Trees in Coalgebra from Generalized Reachability

Thorsten Wißmann
Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

Bálint Kocsis
Radboud University, Nijmegen, The Netherlands

Jurriaan Rot
Radboud University, Nijmegen, The Netherlands

Ruben Turkenburg
Radboud University, Nijmegen, The Netherlands

Abstract

An automaton is called *reachable* if every state is reachable from the initial state. This notion has been generalized coalgebraically in two ways: first, via a universal property on pointed coalgebras, namely, that a reachable coalgebra has no proper subcoalgebra; and second, a coalgebra is reachable if it arises as the union of an iterative computation of successor states, starting from the initial state.

In the current paper, we present corresponding universal properties and iterative constructions for trees. The universal property captures when a coalgebra is a tree, namely, when it has no proper tree unravelling. The iterative construction unravels an arbitrary coalgebra to a tree. We show that this yields the expected notion of tree for a variety of standard examples.

We obtain our characterization of trees by first generalizing the previous theory of reachable coalgebras. Surprisingly, both the universal property and the iterative construction for trees arise as an instance of this generalized notion of reachability.

2012 ACM Subject Classification Mathematics of computing → Trees; Theory of computation → Categorical semantics

Keywords and phrases Trees, Coalgebra, Factorization Systems

Digital Object Identifier 10.4230/LIPIcs.CALCO.2025.15

Category (Co)algebraic pearl

Related Version *Full version with appendix*: `https://arxiv.org/abs/2503.15585`

Funding This research is supported by the NWO grants VI.Vidi.223.096 and OCENW.M20.053.

1 Introduction

Reachability and having a *tree structure* are two fundamental structural properties of state-based systems. For instance, a deterministic automaton is reachable if for every state, there exists a path from the initial state to it, and it is a tree if for every state, there is a *unique* path from the initial state (the latter makes sense in particular if the transition map is partial, otherwise it forces the automaton to be infinite). Analogous notions of reachability and tree structure can be defined for many variations of automata and graphs.

Reachability has been characterized at the level of pointed coalgebras – intuitively, coalgebras with an initial state – providing a general notion that instantiates to a wide variety of systems beyond deterministic automata. There are, in fact, two equivalent characterizations (cf. Figure 1). First, a pointed coalgebra is reachable if it has no proper pointed subcoalgebras [2]. Second, a pointed coalgebra is reachable if, intuitively, the entire state space is obtained when starting at the initial state and iteratively computing next states [27, 7]. The latter construction makes use of the notion of *least bounds* to compute one-step successor states (originally introduced under the name *base* by Blok [9]). While

Notion	Universal Property	Morphism Class	Iterative Construction
Reachable	No proper subcoalgebra [2]	Monic	Union of least bounds [7, 27]
Trees	No proper unravelling	Arbitrary	Coproduct of F-precise maps

Figure 1 Properties of coalgebras and their constructions.

the first approach characterizes reachability abstractly via a universal property, the second approach provides an iterative process that can be used to obtain the reachable part of a pointed coalgebra.

In the current paper, we extend this theory by providing characterizations of when the underlying graph of a pointed coalgebra is a tree. Analogously to the case of reachability, we provide two equivalent characterizations, one via a universal property and one via an iterative construction. These are summarized in the second row of Figure 1.

The first characterization intuitively says that a pointed coalgebra is a tree if it has no proper unravellings, where an unravelling can copy shared nodes and unfold loops. Formally, an unravelling of a coalgebra is captured as a pointed coalgebra homomorphism into it; the difference with the characterization of reachability is that this homomorphism is not required to be monic. The second characterization is again an iterative construction: a coalgebra is a tree if it arises as the *coproduct* of its iteratively computed successors. There is, however, a catch: the classic least bound does not account for avoiding sharing of one-step successors. Therefore, we use the notion of *precise maps* [24], a variation of least bounds whose relevance for tree unravelling has been conjectured [25, Ex. 4.16, Fn 2]. The resulting iterative construction can be used to compute the tree unravelling of a given pointed coalgebra.

We recover expected tree notions for instances such as (partial) automata, coalgebras for various simple polynomial functors, and coalgebras for the bag functor (multisets). Interestingly, coalgebras for the powerset functor are never a tree in our characterization, as this functor allows a form of copying that yields multiple edges between two nodes.

The key idea behind our technical approach is that all characterizations, of reachability and of being a tree, turn out to be instances of a generalized notion of reachability. While reachability is studied in [27] in the setting of an $(\mathcal{E}, \mathcal{M})$-factorization system where $\mathcal{M}$ consists of monos (not necessarily including all of them), here we relax this condition to allow $\mathcal{M}$ to be arbitrary. Then, taking $\mathcal{M}$ to be *all* morphisms is precisely what yields the characterization of trees (see also the morphism class column in Figure 1). In order to make the generalization work, we have to relax the condition that every $\mathcal{M}$-subcoalgebra is an isomorphism to the requirement that it is a split epimorphism. When $\mathcal{M}$ consists of monos, this is equivalent to the isomorphism condition, recovering the classical notion of reachability.

We then show that the iterative construction can also be formulated at this level, and that it is equivalent to the characterization via a universal property, generalizing the corresponding results from [27, 7]. We thus arrive at a general theory that instantiates both to reachability and to the property of being a tree, and provides two equivalent characterizations, one via a universal property and one via an iterative construction.

The outline of this paper is as follows. In Section 2, we recall preliminaries on coalgebras and factorization systems. In Section 3, we recall the notion of least bounds and precise morphisms, and relate the two notions. In Section 4, the generalized notion of reachability is introduced, in terms of a universal property and a matching iterative construction. We instantiate this generalized notion of reachability in Section 5 to obtain an account of trees in coalgebras. The paper concludes with a description of future work in Section 6.

2 Preliminaries

In the following, we assume basic knowledge of category theory (cf. standard textbooks such as [5, 6]). Usually, $\mathcal{C}$ and $\mathcal{D}$ denote categories and F denotes a functor.

▶ **Notation 2.1.** *Product projections are denoted by* $\mathsf{pr}_i\colon \prod_j A_j \to A_i$. *Coproduct injections are called* $\mathsf{inl}\colon A \to A + B$, $\mathsf{inr}\colon B \to A + B$, *and* $\mathsf{in}_i\colon A_i \to \coprod_j A_j$. *We use the notation* $\lceil n \rceil$ *for the set* $\{0, \dots, n-1\}$.

▶ **Definition 2.2.** *Given a category* $\mathcal{C}$ *and an endofunctor* $F\colon \mathcal{C} \to \mathcal{C}$, *an* F-coalgebra *is a pair* (C, c) *consisting of an object* C *(of* $\mathcal{C}$*) and a morphism* $c\colon C \to FC$ *(in* $\mathcal{C}$*). For a fixed object* $I \in \mathcal{C}$, *a* pointed F-coalgebra *is an* F*-coalgebra equipped with a morphism* $i_C\colon I \to C$, *called the* point. *An* F-coalgebra morphism $h\colon (C, c) \to (D, d)$ *between* F*-coalgebras* (C, c) *and* (D, d) *is a morphism* $h\colon C \to D$ *with* $d \cdot h = Fh \cdot c$ *(Figure 2b). A* pointed F-coalgebra morphism $h\colon (C, c, i_C) \to (D, d, i_D)$ *is an* F*-coalgebra morphism that preserves the point:* $i_D = h \cdot i_C$ *(Figure 2a). We denote the category of* F*-coalgebras and* F*-coalgebra morphisms by* $\mathsf{Coalg}(F)$, *and the category of pointed* F*-coalgebras and pointed* F*-coalgebra morphisms by* $\mathsf{Coalg}_I(F)$.

The functor F is dropped from the terminology when understood from context. Intuitively, the *carrier* C of a coalgebra (C, c) is the state space and the point $I \to C$ models an initial state. For $\mathcal{C} = \mathsf{Set}$, we fix $I = 1$ in this paper. Alternatively, one could choose e.g. $I = 2$ to model two initial states. The morphism $c\colon C \to FC$ sends states to their possible next states. The functor of choice F defines how these possible next states FC are structured. We provide a few examples below, and refer to [14, 18] for more examples and a general introduction to the theory of coalgebras.

