A, v \neq w \text{ and } c(v, w) \leq c(w, v)\}$  and the edge coloring  $c_{\mathfrak{A}} := c|_{E_{\mathfrak{A}}}$ , the restriction of  $c$  to  $E_{\mathfrak{A}}$ . Note that  $c(v, w)$  uniquely determines the atomic  $C^2$ -types of  $v, w, (v, w)$  and  $(w, v)$ .

A partition  $\Pi$  of the vertex set of an ec-POG is **equitable** if for all  $P, Q \in \Pi$ , for all  $v, v' \in P$  and for every edge color  $c$ , the vertices  $v$  and  $v'$  have the same number of  $c$ -colored outgoing, incoming and undirected edges connecting them to  $Q$ . An ec-POG is **color-regular** if for each edge color  $c$  every vertex  $v$  has the same  $c$ -indegree, the same  $c$ -outdegree and the same  $c$ -degree for undirected edges, respectively. An edge-colored undirected biregular graph on bipartition  $(P, Q)$  is called **color-biregular** if for every edge color  $c$  the subgraph induced by the edges of color  $c$  is biregular on  $(P, Q)$ . If the graph is partially oriented, vertices in each bipartition class must additionally have the same number of outgoing and incoming edges in each color.

## 3 Inversion

In this section, we treat the so-called **inversion problem** that is closely related to the question which graphs are identified by  $C^2$ . A **complete invariant** of an

equivalence relation  $\equiv$  on a class  $\mathcal{C}$  of structures is a mapping  $\mathcal{I}$  from  $\mathcal{C}$  to some set  $S$ , such that  $\mathfrak{A} \equiv \mathfrak{B}$  if and only if  $\mathcal{I}(\mathfrak{A}) = \mathcal{I}(\mathfrak{B})$ . We say that  $\mathcal{I}$  **admits linear time inversion** if given  $s \in S$  one can construct in linear time a structure  $\mathfrak{A}$  with  $\mathcal{I}(\mathfrak{A}) = s$  or decide that no such structure exists. This algorithmic task describes the inversion problem. Of course, if a structure  $\mathfrak{A}$  is identified then a solution to the inversion problem must construct  $\mathfrak{A}$  when given  $\mathcal{I}(\mathfrak{A})$ .

For  $C^2$ , we show that a natural complete invariant, namely  $\mathcal{I}_C^2$ , admits linear time inversion. Otto [21] proved that this invariant admits polynomial time inversion, not only for simple graphs, but for finite structures in general.

Given a graph  $G$ , one can canonically define a linear ordering  $P_1 \leq \dots \leq P_t$  on the classes of its coarsest equitable partition (see [21]). This ordering allows us to define  $\mathcal{I}_C^2$ , mapping  $G$  to  $(\bar{s}, M)$ , where  $\bar{s}$  is the tuple  $(|P_1|, \dots, |P_t|)$  and  $M$  is a  $t \times t$  matrix, such that every vertex in  $P_i$  has exactly  $M_{ij}$  neighbors in  $P_j$ . It is easy to see that  $\mathcal{I}_C^2$  is a complete invariant of  $C^2$ .

### 3.1 Inversion for Graphs

**Definition 1.** A graph is **circulant** if it has an automorphism that consists of exactly one (permutation) cycle. A graph on vertex set  $P \cup Q$  is **doubly-circulant** with respect to  $P$  and  $Q$  if it has an automorphism with exactly two (permutation) cycles, one on  $P$  and the other on  $Q$ . A graph is **multi-circulant** with respect to a partition  $\{P_1, \dots, P_\ell\}$  of its vertices if it has an automorphism with exactly  $\ell$  (permutation) cycles, each on one of the  $P_i$ .

While circulant graphs are transitive and thus regular, not every regular graph is transitive and not every transitive graph is circulant. Similarly, every doubly-circulant graph is biregular but not every biregular graph is doubly-circulant.

