

Tree decomposition of Reeb graphs, parametrized complexity, and applications to phylogenetics

Anastasios Stefanou ¹

Abstract

Inspired by the interval decomposition of persistence modules and the extended Newick format of phylogenetic networks, we show that, inside the larger category of *ordered Reeb graphs*, every Reeb graph with n leaves and first Betti number s , is equal to a coproduct of at most 2^s trees with $(n + s)$ leaves. Reeb graphs are therefore classified up to isomorphism by their tree decomposition. An implication of this result, is that the isomorphism problem for Reeb graphs is fixed parameter tractable when the parameter is the first Betti number. We propose ordered Reeb graphs as a model for time consistent phylogenetic networks and propose a certain Hausdorff distance as a metric on these structures.

Key words. Coproducts, decomposition, complexity, Betti number, Reeb graphs, phylogenetic networks.

Acknowledgements

A. Stefanou was partially supported by the National Science Foundation through the grant NSF-CCF-1740761 TRIPODS TGDA@OSU, and also by the grant NSF DMS-1440386 Mathematical Biosciences Institute, at The Ohio State University. The author gratefully thanks two anonymous reviewers whose feedback significantly increased the quality of the manuscript. Furthermore, the author thankfully acknowledge F. Mémoli, E. Munch, J. Curry, S. Kurtek, W. KhudaBukhsh, and A. Foroughipour for many helpful discussions during the course of this work.

1 Introduction

Reeb graphs encode the evolution of connected components of a space \mathbb{X} along a real valued map f on \mathbb{X} [31]. Originated from Morse theory, Reeb graphs have been of particular interest to the fields of computational geometry [13], [2], [23], [18], and computational topology [28], [20], [14], [3], and they have found a plethora of applications in computer graphics and computer science [21], [24], [12], [22], [17], [35]. See [4] for a survey. One variation of Reeb graphs that has recently been proposed is Mapper [33] which has been quite successful on big data sets [30], [37].

1.1 Related work

de Silva et al. (2016) showed that any Reeb graph can be identified with a constructible **Set**-valued cosheaf on \mathbb{R} [16]. Thus, Reeb graphs can be thought of as generalized persistence modules in the setting of Bubenik et al. (2015) [8]. A *generalized persistence module* is any functor $\mathbf{F} : \mathcal{P} \rightarrow \mathcal{C}$ from a poset \mathcal{P} to a category \mathcal{C} [8]. When $\mathcal{P} = (\mathbb{R}, \leq)$ the poset of real numbers and $\mathcal{C} = \mathbf{vect}_k$ is the category of finite dimensional k -vector spaces, we obtain the notion of a *pointwise finite dimensional (p.f.d.) persistence module*. Crawley-Boevey (2015) has shown that every p.f.d. persistence module $\mathbf{V} : (\mathbb{R}, \leq) \rightarrow \mathbf{vect}_k$ decomposes into a direct sum of interval persistence modules [15]. The multiset of intervals $B(\mathbf{V})$ associated to \mathbf{V} is called *the barcode of \mathbf{V}* [11]. Because of this decomposition, one can easily check that the isomorphism complexity of p.f.d. persistence modules is polynomial.

However, this is no longer true for arbitrary generalized persistence modules. Bjerkevik et al. (2018) has shown that the isomorphism complexity of Reeb graphs is GI-complete [6], namely deciding if two Reeb graphs are isomorphic it is at least as hard as the graph isomorphism problem. However the graph isomorphism problem has shown to be

¹stefanou.3@osu.edu; 614-688-3198

ORCID id: <https://orcid.org/0000-0002-5408-9317>

Mathematical Biosciences Institute; Department of Mathematics
The Ohio State University

fixed parameter tractable with respect to several parameters, such as: tree-distance width [36], tree-depth [7], and tree-width [26]. Hence it is natural to wonder whether Reeb graphs, like graphs, are fixed parameter tractable with respect to some topologically meaningful parameter.

On the other hand, we can think of any Reeb graph as a weighted directed acyclic graph. Weighted directed acyclic graphs are used as the main method for modelling phylogenetic trees and networks [9], [5], [32], [25]. The isomorphism classes of phylogenetic trees are in one to one correspondence with nested parentheses, a method known today as the *Newick format*, which was already noticed by A. Cayley (1857). For general phylogenetic networks, Cardona et al. (2008) proposed a variant of the Newick format, called the *extended Newick format* [10]. The idea is: given a fixed ordering on the children nodes of a rooted phylogenetic network with n -labelled leaves and s reticulations (Betti number), we can represent that network as a phylogenetic tree with $(n + s)$ -labelled leaves where some of the leaves are allowed to have repeated nodes. A. Dress (2007) proposed a categorical approach to view phylogenetic networks called *X-nets* [19].

1.2 Our contribution

Inspired by the decomposition of p.f.d. persistence modules into interval persistence modules, we show that any Reeb graph decomposes into a coproduct of trees. The construction of each of these trees is an analogue of the extended Newick format in the setting of arbitrary Reeb graphs (not necessary rooted). First, in Sec. 2 and 3 we mention the basic definitions and tools from category theory [27], and the setting of Reeb graphs as studied in [16], which we need in order to formulate properly the tree decomposition of Reeb graphs. In Sec 4 we show that the category of Reeb graphs with a fixed edge structure forms a thin category inside of which every Reeb graph decomposes into a coproduct of trees. As an implication of this decomposition, we show that Reeb graph isomorphism is fixed parameter tractable where the parameter is the first Betti number.

2 Categorical structures

Category theory is fundamentally a language that formalizes mathematical structure having the capability of bridging together different mathematical constructions or theories. In this section we give the basic definitions and tools from category theory that we need.

2.1 Basic definitions

Category theory is a general theory of functions. A general notion of a function is called a *morphism* and the notion of a set is replaced by an *object*. An object can be any mathematical construction and its not necessary to be a set. In contrast with set theory the focus is concentrated in the study of morphisms between objects rather than just study the objects themselves. In particular, we require that morphisms between objects to have a composition operation that is associative and unital. The structure we obtain is said to be a *category*. A good source for an introduction to category theory is [27].

First we define the notion of a category. Here by a *class* we mean a collection of sets that is unambiguously defined by property that all these sets share in common. A class might not be a set and if that is the case is called a *proper class*.

Definition 2.1. A *category* \mathcal{C} consists of

- a class $\mathbf{Ob}\mathcal{C}$ whose elements X, Y, \dots are called *objects*, together with
- for each pair of objects X, Y in \mathcal{C} a set $\mathbf{Mor}_{\mathcal{C}}(X, Y)$, whose elements are called *morphisms* and denoted by $f : X \rightarrow Y$, and each having a unique source X and a unique target Y ,
- for each object X in \mathcal{C} an *identity* morphism $I_X : X \rightarrow X$,
- a binary operation $\circ : \mathbf{Mor}_{\mathcal{C}}(X, Y) \times \mathbf{Mor}_{\mathcal{C}}(Y, Z) \rightarrow \mathbf{Mor}_{\mathcal{C}}(X, Z)$, $(f, g) \mapsto g \circ f$ called *composition* which is associative and unital, i.e.

$$\begin{aligned} h \circ (g \circ f) &= (h \circ g) \circ f \\ f \circ I_X &= I_Y \circ f \end{aligned}$$

for any triple of morphisms $f : X \rightarrow Y$, $g : Y \rightarrow Z$ and $h : Z \rightarrow W$ in \mathcal{C} .

Definition 2.2. A morphism $f : X \rightarrow Y$ is said to be an *isomorphism* from X to Y if there exists a morphism $g : Y \rightarrow X$ (often called the inverse) such that $g \circ f = I_X$ and $f \circ g = I_Y$. Two objects are said to be isomorphic if there exists an isomorphism from X to Y .