▶ **Example 2.3.**

1. The powerset functor $\mathcal{P}\colon \mathsf{Set} \to \mathsf{Set}$ sends each set to its powerset. A $\mathcal{P}$-coalgebra $c\colon C \to \mathcal{P}(C)$ is a transition system, where C is the set of states, and there is a transition from x to y iff $y \in c(x)$. It can also be viewed as a directed graph (where self-loops are allowed).
2. Consider the functor $F\colon \mathsf{Set} \to \mathsf{Set}$ given by $F(X) = X \times X + \{\bot\}$. An F-coalgebra $c\colon C \to C \times C + \{\bot\}$ maps every state to either a pair of next states, or to $\bot$, representing termination.
3. Let A be a fixed set (the alphabet), let $2 = \{0, 1\}$, and let $F\colon \mathsf{Set} \to \mathsf{Set}$ be given by $F(X) = 2 \times (X + \{\bot\})^A$. Coalgebras for F are *partial deterministic automata*. They are conveniently presented as a pair of maps $\langle o, \delta \rangle\colon C \to 2 \times (C + \{\bot\})^A$. Here, C is the set of states, and $o\colon C \to 2$ is thought of as the output function; we say $x \in C$ is an accepting state iff $o(x) = 1$. For $x \in C$ and a letter $a \in A$, if $\delta(x)(a) = \mathsf{inl}(y)$ for some $y \in C$, then we say that there is an a-transition from x to y. If $\delta(x)(a) = \mathsf{inr}(\bot)$, then the a-transition of x is undefined. A point $q_0\colon 1 \to C$ models the initial state.
   We extend δ to words $\delta^*\colon C \to (C + \{\bot\})^{A^*}$ in the evident way.

$$\begin{tikzcd} & C \arrow[d, "h"] \\ I \arrow[ur, "i_C"] \arrow[r, "i_D"'] & D \end{tikzcd}$$

(a) Point preservation.

$$\begin{tikzcd} C \arrow[r, "c"] \arrow[d, "h"] & FC \arrow[d, "Fh"] \\ D \arrow[r, "d"] & FD \end{tikzcd}$$

(b) F-coalgebra morphism.

$$\begin{tikzcd} A \arrow[rr, "f"] \arrow[dr, two heads, "e"'] & & B \\ & Im(f) \arrow[ur, tail, "m"'] & \end{tikzcd}$$

(c) Factorization.

$$\begin{tikzcd} A \arrow[r, two heads, "e"] \arrow[d, "f"'] & B \arrow[d, "g"] \arrow[dl, dashed, "\exists! d"] \\ C \arrow[r, tail, "m"] & D \end{tikzcd}$$

(d) Diagonal fill-in.

Figure 2 Commutative diagrams for Definitions 2.2 and 2.6.

4. A *signature* is a set Σ together with a map $\mathsf{ar}\colon \Sigma \to \mathbb{N}$, sending each symbol $\sigma \in \Sigma$ to its arity $\mathsf{ar}(\sigma) \in \mathbb{N}$. Every signature Σ induces a polynomial Set-functor $F_\Sigma(X) = \coprod_{\sigma\in\Sigma} X^{\mathsf{ar}(\sigma)}$. Pointed F_Σ-coalgebras can be understood as automata for which a state q marked by a symbol $\sigma \in \Sigma$ offers $\mathsf{ar}(\sigma)$-many inputs. Partial deterministic automata for a finite input alphabet A arise as F_Σ-coalgebras for the signature $\Sigma = 2 \times \mathcal{P}(A)$ with $\mathsf{ar}(b, S) = |S|$.

▶ **Definition 2.4.** *A* directed multigraph $G = (V, E, s, t)$ *consists of sets V (of* vertices*) and E (of* edges*), and functions $s, t\colon E \to V$ sending each edge to its source and target vertex, respectively. A pointed (or rooted) graph additionally has a distinguished vertex $v_0 \in V$. We call a graph* locally finite *if each pair of vertices is related by only finitely many edges. A* path *between vertices u and v is a finite sequence of consecutive edges:*

$$\mathsf{Path}(u, v) = \{p \in E^n \mid n \in \mathbb{N}, s(p_1) = u, \forall 1 < k \le n.\, t(p_{k-1}) = s(p_k), t(p_n) = v\}$$

We often drop the adjectives and simply speak of graphs. Examples of graphs are visualized later, e.g. in Figure 11. All commutative diagrams are also examples of graphs, where each object represents a vertex and the morphisms represent edges, possibly allowing multiple morphisms between the same pair of objects.

We model multigraphs as coalgebras for the bag (or multiset) functor $\mathcal{B}$, as defined next, instead of the powerset functor. Our motivation is that the coalgebraic concepts yield the expected graph-theoretic notions in the instantiation for the bag functor, e.g. when characterizing trees in the present paper (see Theorem 5.15 later). Another such example is colour refinement on graphs, which is obtained from coalgebraic bisimilarity on $\mathcal{B}$-coalgebras [26, Ex. 7.18.3]. Colour refinement is an early phase in the approximation of graph isomorphism checking via the Weisfeiler-Leman algorithm [19].

▶ **Example 2.5.** The bag functor $\mathcal{B}\colon \mathsf{Set} \to \mathsf{Set}$ sends each set to the set of its finite multisets:

$$\mathcal{B}X = \{m\colon X \to \mathbb{N} \mid m(x) = 0 \text{ for all but finitely many } x \in X\}.$$

A $\mathcal{B}$-coalgebra $c\colon C \to \mathcal{B}C$ is similar to a transition system, with the difference that there is a number $n \in \mathbb{N}$ of transitions from a given state $x \in C$ to a state $y \in C$.

We can identify $\mathcal{B}$-coalgebras on V with locally-finite multigraphs by translating between coalgebra structures $c\colon V \to \mathcal{B}V$ and edge sets E as follows:

$$E = \coprod_{(u,v)\in V^2} \lceil c(u)(v) \rceil \qquad \begin{aligned} s(\mathsf{in}_{(u,v)}k) &= u \\ t(\mathsf{in}_{(u,v)}k) &= v \end{aligned} \qquad \Longleftrightarrow \qquad c(u)(v) = \big|\{e \in E \mid s(e) = u, t(e) = v\}\big|.$$

From left to right, local finiteness of the graph ensures that the definition of $c(u)(v)$ is a natural number.

▶ **Definition 2.6** [5, Def. 14.1]**.** *Given classes of morphisms $\mathcal{E}$ and $\mathcal{M}$ in $\mathcal{C}$, we say that $\mathcal{C}$ has an* $(\mathcal{E}, \mathcal{M})$-factorization system *provided that:*

1. *$\mathcal{E}$ and $\mathcal{M}$ are closed under composition and contain all isomorphisms. We write $\twoheadrightarrow$ for morphisms $e \in \mathcal{E}$, and $\rightarrowtail$ for morphisms $m \in \mathcal{M}$.*
2. *Every morphism $f\colon A \to B$ in $\mathcal{C}$ has a factorization $f = m \cdot e$ with $e \in \mathcal{E}$ and $m \in \mathcal{M}$ (Figure 2c). We write $Im(f)$ for the intermediate object.*
3. *For each commutative square $g \cdot e = m \cdot f$ with $m \in \mathcal{M}$ and $e \in \mathcal{E}$, there exists a unique diagonal fill-in d with $m \cdot d = g$ and $d \cdot e = f$ (Figure 2d).*

We denote the classes of all epimorphisms, monomorphisms, isomorphisms, and (arbitrary) morphisms in $\mathcal{C}$ by $\mathsf{Epi}, \mathsf{Mono}, \mathsf{Iso}$*, and* Mor*, respectively. A factorization system is called* proper *if $\mathcal{E} \subseteq \mathsf{Epi}$ and $\mathcal{M} \subseteq \mathsf{Mono}$.*

Note that in general, the classes $\mathcal{E}$ and $\mathcal{M}$ are not assumed to be classes of epis, resp. monos.

▶ **Example 2.7.**
1. Set has an $(\mathsf{Epi}, \mathsf{Mono})$-factorization system.
2. Every category $\mathcal{C}$ has an $(\mathsf{Iso}, \mathsf{Mor})$-factorization system, which is not proper in general.

▶ **Definition 2.8.** *For a class $\mathcal{M}$ of morphisms in $\mathcal{C}$ and a functor $F\colon \mathcal{C} \to \mathcal{C}$, let $\mathcal{M}$-carried be the class of those (pointed) F-coalgebra morphisms $h\colon (C, c) \to (D, d)$ with $h \in \mathcal{M}$.*

▶ **Proposition 2.9** [25]**.** *If $F\colon \mathcal{C} \to \mathcal{D}$ preserves $\mathcal{M}$ (that is, $Fm \in \mathcal{M}$ for every $m \in \mathcal{M}$), then every*[1] *$(\mathcal{E}, \mathcal{M})$-factorization system of $\mathcal{C}$ lifts to an $(\mathcal{E}\text{-carried}, \mathcal{M}\text{-carried})$-factorization system on $\mathsf{Coalg}(F)$ and on $\mathsf{Coalg}_I(F)$.*

▶ Remark 2.10. Factorization systems have cancellation properties similar to those satisfied by monomorphisms and epimorphisms. For instance, if a composite $g \cdot f$ is in $\mathcal{M}$ and g is in $\mathcal{M}$, then f is in $\mathcal{M}$, too [5, Prop. 14.9(1)] (in the case of properness, the assumption $g \in \mathcal{M}$ can be dropped).

▶ **Definition 2.11.** *We refer to the pullback of an $\mathcal{M}$-morphism along an $\mathcal{M}$-morphism as a (binary) $\mathcal{M}$-*intersection*.*

▶ **Example 2.12.** If $\mathcal{M}$ is a class of monomorphisms in Set, then an $\mathcal{M}$-intersection is an intersection of subobjects in the usual sense. If $\mathcal{M} = \mathsf{Mor}$, i.e. $\mathcal{M}$ is the class of all morphisms in $\mathcal{C}$, then an $\mathcal{M}$-intersection is a (general) pullback in $\mathcal{C}$.

▶ **Example 2.13.** Pointed F-coalgebras also have $\mathcal{M}$-intersections under mild conditions: if the base category $\mathcal{C}$ has $\mathcal{M}$-intersections and if F preserves them weakly (i.e. it maps pullback squares of $\mathcal{M}$-morphisms to weak pullback squares), then $\mathsf{Coalg}_I(F)$ has pullbacks of $\mathcal{M}$-carried-morphisms.
1. For $\mathcal{C} = \mathsf{Set}$ and $\mathcal{M} \subseteq \mathsf{Mono}$, every functor $\mathsf{Set} \to \mathsf{Set}$ preserves binary intersections up to redefinition of $F\varnothing$ [23, 4, 8] (see [27] for a survey). This redefinition leaves $\mathsf{Coalg}_I(F)$ untouched, because there is at most one pointed coalgebra on $\varnothing$ (and none if $I \neq \varnothing$).
2. For $\mathcal{C} = \mathsf{Set}$ and $\mathcal{M} = \mathsf{Mor}$, this preservation condition boils down to the weak preservation of pullbacks. This criterion is studied at length in the coalgebraic literature [13], as this is the standard sufficient condition for coalgebraic bisimilarity and coalgebraic behavioural equivalence to coincide.