It is well-known that circulant graphs can be constructed by numbering the vertices from 0 to  $n - 1$ , picking an arbitrary set  $S \subseteq \{1, \dots, \lfloor n/2 \rfloor\}$  of distances and inserting all edges between pairs of vertices whose distance of indices in the circular ordering is contained in  $S$ . A  $k$ -regular graph on  $n$  vertices exists exactly if  $k \cdot n$  is even and  $k \leq n - 1$ . In this case a  $k$ -regular circulant graph on  $n$  vertices can be constructed in linear time. We call this the **circulant construction**.

A  $(k, \ell)$ -biregular graph on a bipartition  $(P, Q)$  with  $|P| = m$  and  $|Q| = n$  exists exactly if  $k \cdot m = \ell \cdot n$  as well as  $k \leq n$  and  $\ell \leq m$  (see, for example [16, 21]). In fact, under these conditions there exists such a graph that is doubly-circulant.

**Lemma 1.** For all  $k, \ell, m, n \in \mathbb{N}$  with  $k \leq n, \ell \leq m$  and  $k \cdot m = \ell \cdot n$ , one can construct in  $O(k \cdot m)$  time a  $(k, \ell)$ -biregular graph on a bipartition  $(P, Q)$  with  $|P| = m$  and  $|Q| = n$ , which is doubly-circulant with respect to  $P$  and  $Q$ .

The coarsest equitable partition of a graph is not necessarily its orbit partition. However, one can use Lemma 1 to show that for each  $C^2$ -equivalence class, there is a representative whose coarsest equitable partition is the orbit partition.

**Theorem 1.** For every graph  $G$ , there is a  $C^2$ -equivalent graph  $H$  which is multi-circulant with respect to its coarsest equitable partition.

**Corollary 1.**  $I_C^2$  admits linear time inversion on the class of graphs.

Given an equivalence relation  $\equiv$  on a class  $\mathcal{C}$  of structures, the **canonization problem** for  $\equiv$  is the problem of finding a map  $c : \mathcal{C} \rightarrow \mathcal{C}$  such that for every  $\mathfrak{A} \in \mathcal{C}$  it holds that  $c(\mathfrak{A}) \equiv \mathfrak{A}$  and for all  $\mathfrak{A}, \mathfrak{B} \in \mathcal{C}$  with  $\mathfrak{A} \equiv \mathfrak{B}$  we have  $c(\mathfrak{A}) = c(\mathfrak{B})$ . The map  $c$  is called a **canonization** for  $\equiv$  and  $c(\mathfrak{A})$  is the **canon** of  $\mathfrak{A}$  (with respect to  $c$ ). Typically, the goal is to find such a canonization  $c$  that can be evaluated efficiently. As a consequence of the theorem we obtain two corollaries.

**Corollary 2.** Canonization for  $C^2$  of graphs can be done in  $O((n + m) \log n)$  time.

**Corollary 3.** If a graph  $G$  is identified by  $C^2$ , then its coarsest equitable partition is the orbit partition.

It is natural to ask whether Corollary 3 holds for finite relational structures in general. This is not the case since the unique 1-factorization of  $K_6$  is rigid (i.e., has no non-trivial automorphisms).

### 3.2 Inversion for Finite Relational Structures

Let  $\mathfrak{A} = (A, R_1, \dots, R_\ell)$  be a finite relational structure. We define  $\mathfrak{A}|_2$ , the restriction of  $\mathfrak{A}$  to arity 2, to be the relational structure  $(A, R'_1, \dots, R'_\ell)$  with  $R'_i := \{(v_1, \dots, v_{r_i}) \in R_i \mid \{v_1, \dots, v_{r_i}\} \text{ has at most 2 elements,}\}$  where  $r_i$  is the arity of  $R_i$ . Obviously, two relational structures  $\mathfrak{A}$  and  $\mathfrak{B}$  are  $C^2$ -equivalent if and only if  $\mathfrak{A}|_2$  and  $\mathfrak{B}|_2$  are  $C^2$ -equivalent. Furthermore,  $\mathfrak{A}|_2$  and  $\mathfrak{B}|_2$  are  $C^2$ -equivalent if and only if  $\text{ec-POG}(\mathfrak{A}|_2)$  and  $\text{ec-POG}(\mathfrak{B}|_2)$  are  $C^2$ -equivalent. Hence, the inversion problem for  $I_C^2$  on finite relational structures reduces to the inversion problem for  $I_C^2$  on ec-POGs.