Example 2.3. Examples of categories include:

- the category **Set** whose objects are sets and morphisms are functions between sets
- the category **Top** whose objects are topological spaces and morphisms are continuous maps.
- the category **Groups** whose objects are groups and morphisms are group homomorphisms
- the category **Ab** whose objects are abelian groups and morphisms are group homomorphisms

Definition 2.4. A category \mathcal{D} whose objects and morphisms are in \mathcal{C} and with the same identities and composition operation as of \mathcal{C} is said to be a *subcategory* of \mathcal{C} .

Let \mathcal{C} be a category and let \mathcal{S} be any subset of $\mathbf{Ob}\mathcal{C}$. Then we can consider the same sets $\mathbf{Mor}_{\mathcal{C}}(X, Y)$ of morphisms between $X, Y \in \mathcal{S}$. That way we obtain a category with the same morphisms but fewer objects. We say that \mathcal{S} forms a *full subcategory* of \mathcal{C} . In a full subcategory we only need to specify what are the objects so we often say \mathcal{S} is *the full subcategory of \mathcal{C} whose objects are in \mathcal{S}* . For example the category **Ab** is the full subcategory of **Groups** whose objects are abelian groups.

Example 2.5. The type of categories we work on are the following:

- *slice categories*: given a category \mathcal{C} and an object X we consider the slice category $\mathcal{C} \downarrow X$ whose objects are tuples (Y, f) where $Y \in \mathbf{Ob}\mathcal{C}$ and $f \in \mathbf{Mor}_{\mathcal{C}}(Y, X)$, and morphisms $\varphi : (Y, f) \rightarrow (Z, g)$ are ordinary morphisms $\varphi : Y \rightarrow Z$ in \mathcal{C} such that $g \circ \varphi = f$.
- *thin categories*: a category \mathcal{C} is called thin if for every pair of objects X, Y in \mathcal{C} there exists at most one morphism $f : X \rightarrow Y$ in \mathcal{C} . When a morphism $f : X \rightarrow Y$ exists we write $X \leq Y$. A thin category coincides with the notion of a *preorder*.

Now, we define the notion of maps that preserve the structure of a category.

Definition 2.6. A *functor* $\mathbf{F} : \mathcal{C} \rightarrow \mathcal{D}$ between categories consists of

- a function $\mathbf{F} : \mathbf{Ob}\mathcal{C} \rightarrow \mathbf{Ob}\mathcal{D}$, $X \mapsto \mathbf{F}(X)$, together with
- for each pair of objects X, Y in \mathcal{C} , a function

$$\begin{aligned} \mathbf{F}_{X,Y} : \mathbf{Mor}_{\mathcal{C}}(X, Y) &\rightarrow \mathbf{Mor}_{\mathcal{D}}(\mathbf{F}(X), \mathbf{F}(Y)) \\ f &\mapsto \mathbf{F}[f] \end{aligned}$$

such that for any object X and any morphisms $f : X \rightarrow Y$, $g : Y \rightarrow Z$ in \mathcal{C} :

$$\begin{aligned} \mathbf{F}[g \circ f] &= \mathbf{F}[g] \circ \mathbf{F}[f] \\ \mathbf{F}[I_X] &= I_{\mathbf{F}(X)}. \end{aligned}$$

When $\mathcal{C} = \mathcal{D}$, \mathbf{F} is called an *endofunctor*. A special case is the identity endofunctor $\mathbf{I}_{\mathcal{C}} : \mathcal{C} \rightarrow \mathcal{C}$ that sends each object and morphism to itself.

The collection of all functors from a category \mathcal{C} to a category \mathcal{D} forms a category on its own called a *functor category* and it is denoted by $[\mathcal{C}, \mathcal{D}]$: the objects are functors $\mathbf{F} : \mathcal{C} \rightarrow \mathcal{D}$ and the morphisms are natural transformations $\eta : \mathbf{F} \Rightarrow \mathbf{G}$.

Definition 2.7. A *natural transformation* $\eta : \mathbf{F} \Rightarrow \mathbf{G}$ consists of a family of morphisms $\eta_X : \mathbf{F}(X) \rightarrow \mathbf{G}(X)$ in \mathcal{D} one for each object X in \mathcal{C} , such that the diagram

$$\begin{array}{ccc} \mathbf{F}(X) & \xrightarrow{\eta_X} & \mathbf{G}(X) \\ \mathbf{F}[f] \downarrow & & \downarrow \mathbf{G}[f] \\ \mathbf{F}(Y) & \xrightarrow{\eta_Y} & \mathbf{G}(Y) \end{array}$$

commutes for every morphism $f : X \rightarrow Y$ in \mathcal{C} . In the special case where each η_X is an isomorphism in \mathcal{D} , then η is said to be a *natural isomorphism* and we write $\mathbf{F} \cong \mathbf{G}$.

Every time we write $\mathbf{F} \cong \mathbf{G}$ we mean there exists a natural isomorphism $\eta : \mathbf{F} \Rightarrow \mathbf{G}$.

Definition 2.8. A pair of categories \mathcal{C}, \mathcal{D} are said to be *equivalent* if there exist functors $\mathbf{F} : \mathcal{C} \rightarrow \mathcal{D}$ and $\mathbf{G} : \mathcal{D} \rightarrow \mathcal{C}$ such that $\mathbf{F} \circ \mathbf{G} \cong \mathbf{I}_{\mathcal{D}}$ and $\mathbf{G} \circ \mathbf{F} \cong \mathbf{I}_{\mathcal{C}}$. In the special case where $\mathbf{F} \circ \mathbf{G} = \mathbf{I}_{\mathcal{D}}$ and $\mathbf{G} \circ \mathbf{F} = \mathbf{I}_{\mathcal{C}}$, the categories \mathcal{C} and \mathcal{D} are said to be *isomorphic*.

2.2 Coproducts

We define the notion of a coproduct of objects in a category \mathcal{C} . This is the dual notion of a product [27]. However, here we focus only on the definition of coproducts since this is the only notion we use.

Definition 2.9. Let X_1, \dots, X_n be objects in \mathcal{C} . An object is called *the coproduct* of X_1, X_2, \dots, X_n , written $\coprod X_i$, if there exist morphisms $\iota_j : X_j \rightarrow \coprod X_i$, $j = 1, 2, \dots, n$ satisfying the following *universal property*: for any object Y and any pair of morphisms $f_j : X_j \rightarrow Y$, $j = 1, 2, \dots, n$, there exists a unique morphism $f : \coprod X_i \rightarrow Y$ such that the diagrams

$$\begin{array}{ccc} X_j & \xrightarrow{\iota_j} & \coprod X_i \\ & \searrow f_j & \downarrow f \\ & & Y \end{array}$$

commute for $j = 1, 2, \dots, n$.

Note that by the universal property of coproducts, the morphisms ι_j are uniquely defined up to a unique natural isomorphism. The morphisms $\iota_j : X_j \rightarrow \coprod X_i$ are called *coprojections*.

Definition 2.10. An object X is said to be *decomposable* if it is isomorphic to a coproduct of n objects in \mathcal{C} where $n \geq 2$. Otherwise X is said to be *indecomposable*.

Example 2.11. Here we give some basic examples of categorical coproducts.

- If \mathcal{C} is the category of all sets **Set**, then the coproduct is given by the disjoint union \coprod of sets.
- If \mathcal{C} is the category of groups **Groups** then the coproduct is given by the free product $*$ of groups.
- If \mathcal{C} is the category of abelian groups **Ab** then the coproduct is the direct sum \oplus of abelian groups.

3 Combinatorial structures

In this section we consider the setting of Reeb graphs as developed by V. de Silva et al. (2016) [16]. We define Reeb graphs and examine how they relate to directed acyclic graphs.

3.1 Reeb graphs

The main tool we use to visualize relationships among objects is a *graph*. A graph $G = (\mathbb{V}, \mathbb{E})$ consists of a collection \mathbb{V} of objects called *vertices*, e.g. v_1, v_2, \dots and a set \mathbb{E} of connections e_1, e_2, \dots between vertices called *edges*.