3 Precise Morphisms vs. Least Bounds

Towards a description of trees by an iterative construction on coalgebras, we start by discussing the existing notion of *precise morphisms* [24, Def. 3.1] that intuitively describes the relation between two *levels* of a tree: every node further down is the child of precisely one node in the level above.

▶ **Definition 3.1** [24, Def. 3.1]**.** *For a functor $F\colon \mathcal{C} \to \mathcal{D}$ and a class $\mathcal{M}$ of morphisms in $\mathcal{C}$, a morphism $p\colon P \to FR$ in $\mathcal{D}$ is called F-*precise *(w.r.t $\mathcal{M}$) or simply* precise *if for all $g\colon P \to FC$ (in $\mathcal{D}$) and for all $m\colon R \to D$ and $n\colon C \to D$ in $\mathcal{M}$, the following implication holds:*

$$\begin{tikzcd} P \arrow[r, "p"] \arrow[d, "g"'] & FR \arrow[d, "Fm"] \\ FC \arrow[r, "Fn"'] & FD \end{tikzcd} \quad \overset{\exists d}{\Longrightarrow} \quad \begin{tikzcd} P \arrow[r, "p"] \arrow[d, "g"'] & FR \arrow[dl, "Fd"] \\ FC & \end{tikzcd} \quad \& \quad \begin{tikzcd} & R \arrow[dl, "d"'] \arrow[d, tail, "m"] \\ C \arrow[r, tail, "n"] & D \end{tikzcd}$$

[1] Note that this includes non-proper factorization systems; earlier proofs, e.g. by Kurz [16, 1.3.5-1.3.7], impose restrictions on $\mathcal{E}$ and $\mathcal{M}$.

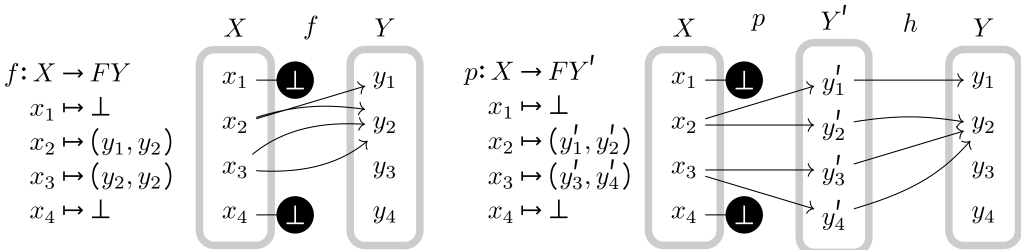

Figure 3 A non-precise map f that factors through the F-precise p for $FX = X \times X + \{\bot\}$ [24].

If $\mathcal{M}$ is part of an $(\mathcal{E}, \mathcal{M})$-factorization system, then by $m, n \in \mathcal{M}$, the diagonal d is necessarily in $\mathcal{M}$ as well (Remark 2.10). The present definition is a mild generalization of the original definition [24], which can be reobtained by setting $\mathcal{C} = \mathcal{D}$ and setting $\mathcal{M}$ to be the class of *all* morphisms.

▶ **Example 3.2** [24]**.** For intuition, we consider the simple case where $\mathcal{M}$ is the class of all morphisms and $\mathcal{C} = \mathcal{D} = \mathsf{Set}$. For a polynomial Set-functor F, a map $p\colon P \to FR$ is F-precise iff every element of R is mentioned precisely once in the definition of the map p. In particular, for the functor $FX = X \times X + \{\bot\}$ on Set, a map $p\colon P \to FR$ is precise iff

$$\forall y \in R,\ \exists!(x,i) \in P \times \{1,2\}\colon p(x) = (y_1, y_2) \text{ and } y_i = y.$$

The map $f\colon X \to FY$ in Figure 3 is not F-precise, because y_2 is used three times (once in $f(x_2)$ and twice in $f(x_3)$), and y_3 and y_4 do not occur in f at all. The map $p\colon X \to FY'$ is F-precise because every element of Y' is used precisely once in p. The equality $Fh \cdot p = f$ witnesses that f is indeed not precise. To see this, consider the commutative square $Fh \cdot p = F\mathsf{id}_Y \cdot f$. There is no diagonal $d\colon Y \to Y'$ with $h \cdot d = \mathsf{id}_Y$. Thus, f does not satisfy the definition of precise maps (Definition 3.1).

▶ Remark 3.3. If $\mathcal{C}$ has weak $\mathcal{M}$-intersections and F preserves them (i.e. weak pullbacks of $\mathcal{M}$-morphisms), then the morphism m in Definition 3.1 can be assumed to be the identity. That is, p is precise already if it fulfils the above definition for $m = \mathsf{id}$ [24, Rem. 3.2], i.e. if for all $g\colon P \to FC$ and for all $n\colon C \to R$ in $\mathcal{M}$ with $p = Fn \cdot g$, there is a $d\colon R \to C$ such that $Fd \cdot p = g$ and $n \cdot d = \mathsf{id}_R$. In particular, n has to be a split epi. If, moreover, $\mathcal{M} \subseteq \mathsf{Mono}$, then p is precise iff for any g and $n \in \mathcal{M}$, $p = Fn \cdot g$ implies that n is an isomorphism.

The lifting property of maps for the identity functor Id boils down to a special case of the diagonal lift of a factorization system:

▶ **Proposition 3.4.** *Suppose $\mathcal{C}$ has an $(\mathcal{E}, \mathcal{M})$-factorization system. Then a map $p\colon P \to R$ is $\mathsf{Id}_\mathcal{C}$-precise (w.r.t. $\mathcal{M}$) iff $p \in \mathcal{E}$.*

As illustrated in Figure 3, certain functors allow factorization of morphisms of the shape $X \to FY$ through precise maps.

▶ **Definition 3.5** [24, Def. 3.4]**.** *We say that $F\colon \mathcal{C} \to \mathcal{D}$* admits precise factorizations *if for every $f\colon P \to FY$, there exists an F-precise morphism $p\colon P \to FR$ and $h\colon R \to Y \in \mathcal{M}$ with $Fh \cdot p = f$ (cf. Figure 4). The triple (R, p, h) is called the* precise factorization *of f.*

The original definition [24] is for endofunctors and parametric in a class $\mathcal{S}$ of objects (requiring $P, R \in \mathcal{S}$), which we omit in the present paper for the sake of simplicity. In any case, even though the diagonal in the definition of precise morphism is not required to be unique, we can still deduce uniqueness of precise factorizations:

$$\begin{tikzcd} P \arrow[r, dashed, "\exists p"] \arrow[dr, "\forall f"'] & FR \arrow[d, "Fh"] \\ & FY \end{tikzcd}$$

Figure 4 Existence of precise factorizations.

$$\begin{tikzcd} X \arrow[r, "p_1"] \arrow[d, "p_2"'] & FY_1 \arrow[d, "Fg"] \arrow[dl, dashed, "Fd"', "\cong"] \\ FY_2 \arrow[r, "Fh"] & FZ \end{tikzcd}$$

Figure 5 Uniqueness of precise factorizations.

▶ **Lemma 3.6.** *If $\mathcal{M}$ is part of an $(\mathcal{E},\mathcal{M})$-factorization system, then precise factorizations are unique up to isomorphism. Concretely, if p_1 and p_2 are precise and $g, h \in \mathcal{M}$ in the commutative square in Figure 5, then the induced morphism d is an isomorphism.*

Many functors admit precise factorizations:

▶ **Proposition 3.7.** $\mathsf{Id}_{\mathcal{C}}$ *admits precise factorizations, if $\mathcal{C}$ has an $(\mathcal{E}, \mathcal{M})$-factorization system.*

▶ **Proposition 3.8** [24, Prop. 3.6]**.** *Let $\mathcal{M} = \mathsf{Mor}$ and $\mathcal{C} = \mathcal{D}$. The following functors admit precise factorizations:*
1. *Constant functors, if $\mathcal{C}$ has an initial object.*
2. *$F \cdot F'$, if F and F' do so.*
3. *$\prod_{i \in I} F_i$, if all $(F_i)_{i \in I}$ do so.*
4. *$\coprod_{i \in I} F_i$, if all $(F_i)_{i \in I}$ do so and $\mathcal{C}$ is I-extensive.*
5. *Right adjoint functors.*

▶ **Corollary 3.9.** *Let $\mathcal{C}$ be an extensive category. Then all polynomial endofunctors on $\mathcal{C}$ admit precise factorizations.*

Proof. Consequence of Proposition 3.7 and Proposition 3.8. ◀

Still for $\mathcal{M} = \mathsf{Mor}$, the powerset functor $\mathcal{P}\colon \mathsf{Set} \to \mathsf{Set}$ does *not* admit precise factorizations, because $\{e\} = \{e, e\} = \{e, e, e\} = \ldots$ for any e. Therefore, any e can be considered to be mentioned unboundedly often. This idea formalizes as follows:

▶ **Example 3.10.** A map $f\colon X \to \mathcal{P}Y$ is $\mathcal{P}$-precise (w.r.t Mor) iff $Y = \emptyset$.

