Recognizability Equals Definability for Partial k-Paths*

Valentine Kabanets

School of Computing Science, Simon Fraser University, Vancouver, Canada

Abstract. We prove that every recognizable family of partial k-paths is definable in a counting monadic second-order logic. We also show the obstruction set of the class of partial k-paths computable for every k.

1 Introduction

In 1960, Büchi [1] showed that a language is regular iff it is definable by some formula in a monadic second-order logic, MS. Here, MS is the extension of the first-order logic that allows quantification over set variables. A set of objects is definable by an MS-formula if the formula is true exactly on the members of the set. Thus Büchi established that recognizability is equivalent to MS-definability for words. Doner [7] then extended this result to ranked trees.

Graphs are algebraic objects since any graph can be constructed from smaller graphs using certain graph operations. They are also logical structures since any graph is completely determined by the set of its vertices and the adjacency relation on this set. Thus the notions of recognizability and definability can be extended to finite graphs. Courcelle [2] proved that every MS-definable set of finite graphs is recognizable, but not conversely. However, he was able to extend the result of Doner to unordered unbounded trees using a counting monadic second-order logic, CMS, an extension of MS that allows modular counting.

The question remained whether there was a sufficiently large class of graphs for which recognizability would imply CMS-definability. In their study of graph minors, Robertson and Seymour [10] introduced the notion of the tree-width of a graph. A graph of tree-width k exhibits certain tree-like structure. Such a graph can be decomposed into subgraphs of size $k + 1$ arranged as nodes of a tree (tree-decomposition) so that the nodes containing a given vertex form a subtree.

The class of graphs of tree-width at most k coincides with that of partial k-trees. Among other classes of graphs of bounded tree-width are trees and forests (tree-width ≤ 1), series-parallel graphs and outerplanar graphs (≤ 2), and Halin graphs (≤ 3).

The class of graphs of bounded tree-width plays an important role for another reason. Courcelle showed in [2] that the MS-theory of the class of partial k-trees

* This research was done while the author was at Simon Fraser University [8]. The author's present address is Department of Computer Science, University of Toronto, Toronto, ON, Canada M5S 3G4; kabanets@cs.utoronto.ca.

is decidable. Seese [11] proved that if the MS-theory of a class of finite graphs $\mathcal{G}$ is decidable, then the graphs in $\mathcal{G}$ have uniformly bounded tree-width. Thus, tree-width "characterizes" classes of finite graphs having decidable MS-theories.

Strictly speaking, the above results hold for so-called MS_2 logic, where MS_2 denotes the monadic second-order language using quantification over both vertex sets and edge sets of graphs; MS_1 is the language that uses quantification over vertex sets only (see [5, 6]). In this paper, we are using MS_2 and CMS_2.

For graphs of tree-width at most k, recognizability is defined using a tree automaton working on the corresponding tree-decompositions: A set $\mathcal{G}$ of partial k-trees G is recognizable if there is a tree automaton that accepts any tree-decomposition of each graph $G \in \mathcal{G}$, and rejects tree-decompositions of graphs not in $\mathcal{G}$. Courcelle [3] showed that a recognizable set of partial k-trees is CMS-definable for $k = 1$ and $k = 2$, and conjectured that recognizability implies CMS-definability of partial k-trees for every k. Kaller [9] proved the case of $k = 3$ and the case of k-connected partial k-trees.

We establish that every recognizable set of partial k-paths is CMS-definable, thereby proving a special case of Courcelle's conjecture. A partial k-path, or graph of bounded path-width, is a partial k-tree for which the corresponding tree-decomposition is a path-decomposition. Partial k-paths are recognized by finite automata working on the corresponding path-decompositions.

Our second result deals with computing the obstruction sets of minor-closed graph families. The class of partial k-trees (k-paths) is minor-closed and its obstruction set can be determined from the MS-formula defining that class [4]. We describe how to construct the MS-formula defining the class of partial k-paths for every given k. As a consequence, the obstruction sets of the classes of partial k-paths are computable for each k.

The remainder of this article is organized as follows: In Sect. 2, we give the necessary definitions. In Sect. 3, we show that recognizability implies CMS-definability for a generalization of the class of k-connected partial k-paths, the class of $(k, 1)$-paths. This is a base case of our solution for arbitrary partial k-paths which is outlined in Sect. 4.

2 Preliminaries

2.1 Partial *k*-Paths

We consider finite and simple graphs $G = (V, E)$, where V is the vertex-set and E is the edge-set of G. A *path-decomposition* (or *decomposition*) of G is a sequence $B = \langle B_1, \ldots, B_m \rangle$ of vertex-subsets, called *bags*, such that

1. every vertex $v \in V$ belongs to some bag B_i $(1 \leq i \leq m)$,
2. for each edge $e \in E$, there is a B_i $(1 \leq i \leq m)$ containing both ends of e,
3. for any $i, l, j \in \{1, \ldots, m\}$ such that $i \leq l \leq j$, $B_i \cap B_j \subseteq B_l$.

The *path-width of a decomposition* $B = \langle B_1, \ldots, B_m \rangle$ is $\max_{1 \leq i \leq m}\{|B_i|\} - 1$. A decomposition of path-width at most k will be called a *k-decomposition*. The *path-width of a graph* G is the minimum path-width over all decompositions of G. A *partial k-path* is a graph of path-width at most k.