Generally we can define a Reeb graph as a connected graph \mathbb{X} together with a real valued map $f : \mathbb{X} \rightarrow \mathbb{R}$ which is strictly monotone when restricted to edges. However with this definition we are not making precise the exact way \mathbb{X} can be constructed in conjunction with the map f being monotone restricted to edges. Making this more precise is what we do in this paragraph.

First we need to talk about the general setting of \mathbb{R} -spaces. An \mathbb{R} -space (\mathbb{X}, f) is a space \mathbb{X} together with a real valued continuous map $f : \mathbb{X} \rightarrow \mathbb{R}$. A *morphism* of \mathbb{R} -spaces $(\mathbb{X}, f), (\mathbb{Y}, g)$ —also called a *function preserving map*—is an ordinary continuous map $\varphi : \mathbb{X} \rightarrow \mathbb{Y}$ such that $g \circ \varphi = f$. The collection of these objects forms the slice category $\mathbf{Top} \downarrow \mathbb{R}$. Now let us return to Reeb graphs.

Definition 3.1. A connected \mathbb{R} -space (\mathbb{X}, f) is said to be a *Reeb graph* if it is constructed by the following procedure, which we call *a structure on* (\mathbb{X}, f) :

Let $S = \{a_1 < \dots < a_k\}$ be an ordered subset of \mathbb{R} with $k \geq 2$.

- For each $i = 1, \dots, k$ we specify a non-empty set \mathbb{V}_i of vertices which lie over a_i ,
- For each $i = 1, \dots, k - 1$ we specify a non-empty set of edges \mathbb{E}_i which lie over $[a_i, a_{i+1}]$
- For $i = 1, \dots, k - 1$, we specify a ‘down map’ $D_i : \mathbb{E}_i \rightarrow \mathbb{V}_i$
- For $i = 1, \dots, k - 1$, we specify an ‘upper map’ $U_i : \mathbb{E}_i \rightarrow \mathbb{V}_{i+1}$.

The space \mathbb{X} is the quotient \mathbb{U}/\sim of the disjoint union

$$\mathbb{U} = \left(\prod_{i=1}^k (\mathbb{V}_i \times \{a_i\}) \right) \coprod \left(\prod_{i=1}^{k-1} (\mathbb{E}_i \times [a_i, a_{i+1}]) \right)$$

with respect to the identifications $(D_i(e), a_i) \sim (e, a_i)$ and $(U_i(e), a_{i+1}) \sim (e, a_{i+1})$, for all $i = 1, \dots, k - 1$, with the map f being the projection onto the second factor.

Remark 3.2. Note that to every Reeb graph corresponds a unique minimal set S , known as the *critical set* of \mathbb{X} [16]. We consider a definition for arbitrary set S because this allows for more flexibility: we can describe the morphisms between Reeb graphs easily if we consider a common set S for both Reeb graphs, e.g. by considering the union of their two S -sets.

The set $\mathbb{V} = \coprod_{i=1}^k \mathbb{V}_i$ is said to be a *vertex-set* for \mathbb{X} and the set $\mathbb{E} = \coprod_{i=1}^{k-1} \mathbb{E}_i$ is said to be a *edge-set* for \mathbb{X}^2 . If we forget the map f associated to a Reeb graph, then topologically \mathbb{X} forms a graph (\mathbb{V}, \mathbb{E}) on its own. See Fig. 1 for an example of a Reeb graph.

Definition 3.3. A *morphism* of Reeb graphs (\mathbb{X}, f) and (\mathbb{Y}, g) is any morphism $\varphi : (\mathbb{X}, f) \rightarrow (\mathbb{Y}, g)$ between these \mathbb{R} -spaces.

Thus, the collection of all Reeb graphs forms a full subcategory **Reeb** of $\mathbf{Top} \downarrow \mathbb{R}$. As shown in [16] Reeb graphs can be identified with constructible **Set**-valued cosheaves on \mathbb{R} . This equivalence of categories allows for one to consider the following combinatorial description of the morphisms of Reeb graphs.

Proposition 3.4 (Prop 3.12 in [16]). Let $(\mathbb{X}, f), (\mathbb{Y}, g)$ be a pair of Reeb graphs with a common set $S = \{a_1 < \dots < a_k\}$, let $\mathbb{V}^{\mathbb{X}}, \mathbb{V}^{\mathbb{Y}}$, and $\mathbb{E}^{\mathbb{X}}, \mathbb{E}^{\mathbb{Y}}$ be their vertex-sets and edge-sets respectively. Any function preserving map $\varphi : (\mathbb{X}, f) \rightarrow (\mathbb{Y}, g)$ of Reeb graphs is completely determined by

- Functions $\varphi_i^{\mathbb{V}} : \mathbb{V}_i^{\mathbb{X}} \rightarrow \mathbb{V}_i^{\mathbb{Y}}$
- Functions $\varphi_i^{\mathbb{E}} : \mathbb{E}_i^{\mathbb{X}} \rightarrow \mathbb{E}_i^{\mathbb{Y}}$, satisfying the
- Consistency conditions: $\varphi_i^{\mathbb{V}} D_i^{\mathbb{X}} = D_i^{\mathbb{Y}} \varphi_i^{\mathbb{E}}$ and $\varphi_{i+1}^{\mathbb{V}} U_i^{\mathbb{X}} = U_i^{\mathbb{Y}} \varphi_i^{\mathbb{E}}$ for all $1 \leq i \leq k-1$.

Any function preserving map $\varphi : (\mathbb{X}, f) \rightarrow (\mathbb{Y}, g)$, since f and g are by definition the projections to the second coordinate of \mathbb{X} and \mathbb{Y} respectively, is given by

$$\begin{aligned} \varphi : (\mathbb{X}, f) &\rightarrow (\mathbb{Y}, g) \\ [(v, t)] &\mapsto [(\varphi_i^{\mathbb{V}}(v), t)], \text{ for all } v \in \mathbb{V}_i^{\mathbb{X}}, \text{ for all } i = 1, \dots, k. \\ [(e, t)] &\mapsto [(\varphi_i^{\mathbb{E}}(e), t)], \text{ for all } e \in \mathbb{E}_i^{\mathbb{X}}, \text{ for all } i = 1, \dots, k-1. \end{aligned}$$

3.2 Reeb graphs viewed as directed acyclic graphs

If we allow the edges e of a graph G to have a direction, i.e. $x \xrightarrow{e} y$ then the resulting graph is said to be a *digraph* and the edges are called *directed edges* or *arrows*. A *directed path of length n* on a digraph is a sequence of arrows $x_0 \rightarrow x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow \dots \rightarrow x_{n-1} \rightarrow x_n$ in G . A directed path that starts and end at the same vertex is called a *directed cycle*. A digraph with no directed cycles is said to be a *directed acyclic graph (DAG)*.

With the map $f : \mathbb{X} \rightarrow \mathbb{R}$, we give a direction to each edge e connecting x_1 and x_2 in \mathbb{X} , by declaring $e : x_1 \rightarrow x_2$ whenever $f(x_1) > f(x_2)$. That way, the underlying graph (\mathbb{V}, \mathbb{E}) of a Reeb graph \mathbb{X} obtains the structure of a directed acyclic graph. Furthermore, each vertex v of \mathbb{X} receives a real weight $w(v) := f(v)$ via f . So, every Reeb graph can be thought of as a vertex-weighted DAG.

Remark 3.5. Note, in particular, for Reeb graphs, because of the vertex-weight $w(v) := f(v)$ each directed edge $e : x_1 \rightarrow x_2$ receives a strictly positive weight $w(e) := f(x_1) - f(x_2) > 0$, where $f(x_1) = a_{i+1}$ and $f(x_2) = a_i$, for some $i = 1, \dots, k-1$, are consecutive critical numbers of (\mathbb{X}, f) .

Let (\mathbb{X}, f) be a Reeb graph with critical set $S = \{a_1, \dots, a_k\}$. Let x be a vertex of \mathbb{X} with $f(x) = a_i$. Then an edge $e \in D_i^{-1}(x)$ ($e \in U_{i-1}^{-1}(x)$) is said to be an edge incident from above (below) of x . Also the number of edges $\mathbf{indeg}(x) = |D_i^{-1}(x)|$ that are incident from above of x is called the *indegree of x* . Similarly, the number of edges $\mathbf{outdeg}(x) = |U_{i-1}^{-1}(x)|$ is called the *outdegree of x* .