The complete invariant  $I_C^2$  for the class of graphs has a well-known extension to finite structures, which translates to the context of ec-POGs as follows: we simply replace the matrix in the original definition of  $I_C^2$  with a matrix  $M$  such that for all  $i, j \in \{1, \dots, k\}$  the entry  $M_{ij}$  is a tuple encoding for each edge color  $d$  the number of  $d$ -colored outgoing edges from a vertex in  $P_i$  to  $P_j$ . Similarly to the case of graphs, in order to solve the inversion problem for ec-POGs it suffices to solve it for the color-regular case (which corresponds to the inversion within one  $C^2$ -partition class) and for the color-biregular case (which corresponds to the inversion between two  $C^2$ -partition classes).

*Color-regular case:* Let  $n$  be the number of vertices. If  $n$  is odd, the degrees in each color in the underlying undirected graph must be even. Consequently, the circulant construction from Subsect. 3.1 can be adapted to perform the inversion for directed and colored edges.

If  $n$  is even and more than one color degree is odd, we cannot apply the circulant construction, which significantly complicates our task. In fact, as a 1-factorization (i.e., a partition of the edge set into perfect matchings) of  $K_6$  shows it might not be possible to construct a transitive graph with the given

color degrees. Still, we can construct a canonical representative for the corresponding  $C^2$ -equivalence class as follows. We use the 1-factorization construction due to Walecki (see [19]). For even  $n$  the construction decomposes the complete graph  $K_n$  into  $n - 1$  disjoint perfect matchings. This decomposition can be made canonical and computed in linear time (see also [23, Example 7.1.2.]). To obtain a graph with specific color degrees we can define a color class to be the union of a suitable number of 1-factors. We call this the **matching construction**. The construction has the property that two matchings of the 1-factorization always yield a Hamiltonian cycle. If we require directed edges, in which case in-degrees must be equal to out-degrees, we pair a suitable number of matchings and orient the obtained Hamiltonian cycles.

*Color-biregular case:* Our doubly-circulant construction for graphs can be altered to handle colored directed edges.

**Corollary 4.**  $\mathcal{I}_C^2$  admits linear time inversion on finite relational structures.

**Corollary 5.** Canonization for  $C^2$  of a relational structure  $\mathfrak{A} = (A, R_1, \dots, R_\ell)$  over a fixed signature can be done in time  $O((n + m) \log n)$  where  $n = |A|$  and  $m = |R_1| + \dots + |R_\ell|$ .

## 4 Characterization of the Graphs Identified by $C^2$

Here we examine the graphs that are identified by the logic  $C^2$  and give a complete characterization of them.

**Definition 2.** Let  $T$  be a tree with a designated vertex  $v$ . For  $i \in \{1, \dots, 5\}$  let  $(T_i, v_i)$  be an isomorphic copy of  $(T, v)$ . Let  $F$  be the disjoint union of the five trees  $(T_i, v_i)$  and  $E$  be the edge set of a 5-cycle on vertex set  $\{v_1, \dots, v_5\}$ . Then we call the graph obtained from  $F$  by inserting the edges in  $E$ , a **bouquet**.

A **bouquet forest** is a disjoint union of vertex-colored trees and non-isomorphic vertex-colored bouquets.

For sets  $P, Q$  we define  $[P, Q] := \{\{v, w\} \mid v \neq w, v \in P, w \in Q\}$ .

**Definition 3.** Let  $G$  be a graph with  $C^2$ -coloring  $\chi$ . We define the **flip of  $G$**  as the vertex-colored graph  $(F, \chi)$  with  $V(F) = V(G)$  and  $E(F) = E(G) \Delta ([P_1, Q_1] \cup \dots \cup [P_t, Q_t])$ , where the  $(P_i, Q_i)$  are all pairs of (not necessarily distinct)  $C^2$ -partition classes of  $G$  which satisfy  $|[P_i, Q_i] \cap E(G)| > |[P_i, Q_i] \setminus E(G)|$ . We say that  $(F, \chi)$  is a **flipped graph**. If  $\chi$  is the  $C^2$ -coloring of  $F$  and  $(F, \chi)$  is a flipped graph, we also say that  $F$  is flipped.