Example 1. Graphs G_1 (Fig. 1) and G_2 (Fig. 2) are partial 1-path and 2-path, respectively, with possible decompositions: $B(G_1) = \langle\{1,2\},\{2,3\},\{3,4\},\{3,5\},\{3,6\}\rangle$ and $B(G_2) = \langle\{1,1',2\},\{1,2,3\},\{2,3,4\},\{2,3,5\},\{2,3,6\}\rangle$.

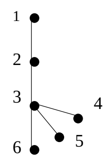

Fig. 1. A partial 1-path G_1.

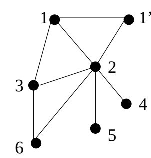

Fig. 2. A partial 2-path G_2.

For a partial k-path $G = (V, E)$ with a decomposition $B = \langle B_1, \ldots, B_m\rangle$, $\text{first}(v)$ is the number of the bag where a vertex $v \in V$ appears for the first time, i.e., $\text{first}(v) = \min_{1\le l\le m}\{l | v \in B_l\}$, $\text{new}(B_i)$ $(i \in \{1,\ldots,m\})$ is the set of vertices in B_i that appear in the decomposition for the first time, i.e., $\text{new}(B_i) = \{u \in B_i | \text{first}(u) = i\}$, and $\text{old}(B_i)$ is the set of vertices in B_i that also appear in some earlier bag, i.e., $\text{old}(B_i) = B_i \setminus \text{new}(B_i)$.

For G and B as above, a vertex $u \in B_r$ $(1 \le r \le m)$ is called a *drop vertex* of B_r iff for every $w \in V \setminus \cup_{i=1}^{r} B_i$, $\{u, w\} \notin E$. The set of all drop vertices of B_r $(1 \le r \le m)$ is denoted by $\text{drop}(B_r)$. The remaining vertices of B_r are called *non-drop vertices* of B_r, the set of which is denoted by $\text{non-drop}(B_r)$.

2.2 CMS-Definability

A graph $G = (V, E)$ can be viewed as a relational structure $(V \cup E, \{\mathbf{p}_v, \mathbf{p}_e, \mathbf{Inc}\})$, where $\mathbf{p}_v$ and $\mathbf{p}_e$ are unary predicates that define the vertex-set and the edge-set, respectively, and $\mathbf{Inc}$ is the ternary incidence predicate, i.e., for any $e \in E$ and $u, v \in V$, $\mathbf{Inc}(e, u, v) = \mathbf{True}$ iff $e = \{u, v\}$.

The language of *counting monadic second-order logic* corresponding to graphs G has the usual logical connectives: $\neg$ ("not"), $\wedge$ ("and"), $\vee$ ("or"), $\Rightarrow$ ("if-then"), and $\Leftrightarrow$ ("if and only if"), universal ($\forall$) and existential ($\exists$) quantifiers, equality symbol $=$, a sequence $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, ..., of individual variables, a sequence $\mathbf{U}, \mathbf{V}, \mathbf{W}, \ldots$, of set variables, the membership symbol $\in$, the unary predicate symbols $\mathbf{mod}_{p,q}$, $p < q$ are non-negative integers, and the predicate symbols $\mathbf{p}_v$, $\mathbf{p}_e$, and $\mathbf{Inc}$. In our interpretation, $\mathbf{mod}_{p,q}(\mathbf{V}) = \mathbf{True}$ iff $|S| = p \bmod q$, where S is the set denoted by the set variable $\mathbf{V}$.

A graph property P is called *CMS-definable* over a class of graphs $\mathcal{G}$ iff there is a CMS-formula Φ such that for each $G \in \mathcal{G}$, G satisfies P iff Φ is true on G.

Example 2. Connectedness of a graph G is an MS-definable property:
$\text{Connected} \equiv \forall\mathbf{V}_1\, \forall\mathbf{V}_2\ \ (\mathbf{V}_1 \ne \emptyset\ \wedge\ \mathbf{V}_2 \ne \emptyset\ \wedge\ \mathbf{V}_1 \cup \mathbf{V}_2 = V) \Rightarrow \text{Adj}(\mathbf{V}_1, \mathbf{V}_2)$,

$\mathrm{Adj}(\mathbf{V}_1, \mathbf{V}_2) \equiv \exists \mathbf{v}_1\ \exists \mathbf{v}_2\ \ \mathbf{v}_1 \in \mathbf{V}_1\ \wedge\ \mathbf{v}_2 \in \mathbf{V}_2\ \wedge\ \mathrm{adj}(\mathbf{v}_1, \mathbf{v}_2)$,
$\mathrm{adj}(\mathbf{v}_1, \mathbf{v}_2) \equiv \exists \mathbf{e}\ \ \mathbf{Inc}(\mathbf{e}, \mathbf{v}_1, \mathbf{v}_2)$,
where $(\mathbf{V}_i \neq \emptyset) \equiv \exists \mathbf{v}\ \ \mathbf{p}_v(\mathbf{v})\ \wedge\ \mathbf{v} \in \mathbf{V}_i\ \ (i = 1, 2)$ and
$(\mathbf{V}_1 \cup \mathbf{V}_2 = V) \equiv \forall \mathbf{v}\ \ \mathbf{p}_v(\mathbf{v}) \Rightarrow (\mathbf{v} \in \mathbf{V}_1\ \vee\ \mathbf{v} \in \mathbf{V}_2)$.