There are three cases that can happen for a node x :

- If $\mathbf{indeg}(x) \leq 1$, then x is said to be a *tree-vertex*.
- If $\mathbf{indeg}(x) \geq 2$ and with x a cycle in the Reeb graph (viewed as a DAG) closes (x is the ‘bottom’ of a cycle), then x is said to be a *reticulation-vertex*.
- If $\mathbf{indeg}(x) \geq 2$ and with x no cycles closes in the Reeb graph, then x again is said to be a *tree-vertex*.

²the reason we call these sets in this way, e.g. we say ‘a vertex-set’ instead of ‘the vertex-set’, is because they depend on the choice of S .

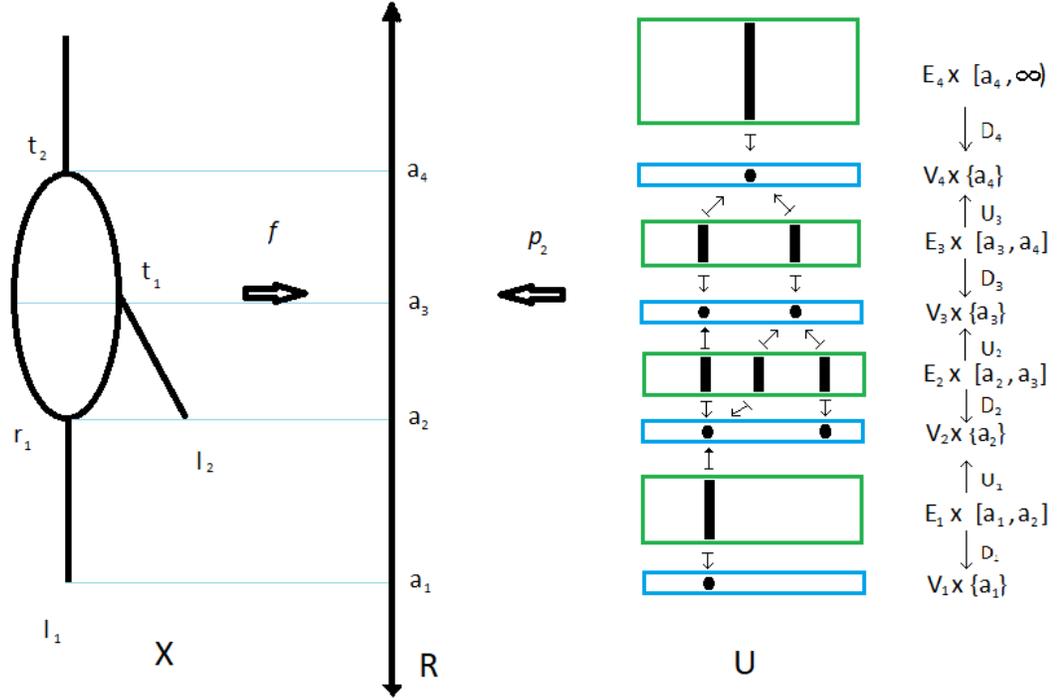


Figure 1: An example of a Reeb graph

A vertex x such that $\mathbf{indeg}(x) = \mathbf{outdeg}(x) = 1$ is called *regular*. We denote by $T(\mathbb{X})$ the set of tree-vertices of \mathbb{X} , and by $R(\mathbb{X})$ the set of all reticulation-vertices of \mathbb{X} . If $(\mathbf{outdeg}(x) = 0)$ or $(\mathbf{outdeg}(x) = 1$ and $\mathbf{indeg}(x) = 0)$, then x is said to be a *leaf*. By definition, it may happen that a leaf is also a reticulation-vertex (i.e. $\mathbf{outdeg}(x) = 0$ and $\mathbf{indeg}(x) \geq 2$).

Remark 3.6. The reason we consider this definition of leaves is so that the tree decomposition of Reeb graphs to work for this type of reticulation vertices. This would be more clear in Ex. 4.12.

We denote by $L(\mathbb{X})$ the set of all leaves of \mathbb{X} . Then, $L(\mathbb{X}) \subset T(\mathbb{X})$.

Viewing Reeb graphs as DAGs is also important for computing the Betti number $s = B_1(\mathbb{X})$ of a Reeb graph \mathbb{X} . The Betti number counts the minimum number of cycles of a graph that generate all possible cycles in the graph. By definition, a Reeb graph (\mathbb{X}, f) is a *tree* if and only if it has no reticulation-vertices (and thus, no cycles). Since \mathbb{X} is always connected by definition, the Euler characteristic provides the formula

$$s = |\mathbb{E}| - |\mathbb{V}| + 1,$$

Furthermore, from the theory of directed graphs we have the *degree sum formula*

$$\sum_{x \in \mathbb{V}} \mathbf{indeg}(x) = \sum_{x \in \mathbb{V}} \mathbf{outdeg}(x) = |\mathbb{E}|$$

By combining these two equations we get

$$s = 1 + \sum_{x \in \mathbb{V}} (\mathbf{indeg}(x) - 1).$$

Proposition 3.7. Let r_1, \dots, r_m be the reticulation-vertices of \mathbb{X} with indegrees $d_1, \dots, d_m \geq 2$ respectively. Then

$$s = \sum_{i=1}^m (d_i - 1).$$

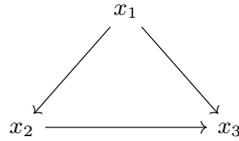
Proof. Thinking of \mathbb{X} as a DAG, then if we remove for each reticulation-vertex r_i all of its incident from above edges except one, from the DAG \mathbb{X} , then we get back a directed tree $\mathbb{T}_{\mathbb{X}} = (\mathbb{V}^{\mathbb{T}_{\mathbb{X}}}, \mathbb{E}^{\mathbb{T}_{\mathbb{X}}})$ having the same tree-vertices and their incident from above edges as in \mathbb{X} plus another m -additional vertices, r_1, \dots, r_m , that now are viewed as tree-vertices, each having indegree 1. This implies that $T(\mathbb{V}_{\mathbb{X}}) \cup \{r_1, \dots, r_m\} = \mathbb{V}^{\mathbb{T}_{\mathbb{X}}}$ and $\sum_{i=1}^m (\text{indeg}_{\mathbb{T}_{\mathbb{X}}}(r_i) - 1) = 0$. We compute

$$\begin{aligned}
s &= 1 + \sum_{x \in \mathbb{V}} (\text{indeg}(x) - 1) \\
&= 1 + \sum_{x \in T(\mathbb{V})} (\text{indeg}(x) - 1) + \sum_{x \in R(\mathbb{V})} (\text{indeg}(x) - 1) \\
&= 1 + \sum_{x \in T(\mathbb{V})} (\text{indeg}(x) - 1) + 0 + \sum_{i=1}^m (d_i - 1) \\
&= 1 + \sum_{x \in T(\mathbb{V})} (\text{indeg}(x) - 1) + \sum_{i=1}^m (\text{indeg}_{\mathbb{T}_{\mathbb{X}}}(r_i) - 1) + \sum_{i=1}^m (d_i - 1) \\
&= 1 + \sum_{x \in \mathbb{V}^{\mathbb{T}_{\mathbb{X}}}} (\text{indeg}(x) - 1) + \sum_{i=1}^m (d_i - 1) \\
&= s^{\mathbb{T}_{\mathbb{X}}} + \sum_{i=1}^m (d_i - 1) \\
&= 0 + \sum_{i=1}^m (d_i - 1) \\
&= \sum_{i=1}^m (d_i - 1).
\end{aligned}$$

□

Finally, note that although any Reeb graph can be thought of as a vertex-weighted DAG, the other direction is not true: not every vertex-weighted DAG is a Reeb graph.