Here, the symbol  $\Delta$  in the definition denotes, the symmetric difference. The notions of a flip and a bouquet forest allow us to obtain the following classification.

**Theorem 2.** A graph is identified by  $C^2$  if and only if its flip is a bouquet forest.

**Corollary 6.** Given a graph with  $n$  vertices and  $m$  edges, we can decide whether it is identified by  $C^2$  in time  $O((m + n) \log n)$ .



A second corollary of Theorem 2 is concerned with vertex colorings of graphs that are identified by  $C^2$ .

**Corollary 7.** *Let  $(G, \chi)$  be a vertex-colored graph which is identified by  $C^2$  and let  $\chi'$  be a vertex coloring of  $G$  which induces a finer partition on  $V(G)$  than  $\chi$  does. Then  $(G, \chi')$  is also identified by  $C^2$ .*

Our classification result actually provides us with some deeper structural insight that we describe next.

For a  $(k, \ell)$ -biregular graph on bipartition  $(P, Q)$  we introduce the three following notations. (1)  $P \square Q \iff k = \ell = 0$ , (2)  $P \doteq Q \iff k = \ell = 1$  and (3)  $P \ll Q \iff k \geq 2$  and  $\ell = 1$ .

For a graph  $G$  with  $C^2$ -partition  $\Pi$ , we define the **skeleton**  $S_G$  of  $G$  as the graph with  $V(S_G) = \Pi$  and  $E(S_G) = \{\{P, Q\} \mid P \doteq Q \text{ or } P \ll Q \text{ in the flip of } G\}$ .

Since they can appear only once per connected component of the skeleton, we call a  $C^2$ -partition class that is a 5-cycle or a matching an **exception**. Our classification of finite relational structures that are identified by  $C^2$ , which we give in the next section, depends on the structural properties that can be proven for identified graphs. These can be summarized as follows.

**Corollary 8.** *A flipped graph  $G$  is identified by  $C^2$  if and only if the following hold: (1) Each  $C^2$ -partition class induces a graph identified by  $C^2$  (i.e., the induced graph has no edges or it is a matching or a 5-cycle), (2) for all  $C^2$ -partition classes  $P$  and  $Q$  we have  $P \square Q$ ,  $P \doteq Q$ ,  $P \ll Q$  or  $Q \ll P$ , (3) the skeleton  $S_G$  is a forest, (4) there is no path  $P_0, P_1, \dots, P_t$  in  $S_G$  with  $P_0 \ll P_1$  and  $P_{t-1} \gg P_t$  in  $G$ , (5) there is no path  $P_0, P_1, \dots, P_t$  in  $S_G$  where  $P_0 \ll P_1$  and  $G[P_t]$  is a 5-cycle or a matching, and (6) in every connected component of  $S_G$  there is at most one exception.*

### 5 General Finite Structures

To generalize our results to finite structures, it suffices to analyze which edge-colored partially oriented graphs (i.e., ec-POGs) are identified.

**Theorem 3.** *Let  $G$  be a color-regular complete ec-POG. Then  $C^2$  identifies  $G$  if and only if  $G$  is (1) an undirected complete graph with only one edge color, (2) undirected and has two edge colors, one of which induces a perfect matching, or (3) one of the exceptions depicted in Fig. 1.*

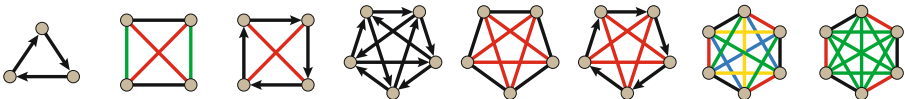


Fig. 1. The special cases that occur in the classification in Theorem 3

We will now describe how to combine such building blocks to form a larger identified ec-POG. By interpreting non-edges as edges of a special color we only need to consider complete graphs.