Using $\mathbf{mod}_{0,2}$, we can express in CMS the property that a given vertex subset of a graph has even cardinality. This cannot be done in MS alone [2].

2.3 Recognizability

We define the notion of recognizability of partial k-paths in terms of deterministic finite automata $A = (\Sigma, Q, \delta, q_0, F)$ working on extended decompositions. A decomposition $\bar{B} = \langle B_1, B_1^-, \ldots, B_m, B_m^- \rangle$ is called *extended* iff dropping old vertices and adding new vertices occur separately, i.e., $B_i^- = \text{non-drop}(B_i)$, $1 \leq i \leq m$.

Example 3. Here is an extended 1-decomposition of the graph G_1: $\bar{B}(G_1) = \langle \{1,2\}, \{2\}, \{2,3\}, \{3\}, \{3,4\}, \{3\}, \{3,5\}, \{3\}, \{3,6\}, \{\} \rangle$.

Let $G = (V, E)$ be a partial k-path with an extended k-decomposition $B = \langle B_1, \ldots, B_m \rangle$. Let $\beta : V \to \{1, \ldots, k+1\}$ be a labeling function such that any two distinct vertices in the same bag or in two consecutive bags have different labels. We call such labeling functions *admissible* by B. It is not difficult to see that $k+1$ labels always suffice in the case of *extended* decompositions. For the labeling function β and any set of vertices $W \subseteq V$, $\beta(W) = \cup_{w \in W} \beta(w)$.

For B and β described above, we define the following string $\sigma_\beta(B)$ of colored undirected graphs on at most $k+1$ vertices: $\sigma_\beta(B) = \langle \sigma_\beta(B_1), \ldots, \sigma_\beta(B_m) \rangle$, where for a bag B_i $(1 \leq i \leq m)$, $\sigma_\beta(B_i) = (V_\beta(B_i), E_\beta(B_i))$ such that $V_\beta(B_i) = \beta(B_i)$, and for every $u, u' \in B_i$, $\{\beta(u), \beta(u')\} \in E_\beta(B_i)$ iff $\{u, u'\} \in E$. Let Σ_g be the set of all colored (with colors $1, \ldots, k+1$) undirected graphs on at most $k+1$ vertices. Clearly, $|\Sigma_g|$ is bounded by a function of k.

A family $\mathcal{G}$ of partial k-paths G is called *recognizable* iff there is an automaton A with the input alphabet Σ_g such that for any G, $G \in \mathcal{G}$ iff $\sigma_\beta(B) \in L(A)$ for any extended k-decomposition B of G and any labeling function β admissible by B, and $G \notin \mathcal{G}$ iff $\sigma_\beta(B) \notin L(A)$ for any B and β as above. Here $L(A)$ denotes the language accepted by A.

3 The Case of $(k,1)$-Paths

3.1 $(k,1)$-Paths and k-Generative Orders

A connected partial k-path is called a $(k,1)$-path if it allows a k-decomposition $B = \langle B_1, \ldots, B_m \rangle$ satisfying the following conditions:

1. $\text{old}(B_i) = \text{non-drop}(B_{i-1})$ for every $i \in \{2, \ldots, m\}$,
2. $\text{drop}(B_i) \neq \emptyset$ for every $i \in \{1, \ldots, m\}$,

3. $|\text{new}(B_i)| = 1$ for every $i \in \{2, \ldots, m\}$.

Here (1) says that vertices are dropped from a bag as soon as possible, (2) that each bag contains at least one drop vertex, and (3) that exactly one new vertex is added to form the next bag. Note that every k-connected partial k-path is a $(k,1)$-path.

Example 4. The graphs G_1 and G_2 described earlier are $(k,1)$-paths.

To show that a recognizable family $\mathcal{G}$ of $(k,1)$-paths G is CMS-definable, it suffices to define in CMS some extended decomposition for every G and then use Büchi's result for sets of words. A decomposition of G can be defined if some linear order on V is known. Let $\leq$ be an arbitrary linear order on V, and let $\langle v_1, \ldots, v_n \rangle$ be the sequence of vertices in V ordered according to $\leq$. We define the sequence $B_{\leq} = \langle B_1, \ldots, B_n \rangle$, where $B_i = \{v_i\} \cup \{v_j | j < i \text{ and there is } j' \geq i \text{ s.t. } \{v_j, v_{j'}\} \in E\}$. Clearly, $B_{\leq}$ is a decomposition of G. For a partial k-path G, a linear order $\leq$ on V is called *k-generative* if $B_{\leq}$ is a k-decomposition. Conversely, from a $(k,1)$-decomposition B of G, one can define a k-generative linear order on G by setting u to be less than v iff $\text{first}(u) < \text{first}(v)$, $u, v \in V$, and ordering the vertices in B_1 arbitrarily.