Example 3.8. Consider the vertex-weighted DAG, $G = (V(G), E(G), w)$, that has edges $e_1 = (x_1, x_2)$, $e_2 = (x_2, x_3)$ connecting x_1 with x_2 and x_2 with x_3 and an edge $e_3 = (x_1, x_3)$ connecting directly x_1 with x_3 with a single edge, and such that $w(x_1) = 2$, and $w(x_2) = w(x_3) = 1$. Then we claim that there exists no Reeb graph (\mathbb{X}, f) such that $V(\mathbb{X}) = V(G)$ and $f|_{V(G)} = w$. Indeed assume the contrary there is one such f . Then $f(x_1) = 2$ and $f(x_2) = f(x_3) = 1$. By definition of G , there exists an edge $e_2 : x_2 \rightarrow x_3$. By Remk. 3.5 we get a strictly positive edge-weight $w(e_2) > 0$. However we compute $w(e_2) := f(x_2) - f(x_3) = w(x_2) - w(x_3) = 1 - 1 = 0$, a contradiction. Hence the weighted DAG G cannot be realized as a Reeb graph. This DAG it is not consistent with ‘time’ in the sense that the edge $e_2 = (x_2, x_3)$ represents a change in the nodes from x_2 to x_3 that happen instantaneously. In other words the edge e_2 is a ‘horizontal’ edge. Reeb graphs cannot have horizontal edges, from their construction.



4 Classifying Reeb graphs up to isomorphism

We show that inside a larger category of Reeb graphs, any Reeb graph is a coproduct of trees.

4.1 Ordered Reeb graphs

Definition 4.1. An *ordered Reeb graph* $(\mathbb{X}, f, \leq_{\mathbb{X}})$ is an ordinary Reeb graph (\mathbb{X}, f) such that its edge sets and its vertex sets as in Defn. 3.1 are in particular partially ordered sets (posets), i.e. $(\mathbb{E}_i, \leq_{\mathbb{E}_i})$, $i = 1, \dots, k-1$, and $(\mathbb{V}_i, \leq_{\mathbb{V}_i})$, $i = 1, \dots, k$, and also the down maps and upper maps preserve the partial orders, i.e. $U_i : (\mathbb{E}_i, \leq_{\mathbb{E}_i}) \rightarrow (\mathbb{V}_{i+1}, \leq_{\mathbb{V}_{i+1}})$, and $D_i : (\mathbb{E}_i, \leq_{\mathbb{E}_i}) \rightarrow (\mathbb{V}_i, \leq_{\mathbb{V}_i})$, for $i = 1, \dots, k-1$. The partial orders of the edge posets and vertex posets induce a partial order both on the disjoint union and the quotient space, namely the Reeb graph, denoted by $\leq_{\mathbb{X}}$. Indeed, it

is known that the category **Pos** of posets, just like sets, admit coequalizers and therefore quotients (see Joy of Cats, pg 119, [1]).

Definition 4.2. We define the category **Reeb**^{ord} whose

- objects $(\mathbb{X}, f, \leq_{\mathbb{X}})$ are ordered Reeb graphs.
- morphisms $\varphi : (\mathbb{X}, f, \leq_{\mathbb{X}}) \rightarrow (\mathbb{Y}, g, \leq_{\mathbb{Y}})$ are ordinary morphisms of Reeb graphs $\varphi : (\mathbb{X}, f) \rightarrow (\mathbb{Y}, g)$ that preserve the partial orders of the edge posets and vertex posets, i.e. $\varphi^{\mathbb{E}_i} : (\mathbb{E}_i^{\mathbb{X}}, \leq_{\mathbb{E}_i^{\mathbb{X}}}) \rightarrow (\mathbb{E}_i^{\mathbb{Y}}, \leq_{\mathbb{E}_i^{\mathbb{Y}}})$, $i = 1, \dots, k-1$, and $\varphi^{\mathbb{V}_i} : (\mathbb{V}_i^{\mathbb{X}}, \leq_{\mathbb{V}_i^{\mathbb{X}}}) \rightarrow (\mathbb{V}_i^{\mathbb{Y}}, \leq_{\mathbb{V}_i^{\mathbb{Y}}})$, $i = 1, \dots, k$, are order preserving maps.
- composition is defined in the obvious way.

Any finite set $A = \{x_1, \dots, x_n\}$ can be trivially thought of as a poset (A, \leq) by considering $x_i \leq x_j \Leftrightarrow x_i = x_j$. Namely, the only inequalities are the identities. Thus, any Reeb graph (\mathbb{X}, f) has its edge sets and vertex sets trivially partially ordered, i.e. the inequalities are only the identities. Hence any Reeb graph (\mathbb{X}, f) is an object of **Reeb**^{ord}. That is, $\mathbf{Ob}(\mathbf{Reeb}) \subset \mathbf{Ob}(\mathbf{Reeb}^{\text{ord}})$. In particular **Reeb** is a full subcategory of **Reeb**^{ord}. Indeed, let $\varphi : (\mathbb{X}, f) \rightarrow (\mathbb{Y}, g)$ be a morphism in **Reeb**. Then, again trivially we can think of φ —when restricted to the edge sets and vertex sets respectively—as an order preserving map between trivially partially ordered sets. Therefore **Reeb** is a full subcategory of **Reeb**^{ord}.

4.2 Ordered Reeb graphs with a fixed edge structure

Fix an ordered subset $S = \{a_1 < \dots < a_k\}$ of \mathbb{R} . Consider an ordered Reeb graph $(\mathbb{X}, f, \leq_{\mathbb{X}})$ with edge posets $(\mathbb{E}, \leq_{\mathbb{E}}) := (\mathbb{E}_1, \leq_{\mathbb{E}_1}) \amalg \dots \amalg (\mathbb{E}_{k-1}, \leq_{\mathbb{E}_{k-1}})$, where \mathbb{E}_i , $i = 1, \dots, k-1$, as in Defn. 3.1. We call $\mathbb{E}_{\bullet} = (\mathbb{E}_1, \dots, \mathbb{E}_{k-1})$ an *edge sequence* of the ordered Reeb graph (\mathbb{X}, f) with respect to S .

Given a common set $S = \{a_1 < \dots < a_k\}$ for a pair of Reeb graphs $(\mathbb{X}, f, \leq_{\mathbb{X}})$ and $(\mathbb{Y}, g, \leq_{\mathbb{Y}})$, we say that their edge sequences $\mathbb{E}_{\bullet}^{\mathbb{X}} = (\mathbb{E}_1^{\mathbb{X}}, \dots, \mathbb{E}_{k-1}^{\mathbb{X}})$ and $\mathbb{E}_{\bullet}^{\mathbb{Y}} = (\mathbb{E}_1^{\mathbb{Y}}, \dots, \mathbb{E}_{k-1}^{\mathbb{Y}})$ are *equivalent*, and denote it by $\mathbb{E}_{\bullet}^{\mathbb{X}} \cong \mathbb{E}_{\bullet}^{\mathbb{Y}}$, if $\mathbb{E}_i^{\mathbb{X}} \cong \mathbb{E}_i^{\mathbb{Y}}$, for all $i = 1, \dots, k-1$, as sets. In other words, $(\mathbb{X}, f, \leq_{\mathbb{X}})$ and $(\mathbb{Y}, g, \leq_{\mathbb{Y}})$ have equivalent edge sequences if $|\mathbb{E}_i| = |\mathbb{E}'_i|$ for all $i = 1, \dots, k-1$, where $|\cdot|$ denotes the cardinality.

Definition 4.3. Let (\mathbb{X}, f) and (\mathbb{Y}, g) be two Reeb graphs with common set S . We say (\mathbb{X}, f) , (\mathbb{Y}, g) have the *same edge structure*, if their edge sequences are equivalent.