▶ **Example 3.11.** The bag functor $\mathcal{B}\colon \mathsf{Set} \to \mathsf{Set}$ admits precise factorizations [24, Prop. 4.3]: a map $p\colon X \to \mathcal{B}Y$ is precise iff $\sum_{x \in X} p(x)(y) = 1$ for every $y \in Y$. For the factorization of a general $f\colon X \to \mathcal{B}Y$, we create for each $x \in X$ and $y \in Y$ as many copies of y as given by the multiplicity $f(x)(y) \in \mathbb{N}$. Let $Z = \coprod_{(x,y) \in X \times Y} \lceil f(x)(y) \rceil$. Then, define $m\colon Z \to Y$ by $m(\mathsf{in}_{(x,y)}(k)) = y$ and $p\colon X \to \mathcal{B}Z$ by $p(x)(\mathsf{in}_{(x',y)}(k)) = 1$ if $x = x'$ and 0 otherwise.

▶ Remark 3.12. The definition of precise morphisms (Definition 3.1) looks similar to that of $(\mathcal{E},\mathcal{M})$-functors [5, Chp. 17], but differs in the following aspects:
1. For precise morphisms, both m and n are required to be in $\mathcal{M}$, which is crucial for the unification with least bounds, as discussed next. In contrast, in the diagram for $(\mathcal{E},\mathcal{M})$-functors [5, Chp. 17], only the "opposite side" n is required to be in $\mathcal{M}$.
2. In contrast to $(\mathcal{E},\mathcal{M})$-functors, the diagonal in Definition 3.1 is not necessarily unique.
3. The factorization of $(\mathcal{E},\mathcal{M})$-functors is defined in terms of *sources*, that is, families of morphisms with a common domain.

If we instantiate $\mathcal{M}$ to be a class of monomorphisms, precise factorizations coincide with the notion of *least bound* (also called *base* [9, 7]) for a functor:

▶ **Definition 3.13** [9, 27]**.** *For a functor $F\colon \mathcal{C} \to \mathcal{D}$ and a class $\mathcal{M}$ of monomorphisms in $\mathcal{C}$, we say that F has* least bounds (w.r.t. $\mathcal{M}$) *if for every morphism $f\colon X \to FY$, there is a least morphism $m\colon Z \rightarrowtail Y$ in $\mathcal{M}$ such that f factors through Fm (via some $g\colon X \to FZ$). Concretely, this means that*

$$\begin{tikzcd} X \arrow[r, dashed, "\exists g"] \arrow[dr, "f"'] & FZ \arrow[d, "Fm"] \\ & FY \end{tikzcd}$$

commutes and for every $m'\colon Z' \rightarrowtail Y$ and $g'\colon X \to FZ'$ with $Fm' \cdot g' = f$, there is a morphism $h\colon Z \to Z'$ with $m' \cdot h = m$. The triple (Z, g, m) is called the least bound *of f.*

In Set, Z intuitively consists of those elements of Y that are actually used by f. Then, m is just the subset inclusion, and since m "omits" only those elements which are not used by f anyway, we can restrict the codomain of f to FZ, yielding $g\colon X \to FZ$.

▶ **Example 3.14.** For Set and $\mathcal{M}$ being the class of all monomorphisms (i.e. injective maps), consider the map $f\colon X \to FY$ from Figure 3 (for $FX = X \times X + \{\bot\}$). The least bound of the map is given by

$$Z = \{y_1, y_2\} \quad m\colon Z \hookrightarrow Y \quad g\colon X \to FZ \quad x_1 \mapsto \bot, \;\; x_2 \mapsto (y_1, y_2), \;\; x_3 \mapsto (y_2, y_2), \;\; x_4 \mapsto \bot$$

Here, m is the subset inclusion and g witnesses that the codomain of f restricts to FZ.

▶ **Example 3.15.** The bag functor $\mathcal{B}$ on sets (and $\mathcal{M}$ = Mono) has least bounds. The least bound of $f\colon X \to \mathcal{B}Y$ is given by those elements of Y that appear with non-zero multiplicity: $Z := \{y \in Y \mid \exists x \in X\colon f(x)(y) \neq 0\}$ and f then factors through $\mathcal{B}m$ for $m\colon Z \hookrightarrow Y$.

A key observation for the generalization of reachability to possibly non-proper factorization systems in Section 4 is that precise factorizations instantiate to least bounds in the following sense:

▶ **Proposition 3.16.** *For every functor $F\colon \mathcal{C} \to \mathcal{D}$, $\mathcal{M} \subseteq$ Mono, and $f\colon X \to FY$, we have:*

1. *If the precise factorization (w.r.t. $\mathcal{M}$) of f exists, then it is also the least bound of f.*
2. *If F preserves $\mathcal{M}$-intersections, if Fm is monic for every $m \in \mathcal{M}$, and if the least bound of f exists, then the least bound of f is also the precise factorization of f.*

In particular, under the given conditions, each of them exists iff the other one does.

▶ **Example 3.17.** Consider the least bound (Z, g, m) of f from Example 3.14. Since $FX = X \times X + \{\bot\}$ preserves intersections and monos, Proposition 3.16.2 implies that (Z, g, m) is also the precise factorization of f (w.r.t. $\mathcal{M}$ = Mono). Thus, the map g is precise w.r.t. $\mathcal{M}$ = Mono but not precise w.r.t. $\mathcal{M}$ = Mor.

4 Generalized Reachability

In the original definition [2, Def. 3.15], a coalgebra is reachable if it has no proper subcoalgebra. Here, a subcoalgebra refers to a (pointed) coalgebra morphism into the coalgebra of interest, carried by a monomorphism. Later [27], this condition was generalized from monomorphisms to a class $\mathcal{M}$ of monomorphisms, not necessarily including all of them. In the present paper, we generalize both the definition and the reachability construction to work for general $(\mathcal{E}, \mathcal{M})$-factorization systems, even those that are not proper. This is crucially used in Section 5, where we instantiate the theory to obtain characterizations of trees.

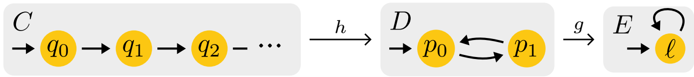

Figure 6 Surjective coalgebra morphisms without splittings.

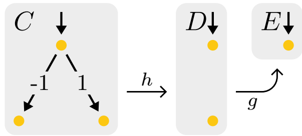

Figure 7 $\mathbb{R}^{(-)}$-coalgebras.

4.1 Universal Property

The universal property of reachability captures that it is not possible to "omit" states. The omission of states is formally encoded by a coalgebra morphism from a class $\mathcal{M}$:

▶ **Definition 4.1.** *Fix a functor $F\colon \mathcal{C} \to \mathcal{C}$ and a class $\mathcal{M}$ of morphisms in $\mathcal{C}$.*

1. *A* pointed $\mathcal{M}$-subcoalgebra *of a pointed coalgebra (C, c, i_C) is a pointed $\mathcal{M}$-carried-morphism $(T, t, i_T) \to (C, c, i_C)$. We simply speak of* subcoalgebras *if $\mathcal{M}$ is clear from the context (even if $\mathcal{M}$ is not assumed to be a class of monomorphisms).*
2. *The coalgebra (C, c, i_C) is called* $\mathcal{M}$-reachable *if every pointed $\mathcal{M}$-subcoalgebra of (C, c, i_C) is a split epimorphism of coalgebras (i.e. the splitting is also a pointed homomorphism).*

Note that the term *subcoalgebra* needs to be understood abstractly because it is parametric in the class $\mathcal{M}$ which may contain morphisms that are not monic. Still, we stick to the term *subcoalgebra* because it conveys the right intuition if $\mathcal{M} \subseteq \mathsf{Mono}$ is assumed. Under this assumption, we recover the usual definition of reachability:

▶ **Lemma 4.2.** *If $\mathcal{M} \subseteq \mathsf{Mono}$, then a coalgebra is reachable iff every pointed subcoalgebra is an isomorphism.*

▶ **Example 4.3.** For $\mathcal{B}$-coalgebras and $\mathcal{M} = \mathsf{Mono}$, a pointed directed multigraph is reachable as a coalgebra (V, c, v_0) iff for every vertex $u \in V$, there is some path from v_0 to u, that is, iff $\mathsf{Path}(v_0, u) \neq \emptyset$ for all $u \in V$.

▶ **Example 4.4.** For $\mathcal{M} = \mathsf{Mono}$, a partial deterministic automaton $\langle o, \delta\rangle\colon C \to 2\times(C+\{\bot\})^A$ considered as a coalgebra with a point $q_0\colon 1 \to C$ is reachable iff for each state $q \in C$, there is an input word $w \in A^*$ such that $\delta^*(q_0)(w) = \mathsf{inl}(q)$.

▶ **Example 4.5.** Let us briefly look at a small example to discuss split epimorphisms of coalgebras. Consider the $(\mathsf{Iso}, \mathsf{Mor})$-factorization system in Set and coalgebras for the identity functor, visualized in Figure 6. The map $g\colon (D, d, p_0) \to (E, e, \ell)$ is in Mor. However, there is no coalgebra morphism $(E, e) \to (D, d)$ (regardless of point preservation), so g is not a split epimorphism of pointed coalgebras. And indeed, E is not a tree (by intuitive means) because it has a cycle. The underlying map $g\colon D \to E$ has two splittings ($s_1\colon \ell \mapsto p_0$ and $s_2\colon \ell \mapsto p_1$), showcasing why we require split epis of coalgebras in Definition 4.1.2. Likewise, D is also not a tree (by intuitive means) because it has a cycle of length 2. And again, D is not Mor-reachable as witnessed by the coalgebra morphism h which has no splitting in coalgebras (though being surjective). This example naturally embeds into coalgebras for polynomial functors with at least one symbol of arity at least 1.