Thus, to show that recognizability implies CMS-definability for $(k,1)$-paths, it would suffice to define in CMS a k-generative linear order for every given $(k,1)$-path. However, there are $(k,1)$-paths for which no linear order can be defined in CMS. Consider the family of $G_n = (\{0, 1, \ldots, n\}, E_n)$, where $E_n = \{\{0, j\} | 1 \leq j \leq n\}$. No linear orders can be CMS-defined on G_n, since these graphs have nontrivial automorphisms, and the size of G_n can be arbitrary large. So, in general, we cannot CMS-define a k-decomposition of a partial k-path.

For a partial k-path G, a partial order on V is called *k-generative* if every completion to a linear order on V is k-generative. We will describe a certain k-generative partial order, which is MS-definable over a suitably colored $(k,1)$-path G^c. Given such a partial order, one can MS-define a *tree-decomposition* of G of a special form. Since we cannot MS-define a path-decomposition but only a tree-decomposition, we need CMS to get the formula for recognizability of G^c, using an extension of Büchi's theorem. To convert the corresponding CMS-formula into a formula for the underlying uncolored $(k,1)$-paths G, we "guess" some coloring of G using a constant number of $\exists$ quantifiers, check in MS if it induces the required structure, and apply our CMS-formula to the colored graph.

To MS-define a k-generative partial order on a $(k,1)$-path G with a $(k,1)$-decomposition $B = \langle B_1, \ldots, B_m \rangle$, we convert G into the directed graph $G^d_B = (V, E^d)$ using the following algorithm. For a bag $B_r = \text{old}(B_r) \cup \text{new}(B_r)$ $(1 < r \leq m)$, where $\text{old}(B_r) = \{u_1, \ldots, u_s\}$ and $\text{new}(B_r) = \{v\}$, if $\{v, u_j\} \in E$, then $(v, u_j) \in E^d$. That is, we direct the edges from new to old vertices. To simplify the notation, we will often omit the superscript in E^d and the subscript in G^d_B.

Now we label G^d as follows. For $v \in \text{new}(B_r)$ and every $u \in \text{old}(B_r) \cap \text{drop}(B_r)$ $(1 < r \leq m)$, we color the arc $v \to u$ with some new color. This colored arc will be denoted as a double arrow $v \Rightarrow u$, and the set of them as $E_{\Rightarrow}$.

If $\{v\} = \text{new}(B_r) = \text{drop}(B_r)$, we color v with some new color, the same color for all such vertices; v will be denoted by having a loop arrow.

Example 5. For G_2 defined earlier, the $(k,1)$-decomposition $B(G_2)$ induces the labeled digraph G_2^d (Fig. 3).

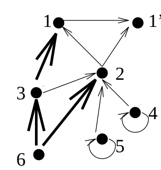

Fig. 3. The labeled digraph G_2^d, with double arrows shown as thick single arrows.

3.2 A k-Generative Partial Order

Given the digraph G^d induced by a $(k,1)$-decomposition B of a $(k,1)$-path G, we define the following binary relation of *strong precedence*, denoted by $\overset{s}{\prec}$, on the set V: for any $u, v \in V$, $u \overset{s}{\prec} v$ iff either $(v,u) \in E$ or there is some $w \in V$ such that $(u,w) \in E$ and $(v,w) \in E_{\Rightarrow}$. The reflexive and transitive closure of $\overset{s}{\prec}$, denoted by $\preceq$, is called *precedence*. Semantically, $u \prec v$ means that $\text{first}(u) < \text{first}(v)$. We extend $\preceq$ so that for any two vertices $u \in B_1$ and $v \notin B_1$ incomparable with respect to $\preceq$, u is less than v. Let $\preceq^1$ denote the transitive closure of that extension. Obviously, $\preceq^1$ is a k-generative partial order on G.

To define the required CMS-formula for recognizability of $(k,1)$-paths, we need a certain refinement of $\preceq^1$. We color G^d so that the precedence relation $\preceq$ is completed to a linear order on the set non-drop(B_1). We do so by coloring the non-drop vertices of B_1 with colors $1, \ldots, k$ so that no two vertices are colored the same. We denote this new colored digraph by G^{d1}.

Using G^{d1} enables us to define the following k sets $P_1, \ldots, P_k$. For any $v \in V$, $v \in P_i$ $(1 \le i \le k)$ iff i is the minimum over the labels of the vertices $u \in \text{non-drop}(B_1)$ such that there is a path of double arrows in the digraph G^{d1} from v to u. The set N of *nodes* is defined as $N = \cup_{i=1}^{k} P_i$, the set L of *leaves* is defined as $L = V \setminus (N \cup B_1)$.

Example 6. The digraph G_2^d from Example 5 can be viewed as G_2^{d1} with the two sets of nodes $P_1 = \{1,3,6\}$ and $P_2 = \{2\}$, and the set of leaves $L = \{4,5\}$.

Since no vertex in G^d can have more than one *incoming* double arrow, each set P_i, $1 \le i \le k$, induces a path of double arrows in G^{d1}. Therefore, each P_i is linearly ordered by $\preceq$. Using this fact, we can MS-define a k-generative partial order on G that is a linear order on the set of nodes N. We denote this partial order by $\preceq^n$. Note that we could MS-define a tree-decomposition of G using $\preceq^n$.