**Definition 4.** Let  $G$  be a vertex-colored ec-POG and let  $P$  and  $Q$  be two disjoint subsets of  $V(G)$ . We introduce the relation  $P \equiv_3^3 Q$  to denote the fact that the graph induced by the edges running between  $P$  and  $Q$  is the graph  $K_{3,3}$  with three edge colors which each induce a perfect matching between  $P$  and  $Q$ .

To define the relations  $P \square Q, P \doteq Q, P \ll Q$  for edge-colored graphs we always consider the graph induced by the edge color class that contains fewer edges and ignore orientations. It is not difficult to see that if the number of edges in the first color is equal to the number of edges in the second color then choosing either induced graph yields the same results. Note that with this convention the relations  $P \square Q, P \doteq Q, P \ll Q$  in particular imply that there are only at most two colors among the edges running between  $P$  and  $Q$ .

**Lemma 2.** Let  $G$  be a vertex/edge-colored undirected graph that is identified by  $C^2$ . If  $P$  and  $Q$  are distinct  $C^2$ -partition classes of  $G$ , then  $P \square Q, P \doteq Q, P \ll Q, Q \ll P$  or  $P \equiv_3^3 Q$ .

In an identified ec-POG, we call every  $C^2$ -partition class which does not induce an undirected complete graph with only one edge color an **exception** (i.e., any class that does not fall under Item 1 of Theorem 3). Similarly, we call every pair of  $C^2$ -partition classes  $P$  and  $Q$  for which  $P \equiv_3^3 Q$  holds an **exception**.

Let  $G$  be a vertex/edge-colored graph. As before, we define the vertices of the **skeleton**  $S_G$  to be the  $C^2$ -partition classes of  $G$ . Two distinct vertices  $P, Q$  in  $S_G$  are adjacent in  $S_G$  if the corresponding classes in  $G$  do not satisfy  $P \square Q$ . For identified structures, we obtain a theorem similar to Corollary 8 for graphs.

**Theorem 4.** Let  $G$  be a vertex-colored ec-POG. Then  $G$  is identified by  $C^2$  if and only if the following hold: (1) Each  $C^2$ -partition class induces a graph identified by  $C^2$ , (2) for all  $C^2$ -partition classes  $P$  and  $Q$  we have  $P \square Q, P \doteq Q, P \ll Q, Q \ll P$  or  $P \equiv_3^3 Q$ , (3) the skeleton  $S_G$  is a forest, (4) there is no path  $P_0, P_1, \dots, P_t$  in  $S_G$  with  $P_0 \ll P_1$  and  $P_{t-1} \gg P_t$ , (5) there is no path  $P_0, P_1, \dots, P_t$  in  $S_G$  where  $P_0 \ll P_1$  and  $P_t$  is an exception, and (6) in every connected component of  $S_G$  there is at most one exception.

**Corollary 9.** Given a finite relational structure  $\mathfrak{A}$  with a universe of size  $n$  over a fixed signature we can decide in time  $O(n^2 \log n)$  whether it is identified by  $C^2$ .

Using the classification we also obtain an extension of Corollary 7 to ec-POGs and, more generally, to finite relational structures.

**Corollary 10.** If a finite relational structure  $\mathfrak{A}$  is identified by  $C^2$ , then every finite relational structure obtained from  $\mathfrak{A}$  by adding unary relations is also identified by  $C^2$ .

## 6 Higher Dimensions

Considering triangular graphs [7, 14] we see that for  $k > 2$  any classification result for graphs identified by  $C^k$  must include an infinite number of non-trivial graphs. This already indicates that the situation for  $k = 2$  is special. We show that statements analogous to Corollaries 3 and 7 do not hold for higher dimensions. For this we use the construction from [6].

**Theorem 5.** *For every  $k > 2$ , there is a graph  $H$  of size  $O(k)$  identified by  $C^3$  for which the  $C^k$ -partition is strictly coarser than the orbit partition. Moreover, not all vertex-colored versions of  $H$  are identified by  $C^k$ .*

Even if a graph is identified and the orbits are correctly determined by  $C^k$ , it may still be the case that this does not hold for all colored versions of the graph.