Note that the relation ‘ $(\mathbb{X}, f) \sim (\mathbb{Y}, g) \Leftrightarrow (\mathbb{X}, f)$, (\mathbb{Y}, g) have the same edge structure’ forms an equivalence relation. This equivalence relation induces a partition on the objects of **Reeb**, i.e. we get

$$\mathbf{Ob}(\mathbf{Reeb}) = \coprod_{a \in A} [(\mathbb{X}_a, f_a)],$$

for some index set A , where $[\]$ denotes the \sim -equivalence class. This fact suggests that each of these blocks can be turned into a category on its own.

Fix an edge-sequence \mathbb{E}_{\bullet} .

Definition 4.4. We define **Reeb**^{ord} $[\mathbb{E}_{\bullet}]$ the category whose

- objects $(\mathbb{X}, f, \leq_{\mathbb{X}}, \mu^{\mathbb{X}})$ are ordered Reeb graphs $(\mathbb{X}, f, \leq_{\mathbb{X}})$ together with a family $\mu^{\mathbb{X}} = \{\mu_i^{\mathbb{X}}\}_{i=1}^{k-1}$ of bijections $\mu_i^{\mathbb{X}} : \mathbb{E}_i^{\mathbb{X}} \rightarrow \mathbb{E}_i$, for all $i = 1, \dots, k-1$, called an *E-edge labelling*, or simply an *edge labelling* if \mathbb{E}_{\bullet} is given.
- the morphisms are ordinary morphisms $\varphi : (\mathbb{X}, f, \leq_{\mathbb{X}}) \rightarrow (\mathbb{Y}, g, \leq_{\mathbb{Y}})$ in **Reeb**^{ord} that preserve the edge-labeling, namely $\mu^{\mathbb{Y}}(\varphi_j^{\mathbb{E}}(e)) = \mu^{\mathbb{X}}(e)$ for all $e \in \mathbb{E}_j^{\mathbb{X}}$ and all $j = 1, \dots, k-1$.
- composition is defined in the obvious way.

Lemma 4.5. The category **Reeb**^{ord} $[\mathbb{E}_{\bullet}]$ is thin.

Proof. Let $\varphi, \psi : (\mathbb{X}, f, \leq_{\mathbb{X}}, \mu^{\mathbb{X}}) \rightarrow (\mathbb{Y}, g, \leq_{\mathbb{Y}}, \mu^{\mathbb{Y}})$ be two morphisms in **Reeb**^{ord} $[\mathbb{E}_{\bullet}]$. By Prop. 3.4 we have to show that the maps agree at the edge posets and vertex posets. By definition $\varphi = \psi = (\mu^{\mathbb{Y}})^{-1} \circ (\mu^{\mathbb{X}})$ when restricted to edge posets, because the labellings are bijections.

Let v be a node in \mathbb{X} , say $v \in \mathbb{V}_i^{\mathbb{X}}$ for some i . We claim that $\varphi_i^{\mathbb{V}}(v) = \psi_i^{\mathbb{V}}(v)$. Now, v is either the down image or the upper image of some edge $e \in \mathbb{E}_i^{\mathbb{X}}$, for some $i = 1, \dots, k-1$. So, we have two cases:

- Case 1: $v = D_i^{\mathbb{X}}(e)$. By Prop. 3.4 we have that

$$\begin{aligned}
\varphi(v) &= \varphi_i^{\mathbb{V}}(D_i^{\mathbb{X}}(e)) \\
&= D_i^{\mathbb{Y}}(\varphi_i^{\mathbb{E}}(e)) \\
&= D_i^{\mathbb{Y}}((\mu^{\mathbb{Y}})^{-1}(\mu^{\mathbb{X}}(e))) \\
&= D_i^{\mathbb{Y}}(\psi_i^{\mathbb{E}}(e)) \\
&= \psi_i^{\mathbb{V}}(D_i^{\mathbb{X}}(e)) \\
&= \psi(v).
\end{aligned}$$

- Case 2: $v = U_i^{\mathbb{X}}(e)$. The proof of case 2 is similar to that of Case 1 and is omitted. □

Remark 4.6. We define the full subcategory $\mathbf{Reeb}[\mathbb{E}_{\bullet}]$ of $\mathbf{Reeb}^{ord}[\mathbb{E}_{\bullet}]$ whose objects are Reeb graphs with the trivial partial order (i.e. ordinary Reeb graphs) and with the same edge structure as \mathbb{E}_{\bullet} . By definition of \sim we observe that (\mathbb{X}, f) and (\mathbb{Y}, g) have the same edge structure if and only if both (\mathbb{X}, f) and (\mathbb{Y}, g) are objects of $\mathbf{Reeb}[\mathbb{E}_{\bullet}]$, for some edge-sequence \mathbb{E}_{\bullet} .

4.3 Tree-decomposition

Theorem 4.7. Fix an edge-sequence \mathbb{E}_{\bullet} . Let $n, s \geq 0$. Inside the larger category $\mathbf{Reeb}^{ord}[\mathbb{E}_{\bullet}]$, any Reeb graph (\mathbb{X}, f, μ) in $\mathbf{Reeb}[\mathbb{E}_{\bullet}]$, with n leaves and Betti number $B_1(\mathbb{X}) = s$, is a coproduct of ordered trees with $(n + s)$ -leaves and same \mathbb{E} -edge labelling, i.e.

$$(\mathbb{X}, f, \mu) \cong \coprod_{\mathbb{T} \in \mathcal{T}(\mathbb{X})} (\mathbb{T}, f_{\mathbb{T}}, \leq_{\mathbb{T}}, \mu),$$

for some set of ordered trees $\mathcal{T}(\mathbb{X})$.

Proof. Let $\mathbb{X} = (\mathbb{X}, f, \mu)$ be a Reeb graph in $\mathbf{Reeb}[\mathbb{E}_{\bullet}]$ with n -leaves and $B_1(\mathbb{X}) = s$. Let $\leq_{\mathbb{X}}$ be its trivial partial order. Let $L(\mathbb{X}) = \{l_1, \dots, l_t\}$, $R(\mathbb{X}) = \{r_1, \dots, r_m\}$ be its sets of leaves, and reticulation-vertices respectively. Let d_i be the indegree of the reticulation-vertex r_i for all $i = 1, \dots, m$. Also for any reticulation-vertex r_i , $i = 1, \dots, m$, let us denote by $\{e_1^{(i)}, \dots, e_{d_i}^{(i)}\}$ the set of all edges that are incident of r_i from above it. The basic idea of the proof is to construct a collection of ordered trees with $(n + s)$ -leaves, and same edge labelling as \mathbb{X} , out of \mathbb{X} ,

$$\mathcal{T}(\mathbb{X}) = \{(\mathbb{T}_{(w_1, \dots, w_m)}, f_{(w_1, \dots, w_m)}, \leq_{(w_1, \dots, w_m)}, \mu) : 1 \leq w_i \leq d_i, \text{ for all } i = 1, \dots, m\}$$

by breaking up in all possible ways the reticulation-vertices (the bottom of the cycles) of \mathbb{X} —without changing the connectivity of the graph—in order to create a tree out of \mathbb{X} , by introducing new leaves. Consider $\mathbb{V}_j^{\mathbb{X}}$ and $\mathbb{E}_j^{\mathbb{X}}$ and the upper and down maps $U_j^{\mathbb{X}}$ and $D_j^{\mathbb{X}}$, $j = 1, \dots, k - 1$ as in Defn. 3.1. For simplicity of the proof, we denote any m -tuple (w_1, \dots, w_m) by w .