We need to order the leaves that are incomparable with respect to $\preceq^n$. By the definition of a $(k,1)$-decomposition, each leaf $w \in L$ has at most k outgoing single arrows pointing to some nodes from *different* sets $P_1, \ldots, P_k$. For a leaf $w \in L$, $P(w)$ denotes the set of nodes to which there are arrows from w, i.e., $P(w) = \{v \in N | (w,v) \in E\}$. We associate with each leaf $w \in L$ its *characteristic vector* $\chi(w) = (\chi_1(w), \ldots, \chi_k(w))$, where for each $1 \leq i \leq k$, $\chi_i(w) = 1$ if $P(w) \cap P_i \neq \emptyset$, and $\chi_i(w) = 0$ otherwise. We extend $\preceq^n$ to a new partial order on V, denoted by $\preceq^{nl}$, by ordering the leaves incomparable with respect to $\preceq^n$ lexicographically according to their characteristic vectors.

For two vertices $w_1, w_2 \in V$, we say that w_1 and w_2 are p-equivalent, denoted by $w_1 \overset{p}{\sim} w_2$, iff $w_1, w_2 \in L$ and $P(w_1) = P(w_2)$. For the quotient graph $G_p = G/\overset{p}{\sim} = (V_p, E_p)$ we extend $\preceq^{nl}$ to the set V_p in the standard way. Clearly, $\preceq^{nl}$ is a linear order on the set $(N \cup L)/\overset{p}{\sim}$. Ordering the drop vertices of B_1 arbitrarily yields a k-generative linear order on G_p, denoted by $\leq_p$. We will denote the digraph G^{d1} with ordered drop vertices of B_1 by $G^{d1'}$.

Example 7. For G_2, the $(k,1)$-decomposition of the corresponding quotient graph is $B'_p = \langle\{[1],[1'],[2]\},\{[1],[2],[3]\},\{[2],[3],[4]\},\{[2],[3],[6]\}\rangle$, where $[u]$ denotes the set of vertices p-equivalent to u, $u \in V$.

3.3 A CMS-Formula

Let $B'_p = \langle B'_1, \ldots, B'_m\rangle$ be the $(k,1)$-decomposition of the graph G_p induced by $\leq_p$. We can construct a $(k,1)$-decomposition of the original graph G as follows. In the sequence B'_p, replace B'_1 with B_1. For every $i \in \{1,\ldots,m\}$, replace $B'_i = \{[u_1]_{\overset{p}{\sim}}, \ldots, [u_{s_i}]_{\overset{p}{\sim}}, [w]_{\overset{p}{\sim}}\}$, where $[w]_{\overset{p}{\sim}}$ is the new vertex of B'_i such that $[w]_{\overset{p}{\sim}} = \{w_1, \ldots, w_{t_i}\}$ $(t_i \geq 1)$, with the sequence of bags $B(w_1) = \{u_1, \ldots, u_{s_i}, w_1\}, \ldots,$ $B(w_{t_i}) = \{u_1, \ldots, u_{s_i}, w_{t_i}\}$. Let B' denote thus constructed decomposition of G.

Example 8. For G_2, two decompositions B' are possible: $\langle\{1,1',2\},\{1,2,3\},$ $\{2,3,4\},\{2,3,5\},\{2,3,6\}\rangle$ or $\langle\{1,1',2\},\{1,2,3\},\{2,3,5\},\{2,3,4\},\{2,3,6\}\rangle$.

Let us convert B'_p into the extended decomposition $\bar{B'}_p$ and color G_p with some labeling function $\beta_p : V_p \rightarrow \{1,\ldots,k+1\}$ admissible by $\bar{B'}_p$. Let us also convert the decomposition B' of G into the extended decomposition $\bar{B'}$ and color the graph G with the labeling function $\beta : V \rightarrow \{1,\ldots,k+1\}$ such that, for every $v \in V$, $\beta(v) = \beta_p([v]_{\overset{p}{\sim}})$. The labeling function β is admissible by $\bar{B'}$ since no leaf appears in two consecutive bags. Note that the symbols in the alphabet Σ_g that correspond to the bags $\bar{B'}(w_1)$ and $\bar{B'}(w_2)$, for any two $\overset{p}{\sim}$-leaves w_1 and w_2, are identical. Let $\sigma_{\beta_p}(\bar{B'}_p) = \langle\sigma_1, \sigma_{1'}, \ldots, \sigma_m, \sigma_{m'}\rangle$. Then $\sigma_\beta(\bar{B'})$ can be obtained from $\sigma_{\beta_p}(\bar{B'}_p)$ by repeating every subsequence $\langle\sigma_i, \sigma_{i'}\rangle$ $(2 \leq i \leq m)$ $|[w]_{\overset{p}{\sim}}|$ times, where $\mathrm{new}(B'_i) = \{[w]_{\overset{p}{\sim}}\}$. It can be shown that $\sigma_{\beta_p}(\bar{B'}_p)$ is MS-definable.