Later, in Example 5.2, we will discuss why Definition 4.1.2 only requires a split epimorphism instead of an isomorphism.

$$\begin{tikzcd} C_k \arrow[r, "c_k"] \arrow[d, tail, "m_k"'] & FC_{k+1} \arrow[d, tail, "Fm_k"] \\ C \arrow[r, "c"] & FC \end{tikzcd}$$

■ **Figure 8** Level k.

$$\begin{tikzcd} 1 \arrow[r, "i'_C"] \arrow[drr, "i_C"'] & C_0 \arrow[r, "\mathsf{in}_0"] & \coprod_{k\in\mathbb{N}} C_k \arrow[r, "\coprod_{k\in\mathbb{N}} c_k"] \arrow[d, "{[m_k]_{k\in\mathbb{N}}}"] & \coprod_{k\in\mathbb{N}} FC_{k+1} \arrow[r, "{[F\mathsf{in}_{k+1}]_{k\in\mathbb{N}}}"] & F\coprod_{k\in\mathbb{N}} C_k \arrow[d, "{F[m_k]_{k\in\mathbb{N}}}"'] \\ & & C \arrow[rr, "c"] & & FC \end{tikzcd}$$

■ **Figure 9** The morphism $[m_k]$ from the coproduct of all levels.

▶ **Lemma 4.6.** *Let $\mathcal{M}$ be part of an ($\mathcal{E}$,$\mathcal{M}$)-factorization system.*

1. *If $h\colon (C, c, i_C) \rightarrowtail (D, d, i_D)$ is an $\mathcal{M}$-morphism between $\mathcal{M}$-reachable coalgebras, then h is an isomorphism.*
2. *If $\mathcal{C}$ has $\mathcal{M}$-intersections and F preserves them weakly, then every coalgebra has at most one $\mathcal{M}$-reachable subcoalgebra. Thus, we may speak of* the *$\mathcal{M}$-reachable subcoalgebra (if it exists).*

The proof of Lemma 4.6 differs from the simple case when $\mathcal{M} \subseteq \mathsf{Mono}$, because the universal property of $\mathcal{M}$-reachability only provides split epimorphisms instead of isomorphisms.

▶ Remark 4.7. In Lemma 4.6.1, it is crucial that h is required to be an $\mathcal{M}$-morphism. Otherwise, the statement is wrong for $\mathcal{M} = \mathsf{Mono}$ and for functors that do not preserve preimages. A concrete example is the functor $FX = \mathbb{R}^{(X)}$ [13], for which successors in a coalgebra can disappear when being identified. Figure 7 shows examples of coalgebra morphisms for this functor. The composition $g \cdot h\colon C \to E$ is a morphism between reachable f-coalgebras, but $g \cdot h$ is not an isomorphism.

If we have factorizations of coalgebra morphisms, the following characterization of reachability in terms of $\mathcal{E}$ is immediate:

▶ **Lemma 4.8** [25, Lem. 4.6]**.** *If $\mathcal{M} \subseteq \mathsf{Mono}$ is part of an $(\mathcal{E}, \mathcal{M})$-factorization system and F preserves $\mathcal{M}$, then a coalgebra C is $\mathcal{M}$-reachable iff every homomorphism $h\colon D \to C$ is in $\mathcal{E}$.*

4.2 Construction

Having shown uniqueness, we now discuss sufficient conditions for the existence of the $\mathcal{M}$-reachable subcoalgebra of a given pointed coalgebra. We do so by generalizing the iterative reachability construction [7, 27] to arbitrary factorization systems:

▶ **Assumption 4.9.** *For the remainder of Section 4.2, we fix a functor $F\colon \mathcal{C} \to \mathcal{C}$ on a category $\mathcal{C}$ with an $(\mathcal{E}, \mathcal{M})$-factorization system such that F preserves $\mathcal{M}$-morphisms and F admits precise factorizations w.r.t. $\mathcal{M}$.*

▶ **Construction 4.10.** *For a pointed F-coalgebra (C, c, i_C), we define a sequence of morphisms $m_k\colon C_k \rightarrowtail C$ in $\mathcal{M}$ and precise morphisms $c_k\colon C_k \to FC_{k+1}$ $(k \in \mathbb{N})$ inductively:*

1. *Take the $(\mathcal{E}, \mathcal{M})$-factorization of $i_C\colon I \to C$ into* $\begin{tikzcd} I \arrow[r, two heads, "i'_C"] \arrow[rr, bend left, "i_C"] & C_0 \arrow[r, tail, "m_0"] & C \end{tikzcd}$.
2. *For $m_k\colon C_k \to C$, let c_k and $m_{k+1} \in \mathcal{M}$ be the precise factorization of $c \cdot m_k$ (Figure 8). We call C_k the k-th level.*

▶ **Lemma 4.11.** *In the setting of Construction 4.10, the coproduct of levels is a coalgebra making $[m_k]_{k\in\mathbb{N}}$ a pointed coalgebra morphism (Figure 9).*

Taking the factorization of this homomorphism yields a pointed coalgebra again:

▶ **Theorem 4.12.** *In the setting of Construction 4.10, the image R of the coalgebra morphism $[m_k]_{k\in\mathbb{N}}$ is an $\mathcal{M}$-reachable coalgebra:*

$$\begin{tikzcd} \coprod_{k\in\mathbb{N}} C_k \arrow[r, two heads, "{[e_k]_{k\in\mathbb{N}}}"] \arrow[rr, bend left, "{[m_k]_{k\in\mathbb{N}}}"] & R \arrow[r, tail, "m'"] & C \end{tikzcd}$$

▶ Remark 4.13. A restricted variant of this result was shown in [7, Theorem 17] with a fixed factorization system (where $\mathcal{E}$ is the class of strong epis and $\mathcal{M} = \mathsf{Mono}$, arising by assumptions on the category), and at a more general level in [27, Thm. 5.20.2], assuming only that $\mathcal{M} \subseteq \mathsf{Mono}$. Here, we make no assumptions about $\mathcal{M}$. Still, we follow the proof of [27, Thm. 5.20.2], but the current proof differs in that we do not (and cannot) use that the morphisms in $\mathcal{M}$ are monic, which was used in the final steps of the earlier proof. To address this, the constructed sequence of morphisms d_k in the proof satisfies a tighter property (which does not involve m').

Proof of Theorem 4.12. For easier comparison, we stick close to the notation of the proof of [27, Thm. 5.20.2] wherever possible. Let $h\colon (S, s, i_S) \rightarrowtail (R, r, i_R)$ be a pointed subcoalgebra, for which we now need to show that it is a split epimorphism. In the following, we will define morphisms $d_k\colon C_k \rightarrowtail S$ (in $\mathcal{M}$) satisfying

$$\begin{tikzcd} C_k \arrow[r, tail, "e_k"] \arrow[d, tail, "d_k"'] & R \\ S \arrow[ur, tail, "h"'] & \end{tikzcd} \quad\text{and}\quad \begin{tikzcd} C_k \arrow[r, "c_k"] \arrow[d, "d_k"'] & FC_{k+1} \arrow[d, "Fd_{k+1}"] \\ S \arrow[r, "s"] & FS \end{tikzcd} \quad\text{for all } k \in \mathbb{N}. \tag{1}$$

Note that $[e_k]_{k\in\mathbb{N}} \in \mathcal{E}$, but e_k (for each $k \in \mathbb{N}$) is in $\mathcal{M}$ because $m_k = m' \cdot e_k$ and $m_k, m' \in \mathcal{M}$. Thus, every such d_k will necessarily be in $\mathcal{M}$ as well.

Formally, we prove by induction that there exist morphisms d_k ($k \in \mathbb{N}$) such that for all $k \in \mathbb{N}$, we have $h \cdot d_k = e_k$ and, if $k \geq 1$, also $s \cdot d_{k-1} = Fd_k \cdot c_{k-1}$. We define $d_0\colon C_0 \rightarrowtail S$ using the diagonal fill-in property:

$$\begin{tikzcd} I \arrow[r, two heads, "i'_C"] \arrow[d, "i_S"'] & C_0 \arrow[d, tail, "e_0"] \arrow[dl, dashed, "d_0"'] \\ S \arrow[r, tail, "h"] & R \end{tikzcd}$$

In this diagram, the outside commutes since $i_R = e_0 \cdot i'_C$ and h preserves the point.