Let $w \in \prod_{i=1}^m \{1, \dots, d_i\}$. We construct an ordered tree $\mathbb{T} = (\mathbb{T}_w, f_w, \leq_w, \mu)$ with $(n + s)$ -leaves, by changing the structure of \mathbb{X} so as to make an ordered tree following the steps below:

1. We define

- for any $j = 1, \dots, k - 1$, the set $\mathbb{E}_j^{\mathbb{T}_w} := \mathbb{E}_j^{\mathbb{X}}$ equipped with the trivial partial order $\leq_{\mathbb{E}_j^{\mathbb{X}}}$. That way, the edge sequence (and thus the edge structure) remains the same, i.e. $\mathbb{E}_j^{\mathbb{T}_w} = \mathbb{E}_j^{\mathbb{X}}$, for all $j = 1, \dots, k - 1$. Moreover consider the same \mathbb{E} -edge labelling $\mu^{\mathbb{T}_w} := \mu$, as for \mathbb{X} .
- for any $j = 1, \dots, k$, the vertex poset $(\mathbb{V}_j^{\mathbb{T}_w}, \leq_{\mathbb{V}_j^{\mathbb{T}_w}})$, where⁴

$$\mathbb{V}_j^{\mathbb{T}_w} := \mathbb{V}_j^{\mathbb{X}} \coprod \left(\prod_{f(r_i)=a_j} \{e_t^{(i)} : 1 \leq t \leq d_i \text{ and } t \neq w_i\} \right),$$

and the partial order $\leq_{\mathbb{V}_j^{\mathbb{T}_w}}$ is given by the identities on the elements, and the additional formal inequalities $r_i <_w e_t^{(i)}$, for all $1 \leq t \leq d_i$ with $t \neq w_i$.

⁴if $f(r_i) \neq a_j$, for all j , then the right component of the direct sum is the emptyset and we get $\mathbb{V}_j^{\mathbb{T}_w} = \mathbb{V}_j^{\mathbb{X}}$.

Remark 4.8. These nontrivial inequalities formalize the idea that the vertex r_i is isolated from the ‘new vertices’ (leaf nodes) that are formed after cutting the reticulation vertex r_i of \mathbb{X} while keeping it connected from above with the edge $e_{w_i}^{(i)}$. This observation is crucial since, on one hand the trees \mathbb{T}_w are distinguished for all choices of w , and on the other hand this makes the coproduct well defined as we will see, e.g. in Ex. 4.12.

- the function $D_j^{\mathbb{T}_w} : \mathbb{E}_j^{\mathbb{T}_w} \rightarrow \mathbb{V}_j^{\mathbb{T}_w}$ given by

$$D_j^{\mathbb{T}_w}(e) = \begin{cases} D_j^{\mathbb{X}}(e), & \text{if } D_j^{\mathbb{X}}(e) \text{ is not a reticulation-vertex} \\ r_i, & \text{if } e = e_{w_i}^{(i)}, \text{ for some } i = 1, \dots, m, \\ e, & \text{if } D_j^{\mathbb{X}}(e) = r_i \text{ and } e \neq e_{w_i}^{(i)}, \text{ for all } i = 1, \dots, m. \end{cases}$$

- the function $U_j^{\mathbb{T}_w} : \mathbb{E}_j^{\mathbb{T}_w} \rightarrow \mathbb{V}_{j+1}^{\mathbb{T}_w}$ given by

$$U_j^{\mathbb{T}_w}(e) = U_j^{\mathbb{X}}(e).$$

Note that since the edge sets are trivially partially ordered, as such, these functions are trivially order preserving, i.e. $D_j^{\mathbb{T}_w} : (\mathbb{E}_j^{\mathbb{T}_w}, \leq_{\mathbb{E}_j^{\mathbb{T}_w}}) \rightarrow (\mathbb{V}_j^{\mathbb{T}_w}, \leq_{\mathbb{V}_j^{\mathbb{T}_w}})$ and $U_j^{\mathbb{T}_w} : (\mathbb{E}_j^{\mathbb{T}_w}, \leq_{\mathbb{E}_j^{\mathbb{T}_w}}) \rightarrow (\mathbb{V}_{j+1}^{\mathbb{T}_w}, \leq_{\mathbb{V}_{j+1}^{\mathbb{T}_w}})$ for all j .

2. \mathbb{T}_w is the quotient of the disjoint union

$$\left(\prod_{j=1}^k \mathbb{V}_j^{\mathbb{T}_w} \times \{a_j\} \right) \amalg \left(\prod_{j=1}^{k-1} (\mathbb{E}_j^{\mathbb{T}_w} \times [a_i, a_{i+1}]) \right)$$

with respect to the identifications $(U_j^{\mathbb{T}_w}(e), a_{j+1}) \sim (e, a_{j+1})$ and $(D_j^{\mathbb{T}_w}(e), a_j) \sim (e, a_j)$. Define f_w to be the projection of \mathbb{T}_w to the second coordinate. Both the disjoint union and the resulting quotient space receive a partial order from the partial orders on $\mathbb{V}_j^{\mathbb{T}_w}$ and $\mathbb{E}_j^{\mathbb{T}_w}$. We denote the resulting partial order on \mathbb{T}_w by $\leq_{\mathbb{T}_w}$.

Remark 4.9. Note also that, by definition of $D_j^{\mathbb{T}_w}$, the identifications $(D_j^{\mathbb{T}_w}(e), a_j) \sim (e, a_j)$ are the trivial ones $(e, a_j) \sim (e, a_j)$ whenever $e \neq e_{w_i}^{(i)}$, for all $i = 1, \dots, m$. That formally expresses the fact that the tree \mathbb{T}_w is constructed from \mathbb{X} by cutting the bottom of all cycles, keeping, for all $1 \leq i \leq m$, only the edge $e_{w_i}^{(i)}$ connected to the reticulation-vertex r_i and disconnecting the others.

To sum up, \mathbb{T}_w forms an ordered tree $(\mathbb{T}_w, f_w, \leq_w, \mu)$ in $\mathbf{Reeb}^{ord}[\mathbb{E}_\bullet]$.

Furthermore, each ordered tree $(\mathbb{T}_w, f_w, \leq_w, \mu)$ is equipped with the obvious quotient map

$$\begin{aligned} q_w : (\mathbb{T}_w, f_w, \leq_w, \mu) &\rightarrow (\mathbb{X}, f, \leq_{\mathbb{X}}, \mu), \text{ where} \\ q_w^{\mathbb{E}} : \mathbb{E}^{\mathbb{T}_w} &\rightarrow \mathbb{E}^{\mathbb{X}} \\ &e \mapsto e, \text{ and} \\ q_w^{\mathbb{V}} : \mathbb{V}^{\mathbb{T}_w} &\rightarrow \mathbb{V}^{\mathbb{X}} \\ &v \mapsto v, \text{ for all } v \in \mathbb{V}^{\mathbb{X}} \\ e_t^{(i)} &\mapsto r_i, \text{ for all } 1 \leq t \leq d_i \text{ with } t \neq w_i, \text{ for all } i = 1, \dots, m. \end{aligned}$$

By definition, the quotient map q_w glues back the new vertices of the tree \mathbb{T}_w —yielded by the w -cut—to form the original Reeb graph \mathbb{X} . Note that no matter what the order of the new vertices is, from the cut of r_i , via the quotient map q_w they are glued to a single vertex r_i , hence q_w trivially preserves the partial orders. Namely, q_w kills the only non-trivial inequalities $r_i \leq e_t^{(i)}$, for all $t \neq w_i$, and $i = 1, \dots, m$. Moreover, by construction, \mathbb{T}_w has the same edge sets as \mathbb{X} , thus making q_w to be trivially \mathbb{E} -labelling preserving. So indeed q_w is a morphism in $\mathbf{Reeb}^{ord}[\mathbb{E}_\bullet]$.

Now we claim that the coproduct of all the trees $(\mathbb{T}_w, f_w, \leq_w, \mu)$ in $\mathbf{Reeb}^{ord}[\mathbb{E}_\bullet]$ is isomorphic to the ordered Reeb graph $(\mathbb{X}, f, \leq_{\mathbb{X}}, \mu)$ (where, again, $\leq_{\mathbb{X}}$ is the trivial order of the Reeb graph) with the coprojections morphisms being the quotient maps q_w .