**Theorem 6.** *For every  $k > 2$ , there is a graph  $H$  of size  $O(k)$  which is identified by  $C^3$  such that the  $C^k$ -partition classes are the orbits of  $H$  but there are vertex-colored versions of  $H$  that are not identified by  $C^k$  and for which the  $C^k$ -partition classes are not the orbits of  $H$ .*

## References

1. Arvind, V., Köbler, J., Rattan, G., Verbitsky, O.: On the power of color refinement. In: FCT 2015, (to appear 2015)
2. Atserias, A., Maneva, E.N.: Sherali-adams relaxations and indistinguishability in counting logics. *SIAM J. Comput.* **42**(1), 112–137 (2013)
3. Babai, L., Erdős, P., Selkow, S.M.: Random graph isomorphism. *SIAM J. Comput.* **9**(3), 628–635 (1980)
4. Babai, L., Kucera, L.: Canonical labelling of graphs in linear average time. In: FOCS 1979, pp. 39–46. IEEE Computer Society (1979)
5. 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)
6. Cai, J., Fürer, M., Immerman, N.: An optimal lower bound on the number of variables for graph identifications. *Combinatorica* **12**(4), 389–410 (1992)
7. Chang, L.-C.: The uniqueness and nonuniqueness of the triangular association scheme. *Sci. Record.* **3**, 604–613 (1959)
8. Dawar, A., Holm, B.: Pebble games with algebraic rules. In: Czumaj, A., Mehlhorn, K., Pitts, A., Wattenhofer, R. (eds.) ICALP 2012, Part II. LNCS, vol. 7392, pp. 251–262. Springer, Heidelberg (2012)
9. Grädel, E., Otto, M.: On logics with two variables. *Theor. Comput. Sci.* **224**(1–2), 73–113 (1999)
10. Grädel, E., Otto, M., Rosen, E.: Two-variable logic with counting is decidable. In: LICS 1997, pp. 306–317. IEEE Computer Society (1997)
11. Grohe, M.: Finite variable logics in descriptive complexity theory. *Bull. Symbolic Log.* **4**(4), 345–398 (1998)
12. Grohe, M., Otto, M.: Pebble games and linear equations. In: Cégielski, P., Durand, A. (eds.) CSL 2012, of LIPIcs, vol. 16, pp. 289–304 (2012)
13. Hella, L.: Logical hierarchies in PTIME. *Inf. Comput.* **129**(1), 1–19 (1996)

14. Hoffman, A.J.: On the uniqueness of the triangular association scheme. *Ann. Math. Statist.* **31**(2), 492–497 (1960)
15. Kiefer, S., Schweitzer, P., Selman, E.: Graphs identified by logics with counting. CoRR, abs/1503.08792 (2015). full version of the paper
16. Kopczynski, E., Tan, T.: Regular graphs and the spectra of two-variable logic with counting. CoRR, abs/1304.0829 (2013)
17. Krebs, A., Verbitsky, O.: Universal covers, color refinement, and two-variable logic with counting quantifiers: Lower bounds for the depth. CoRR, abs/1407.3175 (2014)
18. Kucera, L.: Canonical labeling of regular graphs in linear average time. In: FOCS 1987, pp. 271–279. IEEE Computer Society (1987)
19. Lucas, E.: *Récréations mathématiques*. 2ième éd., nouveau tirage. Librairie Scientifique et Technique Albert Blanchard, Paris (1960)
20. McKay, B.D., Piperno, A.: Practical graph isomorphism, II. *J. Symb. Comput.* **60**, 94–112 (2014)
21. Otto, M.: Canonization for two variables and puzzles on the square. *Ann. Pure Appl. Logic* **85**(3), 243–282 (1997)
22. Pratt-Hartmann, I.: Complexity of the two-variable fragment with counting quantifiers. *J. Log. Lang. Inf.* **14**(3), 369–395 (2005)
23. West, D.B.: *Introduction to Graph Theory*, 2nd edn. Pearson, Upper Saddle River (2000)