Let $A = (\Sigma_g, Q, \delta, q_0, F)$ be the automaton recognizing a family $\mathcal{G}$ of $(k,1)$-paths G. To obtain the required CMS- formula for recognizability of $\mathcal{G}$, we use an extension of Büchi's result to words that are defined as sequences of substrings given with their multiplicities (in our case, the sequences $\sigma_{\beta_p}(\bar{B'}_p)$ with the cardinalities of the corresponding p-equivalence classes). By finiteness of A, to determine the behavior of A on a substring ω repeated t times, it suffices to know t mod a for some constant a dependent on A. Therefore, every recognizable family of colored $(k,1)$-paths $G^{d1'}$ is CMS-definable.

Let Φ be the CMS-formula checking the recognizability of suitably colored $(k,1)$-paths. We state without proof that there is an MS-formula Φ_{adm} verifying that a given coloring c of a $(k,1)$-path G is such that G is recognized by A iff Φ holds for G colored by c. Then the required CMS-formula for uncolored $(k,1)$-paths G is the following: $\exists$ "coloring c of G" $\Phi_{\text{adm}}(c) \wedge \Phi(G^c)$.

Theorem 1. *Every recognizable family of $(k,1)$-paths is CMS-definable.*

4 The General Case

4.1 Nice Decompositions

In general, a partial k-path is not necessarily a $(k,1)$-path; consider the partial 2-path G_2 from Example 1 with the new edge connecting vertices 4 and 5. We generalize our definition of $(k,1)$-decomposition as follows. A decomposition $B = \langle B_1, \ldots, B_m \rangle$ of G is called *nice* iff all of the following conditions hold:

1. $\text{old}(B_i) = \text{non-drop}(B_{i-1})$ for every $i \in \{2, \ldots, m\}$,
2. $\text{drop}(B_i) \neq \emptyset$ for every $i \in \{1, \ldots, m\}$,
3. for any $i \in \{2, \ldots, m\}$, if $|\text{new}(B_i)| > 1$, then
   (a) for any $v \in \cup_{j=i}^{m} \text{new}(B_j)$, each decomposition $\langle B_1, \ldots, B_{i-1}, \text{old}(B_i) \cup \{v\}, C_1, \ldots, C_s \rangle$ of G is such that $\text{drop}(\text{old}(B_i) \cup \{v\}) = \emptyset$, and
   (b) for any subset $S \subset \text{new}(B_i)$, each decomposition $\langle B_1, \ldots, B_{i-1}, \text{old}(B_i) \cup S, C_1, \ldots, C_s \rangle$ of G is such that $\text{drop}(\text{old}(B_i) \cup S) = \emptyset$.

Here (1) and (2) are as those for $(k,1)$-decompositions, and (3) says that if more than one new vertex is added to form B_i, then both (a) there was no single non-added vertex to choose instead of the set $\text{new}(B_i)$ so that B_i contained a drop vertex and (b) $\text{new}(B_i)$ is a minimal set with respect to set inclusion such that B_i contains a drop vertex.

It is not difficult to show that every k-decomposition can be converted into a nice k-decomposition. We call a nice k-decomposition $B = \langle B_1, \ldots, B_m \rangle$ a *(k,p)-decomposition* for some $1 \leq p \leq k$ iff $|\text{new}(B_i)| \leq p$ for all $1 < i \leq m$. A partial k-path allowing a (k,p)-decomposition will be called a *(k,p)-path*.

Let $B = \langle B_1, \ldots, B_m \rangle$ be a nice k-decomposition of a partial k-path G. The family of sets $\text{new}(B_i)$ $(1 \leq i \leq m)$ forms a partitioning of the vertex-set V of G. We call the corresponding equivalence on V the *1-equivalence*, denoted by $\stackrel{1}{\sim}$. The decomposition B also induces a linear order on the quotient set $V/\stackrel{1}{\sim}$, denoted by $\leq_1$. Clearly, given the pair $(\stackrel{1}{\sim}, \leq_1)$, we can reconstruct the decomposition B of G. Although we can MS-define the 1-equivalence when G is suitably colored, it is impossible to MS-define $\leq_1$.

We will divide a k-decomposition of a partial k-path G into a sequence of monotonic pieces whose structure resembles that of $(k,1)$-decompositions. Formally, a contiguous subsequence $\langle B_i,\ldots,B_{i+l}\rangle$ $(1 \le i, i+l \le m)$ of a decomposition $B = \langle B_1,\ldots,B_m\rangle$ is called *monotonic* iff $|\text{new}(B_i)| > 1$ and $|\text{new}(B_r)| = 1$ for each $i < r \le i+l$. The nice decomposition B can then be viewed as a sequence of monotonic pieces $\langle M_1,\ldots,M_d\rangle$, where $M_s = \langle B_{i_s},\ldots,B_{j_s}\rangle$ for each $1 \le s \le d$. Note that a nice decomposition is defined so that it is monotonic as long as possible, then there is a "jump" — more than one new vertex is added to a bag — which starts a new monotonic piece, and so on.

We define the sets $\text{new}(M_s) = \cup_{r=i_s}^{j_s}\text{new}(B_r)$ $(1 \le s \le d)$ the family of which forms a partitioning of the vertex-set V of G. The corresponding equivalence on V is called *2-equivalence* and denoted by $\overset{2}{\sim}$. This sequence of monotonic pieces also induces a linear order on the quotient set $V/\overset{2}{\sim}$, denoted by $\le_2$. Some k-decomposition of G (possibly different from B) can be constructed given $\overset{1}{\sim}$, $\overset{2}{\sim}$, and $\le_2$. Again, we can MS-define the 2-equivalence on a suitably colored graph, but not $\le_2$.