In the inductive step, given $d_k\colon C_k \rightarrowtail S$ with $h \cdot d_k = e_k$, the left-hand diagram below commutes. Using that c_k is precise and $h, e_{k+1} \in \mathcal{M}$, we obtain the diagonal fill-in d_{k+1} (on the right) satisfying the two desired equalities.

$$\begin{tikzcd} C_k \arrow[rr, "c_k"] \arrow[d, tail, "d_k"'] \arrow[dr, bend left, "e_k"] \arrow[dr, phantom, "\text{I.H.}"'] & & FC_{k+1} \arrow[dd, "Fe_{k+1}"] \arrow[ddl, phantom, "\text{Def. } R"] \\ S \arrow[r, "h"] \arrow[d, "s"'] \arrow[drr, phantom, "\text{Hom.}"] & R \arrow[dr, "r"] & \\ FS \arrow[rr, "Fh"] & & FR \end{tikzcd} \overset{\exists d_{k+1}}{\Longrightarrow} \begin{tikzcd} C_k \arrow[r, "c_k"] \arrow[d, "d_k"'] & FC_{k+1} \arrow[ddl, "Fd_{k+1}"] \\ S \arrow[d, "s"'] & \\ FS & \end{tikzcd} \;\&\; \begin{tikzcd} & C_{k+1} \arrow[d, "e_{k+1}"] \arrow[dl, dashed, "d_{k+1}"'] \\ S \arrow[r, "h"] & R \end{tikzcd}$$

Thus, (1) and the definition of d_0 yield that $[d_k]_{k\in\mathbb{N}}$ is a pointed coalgebra morphism, making the left-hand diagram commute:

$$
\begin{tikzcd}
I \arrow[r, "\mathsf{in}_0 \cdot i'_C"] \arrow[dr, "i_S"'] & \coprod_{k\in\mathbb{N}} C_k \arrow[rr, "\coprod_{k\in\mathbb{N}} c_k"'] \arrow[d, "{[d_k]_{k\in\mathbb{N}}}"] & & \coprod_{k\in\mathbb{N}} FC_k \arrow[r, "{[F\mathsf{in}_{k+1}]_{k\in\mathbb{N}}}"'] & F\coprod_{k\in\mathbb{N}} C_k \arrow[d, "{F[d_k]_{k\in\mathbb{N}}}"'] \\
& S \arrow[rrr, "s"] & & & FS
\end{tikzcd}
\Rightarrow
\begin{tikzcd}
\coprod_{k\in\mathbb{N}} C_k \arrow[r, two heads, "{[e_k]_{k\in\mathbb{N}}}"] \arrow[d, "{[d_k]_{k\in\mathbb{N}}}"'] & R \arrow[d, equal] \arrow[dl, dashed, "\exists! g"'] \\
S \arrow[r, tail, "h"] & R
\end{tikzcd}
$$

Thus, the right-hand square is a commutative square in the *category of pointed coalgebras*. Using the factorization system, we obtain a diagonal fill-in $g\colon R \to S$ that satisfies $h \cdot g = \mathsf{id}_R$, showing that h is indeed a split epimorphism of coalgebras. ◀

▶ **Corollary 4.14.** *A coalgebra is $\mathcal{M}$-reachable iff it is isomorphic to the reachable coalgebra R from Theorem 4.12.*

▶ **Example 4.15.** For every set functor $F\colon \mathsf{Set} \to \mathsf{Set}$, there is a canonical way to transform F-coalgebras into coalgebras for the powerset functor, which is interpreted as a directed graph and called the *canonical graph* [12]. Whenever F has least bounds for $\mathcal{M} \subseteq \mathsf{Mono}$, then an F-coalgebra is reachable iff its canonical graph is reachable (in the usual graph-theoretic sense) [27, Ex. 2.7, 5.26]. In such a case, the reachable subcoalgebra can simply be computed on the canonical graph.

5 Trees

We instantiate $\mathcal{M}$-reachability for the non-proper $(\mathsf{Iso}, \mathsf{Mor})$-factorization system that exists on every category $\mathcal{C}$. For simplicity, we spell out the definitions and constructions explicitly.

5.1 Universal Property

For $\mathcal{M} := \mathsf{Mor}$, reachability (Definition 4.1) instantiates to the following, which we call *tree*:

▶ **Definition 5.1.**
1. *An* unravelling *of a pointed coalgebra (C, c, i_C) is a pointed homomorphism $h\colon (T, t, i_T) \to (C, c, i_C)$.*
2. *The coalgebra (C, c, i_C) is called a* tree *if every unravelling h is a split epimorphism of pointed coalgebras.*
3. *A* tree unravelling *of a coalgebra is an unravelling that is a tree.*

Intuitively, the definition expresses that a tree (in the usual graph-theoretic sense) is a graph that cannot be unravelled any further, because it has no loops and no nodes share descendants. Unsurprisingly, if a coalgebra contains a cycle, then its tree unravelling is infinite (see e.g. Figures 6 and 14).

▶ **Example 5.2.** Figure 10 shows a simple unravelling of E that sends both q and r in D to the same node q in E. There is no coalgebra morphism $E \to D$, which shows that E is not a tree by the above definition. The coalgebra D is a tree, however. For instance, the coalgebra C contains two copies of D, so $C \cong D + D$ (as plain coalgebras), and the homomorphism $\nabla = [\mathsf{id}_D, \mathsf{id}_D]\colon D + D \to D$ has a right inverse, namely, inl (when assuming that the left-hand tree corresponds to the left-hand component; $\mathsf{inr}\colon D \to D + D$ does not preserve the point). Observe that ∇ is not an isomorphism, which showcases why Definition 4.1 of $\mathcal{M}$-reachability only requests $\mathcal{M}$-subcoalgebras to be split epimorphisms instead of isomorphisms.

▶ Remark 5.3. Note that E from Figure 10 shows that even if the canonical graph of a coalgebra is a tree (in the graph-theoretic sense), the coalgebra itself might not be a tree. The canonical graph of E is simply $\to p \longrightarrow q$, which is a tree.

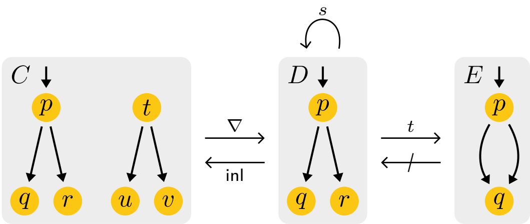

Figure 10 Unravelling for $FX = X \times X + \{\bot\}$.

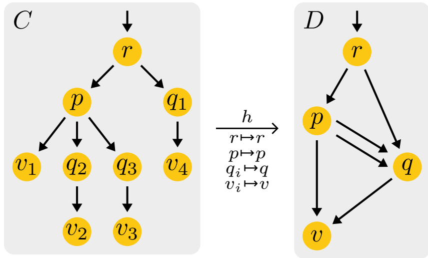

Figure 11 Unravelling of a $\mathcal{B}$-coalgebra.

In the example above, $\mathsf{inl}\colon D \to C$ is not surjective and thus not a split epimorphism of coalgebras, which shows that C is not a tree. This is no coincidence, because trees are always reachable:

▶ **Lemma 5.4.** *For any $(\mathcal{E}, \mathcal{M})$-factorization system, every tree is $\mathcal{M}$-reachable.*

Proof. Let (C, c, i_C) be a tree, and $h\colon (T, t, i_T) \rightarrowtail (C, c, i_C)$ a pointed $\mathcal{M}$-subcoalgebra. Then h is immediately a split epimorphism of pointed coalgebras as (C, c, i_C) is a tree. ◀

There is a slight variation of the definition of trees in terms of reachable coalgebras:

▶ **Definition 5.5.** *For a class of morphisms $\mathcal{M}$, a pointed coalgebra $C \in \mathsf{Coalg}_I(F)$ is called* ($\mathcal{M}$-)reachably unfolded *if every pointed coalgebra morphism $h\colon R \to C$ from an $\mathcal{M}$-reachable coalgebra $R \in \mathsf{Coalg}_I(F)$ is an isomorphism.*

Under some conditions, trees coincide with $\mathcal{M}$-reachably unfolded coalgebras:

▶ **Theorem 5.6.**

1. *If $\mathsf{SplitMono} \subseteq \mathcal{M}$, then every tree is $\mathcal{M}$-reachably unfolded.*
2. *If every pointed coalgebra admits a morphism from some $\mathcal{M}$-reachable coalgebra, then every $\mathcal{M}$-reachably unfolded coalgebra is a tree (and in particular, is also $\mathcal{M}$-reachable).*

Proof.

1. Let (C, c, i_C) be a tree, and $h\colon (R, r, i_R) \to (C, c, i_C)$ a homomorphism from an $\mathcal{M}$-reachable (R, r, i_R). As C is a tree, h is a split epimorphism so that it has a section $s\colon (C, c, i_C) \to (R, r, i_R)$. Since sections are split mono by definition and s is a pointed coalgebra homomorphism, s is a pointed $\mathcal{M}$-subcoalgebra of R. Thus, s is also split epi, and so an isomorphism. Then, also h must be an isomorphism. Hence, C is reachably unfolded.
2. Let (C, c, i_C) be reachably unfolded and $h\colon (T, t, i_T) \to (C, c, i_C)$ some unravelling. Let $h'\colon (R, r, i_R) \to (T, t, i_T)$ be a morphism from an $\mathcal{M}$-reachable coalgebra R. By assumption, $h \cdot h'$ must be an isomorphism, so that it has an inverse, say $s\colon C \to R$. Then, $h' \cdot s$ is a section of h as $h \cdot (h' \cdot s) = (h \cdot h') \cdot s = \mathsf{id}_C$. Further, h' and s are pointed coalgebra homomorphisms, so their composition is as well.
   For the additional comment on $\mathcal{M}$-reachability, note that since (C, c, i_C) is a tree, it is also $\mathcal{M}$-reachable by Lemma 5.4. ◀

▶ Remark 5.7. The notion of $\mathcal{M}$-reachable unfoldedness degrades if the assumptions of Theorem 5.6 are not met, and then, $\mathcal{M}$-reachable unfoldedness differs from being a tree:

1. Let $\mathcal{M} = \mathsf{Iso}$ (note that $\mathsf{SplitMono} \nsubseteq \mathcal{M}$). Then every coalgebra is $\mathcal{M}$-reachable and thus a pointed coalgebra D is reachably unfolded iff every coalgebra morphism $C \to D$ is an isomorphism. Thus, no coalgebra in Set is reachably unfolded for $\mathcal{M} = \mathsf{Iso}$, because one can always put $C := D + D$ (and $D + D \ncong D$), see for example $\nabla\colon C \to D$ in Figure 10.
2. Let $\mathcal{M} = \mathsf{Mor}$, i.e. the class of all morphisms in $\mathcal{C}$ (note that Definitions 4.1 and 5.5 impose no assumptions on $\mathcal{M}$). Then, $\mathcal{M}$-reachability is the property of being a tree. There are functors F for which there is a F-coalgebra (C, c, i_C) without any tree-unravelling (see Example 5.17 later and Figure 7 for concrete examples). Such a (C, c, i_C) is in particular not a tree, but is $\mathcal{M}$-reachably unfolded, for the trivial reason that no morphism from a tree (i.e. $\mathcal{M}$-reachable coalgebra) exists.