4.2 k-Generative Structures

For a partial k-path G, a triple $(\overset{1}{\sim}', \overset{2}{\sim}', \le_2')$, where $\overset{1}{\sim}'$ and $\overset{2}{\sim}'$ are equivalences on V and $\le_2'$ is a linear order on $V/\overset{2}{\sim}'$, is called a *linear k-generative structure on G* iff there exists some nice k-decomposition B of G such that $\overset{1}{\sim}'$ and $\overset{2}{\sim}'$ are the 1-equivalence and 2-equivalence, respectively, induced by B, and $\le_2'$ is the linear order on 2-equivalence classes induced by B. For a partial k-path G, a triple $(\overset{1}{\sim}', \overset{2}{\sim}', \preceq_2')$, where $\overset{1}{\sim}'$ and $\overset{2}{\sim}'$ are equivalences on V and $\preceq_2'$ is a partial order on $V/\overset{2}{\sim}'$, is called a *partial k-generative structure on G* iff any completion of $\preceq_2'$ to a linear order yields a linear k-generative structure on G.

Let $\overset{1}{\sim}$ and $\overset{2}{\sim}$ be the 1-equivalence and 2-equivalence, respectively, induced by some nice k-decomposition of a partial k-path G. Let $\preceq$ be the precedence relation defined similarly to the case of $(k,1)$-paths, and let $\overset{2}{\preceq}$ be the extension of $\preceq$ to the quotient set $V/\overset{2}{\sim}$ in the standard way. The triple $(\overset{1}{\sim}, \overset{2}{\sim}, \overset{2}{\preceq})$ is not necessarily a partial k-generative structure on G. One reason is that each $\overset{2}{\sim}$-class $[u]_{\overset{2}{\sim}}$ $(u \in V)$ contains several vertices all of which must be put in the same bag. The other reason is that $[u]_{\overset{2}{\sim}}$ can "contribute" more non-drop vertices than drop vertices. We did not have the latter problem in the case of $(k,1)$-paths, because there adding a new vertex always produced at least one drop vertex.

To get around these problems, we put consecutive monotonic pieces of the k-decomposition B of G into sequences of minimal length such that the number of non-drop vertices produced by each sequence, except the first one, is at most that of drop vertices. More formally, let $\mu = \langle M_s,\ldots,M_t\rangle$ be a contiguous subsequence of a nice k-decomposition B that corresponds to the sequence of bags $\langle B_{i_s},\ldots,B_{j_t}\rangle$. We define the *balance* of μ, $\text{bal}(\mu)$, as $\text{bal}(\mu) =$

$|\text{non-drop}(B_{j_t})| - |\text{old}(B_{i_s})|$. A contiguous subsequence μ of monotonic pieces is called *balanced* if $\text{bal}(\mu) \leq 0$ and no proper non-empty prefix of μ is of non-positive balance.

Let $B = \langle M_1, \ldots, M_d \rangle$, where M_s, $1 \leq s \leq d$, is a monotonic piece. We divide B into disjoint subsequences of monotonic pieces $\mu_1, \ldots, \mu_r$ such that $B = \mu_1 \ldots \mu_r$, $\mu_1 = \langle M_1 \rangle$, and each μ_i, $2 \leq i \leq r$, is balanced. It can be shown that every μ_i, $2 \leq i \leq r$, corresponds to a $(k, k-1)$-subdecomposition of G. The sets $\text{new}(\mu_i)$, $1 \leq i \leq r$, defined in an obvious way induce a partitioning of V. The corresponding equivalence is called 3_1-equivalence and is denoted by $\overset{3_1}{\sim}$. Recursively, we partition each μ_i, $1 \leq i \leq r$, into $\mu^i_1, \ldots, \mu^i_s$ and define 3_2-equivalence classes. Each μ^i_j, $2 \leq j \leq s$, corresponds to a $(k, k-2)$-subdecomposition of G. We stop after k steps when every (not necessarily balanced) sequence μ consists of a single monotonic piece and corresponds to a $(k, 1)$-subdecomposition of G; also note that 3_k-equivalence coincides with 2-equivalence.

Then we define partial orders on these 3_i-equivalence classes, denoted by $\overset{3_i}{\preceq}$, $1 \leq i \leq k$, satisfying the following condition: for any completions of $\overset{3_i}{\preceq}$ to linear orders $\leq^i$, $1 \leq i \leq k$, such that $\leq^j$ is a refinement of $\leq^i$ for every $j > i$ (i.e., the restriction of $\leq^j$ to $V/\overset{3_i}{\sim}$ coincides with $\leq^i$), the triple $(\overset{1}{\sim}, \overset{2}{\sim}, \leq^k)$ is a linear k-generative structure on G. These partial orders as well as 3_i-equivalences can be MS-defined for suitably colored connected partial k-paths thanks to the properties of nice decompositions.

4.3 Defining a CMS-Formula

We partition our set of 3_i-equivalence classes into the sets of 3_i-nodes and 3_i-leaves, $1 \leq i \leq k$. Then we refine each partial order $\overset{3_i}{\preceq}$, $1 \leq i \leq k$, to a linear order on the set of 3_i-nodes within each 3_{i-1}-equivalence class; every two vertices of G are 3_0-equivalent. However, we cannot order leaves in the same way as we did in the case of $(k, 1)$-paths, because now they are not necessarily single vertices but instead correspond to sequences of bags, and hence to words over Σ_g.

Let $A = (\Sigma_g, Q, \delta, q_0, F)$ be an automaton recognizing our family of partial k-paths. We call two incomparable 3_i-leaves within the same 3_{i-1}-equivalence class, $1 \leq i \leq k$, *δ_i-equivalent* if the corresponding words ω_1 and ω_2 over Σ_g are such that for each $q \in Q$, $\delta^*(q, \omega_1) = \delta^*(q, \omega_2)$, where δ^* is the extended transition function of A. To determine if two leaves are δ_i-equivalent, we need to know the behavior of A on the sequences of bags corresponding to those leaves.

The above discussion suggests the following "bottom-up" procedure which can be encoded in CMS. We define the sequence of bags corresponding to each 3_k-equivalence class as in the case of $(k, 1)$-paths, since each 3_k-equivalence class is the set of new vertices of a *monotonic* piece. Then we convert this sequence into the word ω over Σ_g and compute the behavior of A on ω. This behavior is a map from Q to Q, which can be presented as a *state-vector* $q(\omega)$ of length $|Q|$. For each 3_{k-1}-equivalence class C, two 3_k-leaves C' and C'' in $C/\overset{3_k}{\sim}$ are δ_k-equivalent iff $q(C') = q(C'')$. We extend the partial order on the set $C/\overset{3_k}{\sim}$ to a linear order on $C_\delta = (C/\overset{3_k}{\sim})/\overset{\delta_k}{\sim}$ by ordering incomparable leaves lexicographically according to their state-vectors. Let $\langle C_1, \dots, C_s \rangle$ be thus ordered sequence of elements of C_δ. The behavior of A on C is defined as $q(C) = q(C_1)^{t_1} \circ \cdots \circ q(C_s)^{t_s}$, where $t_i = |C_i|$, $1 \leq i \leq s$, and $\circ$ is the composition. By finiteness of Q, $q(C)$ can be defined in CMS. Continuing in this manner will give us, after k steps, the vector $q(G)$ describing the behavior of A on the entire k-decomposition of G. The graph G is recognized by A iff $q(G)$ maps q_0 to some final state of A.

Thus, we can define a CMS-formula for recognizability of suitably colored connected partial k-paths. As in the case of $(k, 1)$-paths, there is an MS-formula Φ'_{adm} so that recognizability implies CMS-definability for connected partial k-paths. Note that the formula $\exists$ "coloring c of G" $\Phi'_{\mathrm{adm}}(c)$ is true on G iff G is a partial k-path, so the obstruction set of the class of partial k-paths is computable.

For a disconnected partial k-path G, we compute the state-vectors for its connected components, order these vectors lexicographically, and compute their composition in CMS. Together with Courcelle's result this yields our main claim.

Theorem 2. *Recognizability equals definability for partial k-paths.*

Acknowledgements. I am indebted to my supervisor Arvind Gupta at SFU for suggesting this topic and for his encouragement and support. I want to thank David Mould for his assistance in preparing this paper. I am also grateful to the anonymous referees for their comments.

References

1. J. Büchi. Weak second-order arithmetic and finite automata. *Zeitschr. j. math. Logik und Grundlagen d. Math.*, 6:66–92, 1960.
2. B. Courcelle. The monadic second-order logic of graphs. I. Recognizable sets of finite graphs. *Information and Computation*, 85:12–75, 1990.
3. B. Courcelle. The monadic second-order logic of graphs. V. On closing the gap between definability and recognizability. *Theoret. Comput. Sci.*, 80:153–202, 1991.
4. B. Courcelle. The monadic second-order logic of graphs. III. Tree-decompositions, minors and complexity issues. *Informatique théorique et Appl.*, 26:257–286, 1992.
5. B. Courcelle. The monadic second-order logic of graphs. VI. On several representations of graphs by relational structures. *Discr. Appl. Math.*, 54:117–149, 1994.
6. B. Courcelle. The monadic second-order logic of graphs. VIII. Orientations. *Ann. Pure Appl. Logic*, 72:103–143, 1995.
7. J. Doner. Tree acceptors and some of their applications. *J. Computer and System Sciences*, 4:406–451, 1970.
8. V. Kabanets. Recognizability equals definability for partial k-paths. Master's thesis, Simon Fraser University, June 1996.
9. D. Kaller. Definability equals recognizability of partial 3-trees, 1996. Workshop on Graph-Theoretic Concepts in Computer Science (WG '96).
10. N. Robertson and P. Seymour. Graph minors. II. Algorithmic aspects of tree-width. *J. Algorithms*, 7:309–322, 1986.
11. D. Seese. The structure of the models of decidable monadic theories of graphs. *Ann. Pure Appl. Logic*, 53:169–195, 1991.