From the generalized uniqueness result (Lemma 4.6.1), we obtain that tree unravellings are unique up to isomorphism:

▶ **Lemma 5.8.**
1. *Every coalgebra morphism $h\colon C \to D$ between trees is an isomorphism.*
2. *If $\mathcal{C}$ has pullbacks and F preserves them weakly, then every coalgebra has at most one tree unravelling.*

In contrast to other categorical uniqueness results (e.g. that of (co)limits), tree unravellings are not unique up to *unique* isomorphism, because isomorphic but distinct subtrees can always be permuted by automorphisms. For example, consider the automorphism $s\colon D \to D$ in Figure 10 that swaps q and r; since $s \neq \mathsf{id}_D$, the tree unravelling D of E is not unique up to unique isomorphism.

Let us look at a few more examples of trees. We start with partial deterministic automata.

▶ **Definition 5.9.** *For a partial deterministic automaton $\langle o, \delta\rangle\colon C \to O \times (C + \{\bot\})^A$ with a point $q_0\colon 1 \to C$, define the* coalgebra of defined inputs *$\langle \bar{o}, \bar{\delta}\rangle$ by*

$$P = \{w \in A^* \mid \exists q \in C\colon \delta^*(q_0)(w) = \mathsf{inl}(q)\}, \quad \bar{o}\colon P \to O \quad \bar{o}(w) = o(\delta^*(q_0)(w)) \quad \varepsilon\colon 1 \to P$$

$$\bar{\delta}\colon P \to (P + \{\bot\})^A \qquad \bar{\delta}(w)(a) = \begin{cases} \mathsf{inl}(w\,a) & \text{if } \delta(\delta^*(q_0)(w)) = \mathsf{inl}(q) \text{ for some } q \in C \\ \mathsf{inr}(\bot) & \text{otherwise.} \end{cases}$$

In other words, P is the domain of the partial map $\delta^*(q_0)\colon A^* \rightharpoonup C$. Thus, $\delta^*(q_0)$ restricts to a total map $\delta^*(q_0)\colon P \to C$, which is a pointed coalgebra morphism by construction.

▶ **Lemma 5.10.** *For every partial DFA, the coalgebra of defined inputs is a tree.*

▶ **Corollary 5.11.** *The tree unravelling of a partial DFA is given by its defined inputs.*

We now take a look at graphs.

▶ **Definition 5.12.** *For a pointed multigraph $G = (V, E, s, t, v_0)$, define the* coalgebra of rooted paths *by*

$$\bar{E} := \bigcup_{v \in V} \mathsf{Path}(v_0, v) \qquad d\colon \bar{E} \to \mathcal{B}\bar{E} \qquad d(p)(p') = \begin{cases} 1 & \text{if there is } e \in E \text{ with } (p, e) = p' \\ 0 & \text{otherwise.} \end{cases}$$

The distinguished point $i_D := \varepsilon$ is the path of length 0 (from v_0 to v_0).

This coalgebra satisfies the property of being a tree (Definition 5.1):

▶ **Lemma 5.13.** *The coalgebra of rooted paths is a tree (in the coalgebraic sense).*

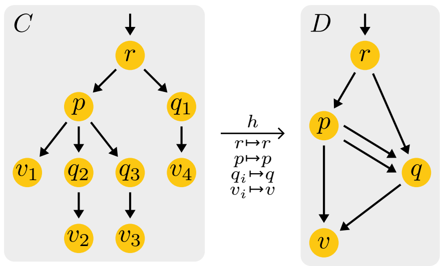

Figure 12 Unravelling a $\mathcal{P}$-coalgebra.

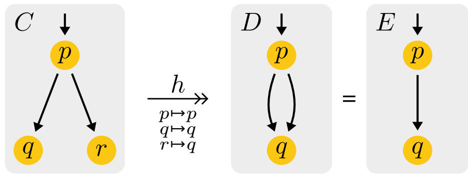

Figure 13 Unravelling a $\mathcal{B}$-coalgebra.

▶ **Proposition 5.14.** *For each pointed multigraph G, the map $t\colon \bar{E} \to V$ sending each path to its target vertex is a pointed coalgebra morphism from the coalgebra of rooted paths to G considered as a $\mathcal{B}$-coalgebra.*

▶ **Theorem 5.15.** *For every $\mathcal{B}$-coalgebra (V, c, i_V), the following are equivalent:*

1. *(V, c, i_V) is a tree (in the coalgebraic sense).*
2. *The morphism $t\colon \bar{E} \to V$ from the coalgebra of rooted paths is an isomorphism.*
3. *For every $v \in V$, there is precisely one path from v_0 to v.*

Proof.

- $1 \Rightarrow 2$: Since $\bar{E}$ and V are both trees, t is an isomorphism by Lemma 5.8.1.
- $2 \Rightarrow 1$: V is isomorphic to its coalgebra of rooted paths $\bar{E}$, which is a tree (Lemma 5.13). Thus, V is also a tree.
- $2 \Leftrightarrow 3$: Follows immediately from the definition of $t\colon \bar{E} \to V$ and the characterization of bijections as having unique preimages. ◀

▶ **Example 5.16.** In Figure 11, the coalgebra morphism $h\colon C \to D$ is the tree unravelling of D. Note that for every vertex $x \in D$, the number of paths from the root r to x determines the number of copies of x in the tree unravelling C. For instance, there are three paths to q: one direct edge from r and two paths via p.

One might first be tempted to model graphs as $\mathcal{P}$-coalgebras. However, $\mathcal{P}$-coalgebras do not provide the right notion of trees:

▶ **Example 5.17.** If a $\mathcal{P}$-coalgebra is a tree, then it has one vertex and no edges. In other words, no $\mathcal{P}$-coalgebra with an edge is a tree. The intuitive reason is that $\{x\} = \{x, x\}$, so it is always possible to duplicate siblings. For a concrete example, see Figure 12. Since $\{q\} = \{q, q\}$, the coalgebras D and E are identical (in a pure set-theoretic sense). The coalgebra D is not a tree, as witnessed by the unravelling to C. Now C is still not a tree, because we can apply the same principle again to duplicate the transition from p to q. As a consequence, no $\mathcal{P}$-coalgebra with an edge has a tree unravelling.

In summary, a $\mathcal{B}$-coalgebra is a tree iff it is a tree in the usual sense of a directed graph. In this case, the above issue of transition duplication can not happen, as visualized in Figure 13. In the coalgebra E, the transition from p to q has multiplicity 2, which we understand as a multigraph with two edges (visualized as D). Thus, q can be duplicated at most once, resulting in the coalgebra C. Every coalgebra morphism from a reachable coalgebra to C is then necessarily an isomorphism.

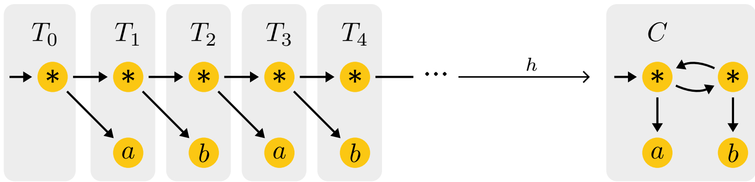

Figure 14 Tree unravelling h of an F_Σ-coalgebra, for $\Sigma = \{a/0, b/0, */2\}$.

5.2 Construction

The generalized reachability construction (Construction 4.10) instantiates nicely to trees for $\mathcal{E} = \mathsf{Iso}$ and $\mathcal{M} = \mathsf{Mor}$:

▶ **Construction 5.18.** *For a pointed coalgebra (C, c, i_C), we define the following sequence of maps $h_k\colon T_k \to C$ and precise maps $t_k\colon T_k \to FT_{k+1}$ $(k \in \mathbb{N})$:*
1. *$T_0 = I$, $h_0 = i_C\colon T_0 \to C$.*
2. *t_k and h_{k+1} are the precise factorization of $c \cdot h_k$:*

$$\begin{tikzcd} T_k \arrow[r, dashed, "t_k"] \arrow[d, "h_k"'] & FT_{k+1} \arrow[d, "Fh_{k+1}"] \\ C \arrow[r, "c"] & FC \end{tikzcd}$$

We call T_k the k-th level and $[h_k]_{k\in\mathbb{N}}\colon \coprod_k T_k \longrightarrow C$ the coproduct of levels *of (C, c, i_C).*

▶ **Theorem 5.19.** *For every coalgebra, the coproduct of its levels is a tree unravelling.*

▶ **Corollary 5.20.** *A coalgebra is a tree iff it is isomorphic to the coproduct of its levels.*

▶ **Example 5.21.** Adámek and Porst [3] consider the elements of the *final coalgebra* as *trees*. However, no coalgebraic definition of being a tree for a general functor is provided. Instead, they define a *tree coalgebra* A_t for a polynomial functor $F_\Sigma\colon \mathsf{Set} \to \mathsf{Set}$ and an element t of the final F_Σ-coalgebra (T_Σ, τ) [3, II.3]. Such an element $t \in T_\Sigma$ is a partial function $t\colon \mathbb{N}^* \rightharpoonup \Sigma$ that, intuitively speaking, returns the symbol $\sigma \in \Sigma$ whose location in the tree is described by $p \in \mathbb{N}^*$. For $t \in T_\Sigma$, the tree coalgebra A_t then consists of the domain of the partial function t, which is intuitively the set of rooted paths in the tree, analogous to the coalgebras of rooted paths and defined inputs considered above. As they observe already, the coalgebra A_t is not isomorphic to the sub-coalgebra of T_Σ generated by t. With our coalgebraic definition of tree, we now see that A_t is not only a tree in our sense, but also *the* tree unravelling of the pointed F_Σ-coalgebra (T_Σ, τ, t).

▶ **Example 5.22.** An example of an F_Σ-coalgebra is visualized in Figure 14. Its tree unravelling with levels $(T_k)_{k\in\mathbb{N}}$ arising from Construction 5.18 is shown on the left. The resulting coalgebra is the same as the tree coalgebra A_t [3, II.3]. Note that there is a unique pointed coalgebra morphism $h\colon T \to C$.

6 Conclusions and Future Work

Having characterized what it means for a coalgebra to be a tree, it remains open how other graph-based notions can be described by universal properties. Here, we think of notions like directed acyclic graphs (DAGs), shortest paths, and maximum flow.

What also remains for future work is to define a coalgebraic recursion principle on trees that works for possibly infinite trees. Established coalgebraic recursion principles known from well-founded coalgebras and recursive coalgebras restrict to finite-depth structures due to their "bottom up" strategy [17, 20, 22, 21, 15, 1].

It also remains for future work to connect our coalgebraic tree notion to trees considered in the open-map approach to bisimilarity, where a tree is *a colimit of paths* [10], covering non-deterministic systems and weighted systems [11].

Natural transformations $\alpha\colon F \to G$ canonically extend to functors $\mathsf{Coalg}_I(F) \to \mathsf{Coalg}_I(G)$. It is an open question which sufficient (or even necessary) conditions on the natural transformation α ensure that it preserves the property of being a tree. This may also give insights into how to relate the coalgebraic notion of trees in Set to properties of the canonical graph.

References

1. Jirí Adámek, Stefan Milius, and Lawrence S. Moss. On well-founded and recursive coalgebras. In Jean Goubault-Larrecq and Barbara König, editors, *Proc. 23rd International Conference on Foundations of Software Science and Computation Structures (FOSSACS 2020)*, volume 12077 of *Lecture Notes in Computer Science*, pages 17–36. Springer, 2020. `doi:10.1007/978-3-030-45231-5_2`.
2. Jirí Adámek, Stefan Milius, Lawrence S. Moss, and Lurdes Sousa. Well-pointed coalgebras. *Log. Methods Comput. Sci.*, 9(3), 2013. `doi:10.2168/LMCS-9(3:2)2013`.
3. Jirí Adámek and Hans-E. Porst. On tree coalgebras and coalgebra presentations. *Theor. Comput. Sci.*, 311(1-3):257–283, 2004. `doi:10.1016/S0304-3975(03)00378-5`.
4. Jiří Adámek and Věra Trnková. *Automata and Algebras in Categories*, volume 37 of *Mathematics and its Applications*. Kluwer Academic Publishers, 1990.
5. Jiří Adámek, Horst Herrlich, and George E. Strecker. Abstract and concrete categories. the joy of cats, 2004.
6. Steve Awodey. *Category Theory*. Oxford Logic Guides. OUP Oxford, 2010. URL: `http://books.google.de/books?id=-MCJ6x2lC7oC`.
7. Simone Barlocco, Clemens Kupke, and Jurriaan Rot. Coalgebra learning via duality. In *Foundations of Software Science and Computation Structures (FoSSaCS)*, pages 62–79. Springer, 2019. `doi:10.1007/978-3-030-17127-8_4`.
8. Michael Barr. Terminal coalgebras in well-founded set theory. *Theoretical Computer Science*, 114(2):299–315, 1993. `doi:10.1016/0304-3975(93)90076-6`.
9. Alwin Blok. Interaction, observation and denotation. Master's thesis, Universiteit van Amsterdam, 2012.
10. Jérémy Dubut, Eric Goubault, and Jean Goubault-Larrecq. Bisimulations and unfolding in p-accessible categorical models. In Josée Desharnais and Radha Jagadeesan, editors, *27th International Conference on Concurrency Theory (CONCUR 2016)*, volume 59 of *LIPIcs*, pages 25:1–25:14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2016. `doi:10.4230/LIPICS.CONCUR.2016.25`.
11. Jérémy Dubut and Thorsten Wißmann. Weighted and branching bisimilarities from generalized open maps. In Orna Kupferman and Pawel Sobocinski, editors, *Proc. Foundations of Software Science and Computation Structures (FoSSaCS 2023)*, volume 13992 of *Lecture Notes in Computer Science*, pages 308–327. Springer, 2023. `doi:10.1007/978-3-031-30829-1_15`.
12. H. Peter Gumm. From t-coalgebras to filter structures and transition systems. In José Luiz Fiadeiro, Neil Harman, Markus Roggenbach, and Jan Rutten, editors, *Algebra and Coalgebra in Computer Science*, pages 194–212, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg. `doi:10.1007/11548133_13`.
13. H. Peter Gumm and Tobias Schröder. Types and coalgebraic structure. *algebra universalis*, 53(2):229–252, 2005. `doi:10.1007/s00012-005-1888-2`.

14 Bart Jacobs. *Introduction to Coalgebra: Towards Mathematics of States and Observation*, volume 59 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 2016. doi:10.1017/CBO9781316823187.

15 Jean-Baptiste Jeannin, Dexter Kozen, and Alexandra Silva. Well-founded coalgebras, revisited. *Math. Struct. Comput. Sci.*, 27(7):1111–1131, 2017. doi:10.1017/S0960129515000481.

16 Alexander Kurz. *Logics for Coalgebras and Applications to Computer Science*. PhD thesis, Ludwig-Maximilians-Universität München, July 2000. URL: https://www.cs.le.ac.uk/people/akurz/LMU/Diss/all-s.ps.gz.

17 Gerhard Osius. Categorical set theory: A characterization of the category of sets. *Journal of Pure and Applied Algebra*, 4(1):79–119, 1974. doi:10.1016/0022-4049(74)90032-2.

18 Jan J. M. M. Rutten. Universal coalgebra: a theory of systems. *Theor. Comput. Sci.*, 249(1):3–80, 2000. doi:10.1016/S0304-3975(00)00056-6.

19 Nino Shervashidze, Pascal Schweitzer, Erik Jan van Leeuwen, Kurt Mehlhorn, and Karsten M. Borgwardt. Weisfeiler-lehman graph kernels. *J. Mach. Learn. Res.*, 12:2539–2561, 2011. doi:10.5555/1953048.2078187.

20 Paul Taylor. Intuitionistic sets and ordinals. *J. Symb. Log.*, 61(3):705–744, 1996. doi:10.2307/2275781.

21 Paul Taylor. *Practical Foundations of Mathematics*, volume 59 of *Cambridge studies in advanced mathematics*. Cambridge University Press, 1999.

22 Paul Taylor. Well founded coalgebras and recursion. accessed on 2024-01-27, 2023. URL: https://www.paultaylor.eu/ordinals/welfcr.pdf.

23 Věra Trnková. On a descriptive classification of set functors I. *Commentationes Mathematicae Universitatis Carolinae*, 12(1):143–174, 1971.

24 Thorsten Wißmann, Jérémy Dubut, Shin-ya Katsumata, and Ichiro Hasuo. Path category for free. In Mikołaj Bojańczyk and Alex Simpson, editors, *Proc. 22nd International Conference on Foundations of Software Science and Computation Structures (FoSSaCS 2019)*, volume 11425 of *LNCS*, pages 523–540, Cham, April 2019. Springer International Publishing. doi:10.1007/978-3-030-17127-8_30.

25 Thorsten Wißmann. Minimality notions via factorization systems and examples. *Logical Methods in Computer Science*, Volume 18, Issue 3, September 2022. doi:10.46298/lmcs-18(3:31)2022.

26 Thorsten Wißmann, Ulrich Dorsch, Stefan Milius, and Lutz Schröder. Efficient and modular coalgebraic partition refinement. *Logical Methods in Computer Science*, 16:1:8:1–8:63, January 2020. Selected Papers of the 28th International Conference on Concurrency Theory (CONCUR 2017). doi:10.23638/LMCS-16(1:8)2020.

27 Thorsten Wißmann, Stefan Milius, Shin-ya Katsumata, and Jérémy Dubut. A coalgebraic view on reachability. *Commentationes Mathematicae Universitatis Carolinae*, 60:4:605–638, December 2019. doi:10.14712/1213-7243.2019.026.