Let $\varphi_w : (\mathbb{T}_w, f_w, \leq_w, \mu) \rightarrow (\mathbb{Y}, g, \leq_{\mathbb{Y}}, \mu^{\mathbb{Y}})$ be any w -indexed family of morphisms in $\mathbf{Reeb}^{ord}[\mathbb{E}_\bullet]$. Then, in particular

$$\begin{aligned} \varphi_w : (\mathbb{T}_w, f_w, \leq_w) &\rightarrow (\mathbb{Y}, g, \leq_{\mathbb{Y}}) \\ [(v, t)] &\mapsto [(\varphi_{w,j}^{\mathbb{V}}(v), t)], \text{ for all } v \in \mathbb{V}_j^{\mathbb{X}} \\ [(e, t)] &\mapsto [(\varphi_{w,j}^{\mathbb{E}}(e), t)], \text{ for all } e \in \mathbb{E}_j^{\mathbb{X}}, \end{aligned}$$

for some order preserving maps $\varphi_{w,j}^{\mathbb{V}}$ and $\varphi_{w,j}^{\mathbb{E}}$ satisfying the consistency conditions as described in Prop. 3.4. We claim that there exists a unique morphism $\varphi : (\mathbb{X}, f, \leq_{\mathbb{X}}, \mu) \rightarrow (\mathbb{Y}, g, \leq_{\mathbb{Y}}, \mu_{\mathbb{Y}})$ in $\mathbf{Reeb}^{ord}[\mathbb{E}]$ such that each of the diagrams

$$\begin{array}{ccc} (\mathbb{T}_w, f_w, \leq_w, \mu) & \xrightarrow{q_w} & (\mathbb{X}, f, \leq_{\mathbb{X}}, \mu) \\ & \searrow \varphi_w & \downarrow \varphi \\ & & (\mathbb{Y}, g, \leq_{\mathbb{Y}}, \mu_{\mathbb{Y}}) \end{array}$$

commutes. Since the category $\mathbf{Reeb}^{ord}[\mathbb{E}_{\bullet}]$ is thin, all diagrams commute by default and if φ exists then it is unique. Therefore, we only need to show that there exists a morphism

$$\begin{aligned} \varphi : (\mathbb{X}, f, \leq_{\mathbb{X}}, \mu) &\rightarrow (\mathbb{Y}, g, \leq_{\mathbb{Y}}, \mu_{\mathbb{Y}}) \\ [(v, t)] &\mapsto [(\varphi_j^{\mathbb{V}}(v), t)], \text{ for all } v \in \mathbb{V}_j^{\mathbb{X}} \\ [(e, t)] &\mapsto [(\varphi_j^{\mathbb{E}}(e), t)], \text{ for all } e \in \mathbb{E}_j^{\mathbb{X}}, \end{aligned}$$

for some order preserving maps $\varphi_j^{\mathbb{V}}$ and $\varphi_j^{\mathbb{E}}$ satisfying the consistency conditions as described in Prop. 3.4.

Lets return to φ_w . Each map $\varphi_{w,j}^{\mathbb{E}}$ is trivially order preserving; the edge posets are just edge sets, and therefore they have the trivial partial order given only by the identities. Since φ_w is \mathbb{E} -edge labelling preserving, we get $\varphi_{w,j}^{\mathbb{E}}(e) = (\mu^{\mathbb{V}})^{-1}\mu(e)$, for all $e \in \mathbb{E}_j^{\mathbb{T}w}$. Thus $\varphi_{w,j}^{\mathbb{E}}$ is independent of w . So, naturally, we define $\varphi_j^{\mathbb{E}}$ by

$$\varphi_j^{\mathbb{E}}(e) = \varphi_{w,j}^{\mathbb{E}}(e) = (\mu^{\mathbb{V}})^{-1}\mu(e), \text{ for all } e \in \mathbb{E}_j^{\mathbb{X}} = \mathbb{E}_j^{\mathbb{T}w}.$$

Now lets focus on the order preserving maps $\varphi_{w,j}^{\mathbb{V}} : (\mathbb{V}_j^{\mathbb{T}w}, \leq_{\mathbb{V}_j^{\mathbb{T}w}}) \rightarrow (\mathbb{V}_j^{\mathbb{Y}}, \leq_{\mathbb{V}_j^{\mathbb{Y}}})$, for $j = 1, \dots, k$. We compute

$$\begin{aligned} \varphi_{w,j}^{\mathbb{V}}(r_i) &= \varphi_{w,j}^{\mathbb{V}}(D_j^{\mathbb{T}w}(e_{w_i}^{(i)})) \\ &= D_j^{\mathbb{Y}}\varphi_{w,j}^{\mathbb{E}}(e_{w_i}^{(i)}) \\ &= D_j^{\mathbb{Y}}\varphi_j^{\mathbb{E}}(e_{w_i}^{(i)}), \end{aligned}$$

and

$$\begin{aligned} \varphi_{w,j}^{\mathbb{V}}(e_t^{(i)}) &= \varphi_{w,j}^{\mathbb{V}}(D_j^{\mathbb{T}w}(e_t^{(i)})) \\ &= D_j^{\mathbb{Y}}\varphi_{w,j}^{\mathbb{E}}(e_t^{(i)}) \\ &= D_j^{\mathbb{Y}}\varphi_j^{\mathbb{E}}(e_t^{(i)}), \end{aligned}$$

for all $1 \leq t \leq d_i$ with $t \neq w_i$.

On the other hand, the inequalities $r_i \leq_{\mathbb{V}_j^{\mathbb{T}w}} e_t^{(i)}$ yield the innequalities $\varphi_{w,j}^{\mathbb{V}}(r_i) \leq_{\mathbb{V}_j^{\mathbb{Y}}} \varphi_{w,j}^{\mathbb{V}}(e_t^{(i)})$. Therefore

$$D_j^{\mathbb{Y}}\varphi_j^{\mathbb{E}}(e_{w_i}^{(i)}) \leq_{\mathbb{V}_j^{\mathbb{Y}}} D_j^{\mathbb{Y}}\varphi_j^{\mathbb{E}}(e_t^{(i)}).$$

However note that since w_i was chosen in random, we can as well get the inequality

$$D_j^{\mathbb{Y}}\varphi_j^{\mathbb{E}}(e_t^{(i)}) \leq_{\mathbb{V}_j^{\mathbb{Y}}} D_j^{\mathbb{Y}}\varphi_j^{\mathbb{E}}(e_{w_i}^{(i)}),$$

by formally replacing w_i with t . Therefore we obtain the equalities

$$D_j^{\mathbb{Y}}\varphi_j^{\mathbb{E}}(e_t^{(i)}) = D_j^{\mathbb{Y}}\varphi_j^{\mathbb{E}}(e_s^{(i)}), \text{ for all } 1 \leq t, s \leq d_i, \text{ for all } i = 1, \dots, m. \quad (1)$$

Thus we see that

$$\varphi_{w,j}^{\mathbb{V}}(r_i) = \varphi_{w,j}^{\mathbb{V}}(e_t^{(i)}), \text{ for all } 1 \leq t \leq d_i \text{ with } t \neq w_i, \quad (2)$$

and, in particular by Eqn. 1, we see that this is independent of w . So, naturally, setting $\varphi_j^{\mathbb{V}}(r_i) := \varphi_{w,j}^{\mathbb{V}}(r_i)$ is well defined.

Now, for any vertex $v \in \mathbb{V}_j^{\mathbb{X}}$ which is not a reticulation-vertex, by construction of \mathbb{T}_w , we have $v \in \mathbb{V}_j^{\mathbb{T}w}$. In particular, $v \in \mathbb{V}_j^{\mathbb{T}w}$ is either the down image of some $e \in \mathbb{E}_j^{\mathbb{T}w}$ (via the map $D_j^{\mathbb{T}w}$) or the upper image of some $e \in \mathbb{E}_j^{\mathbb{T}w-1}$ (via the map $U_j^{\mathbb{T}w}$). Because of that, we can, again show that $\varphi_{w,j}^{\mathbb{V}}(v)$ is independent of the choice of w . So we define $\varphi_j^{\mathbb{V}}(v) := \varphi_{w,j}^{\mathbb{V}}(v)$. To sum up, $\varphi_j^{\mathbb{V}}(v) := \varphi_{w,j}^{\mathbb{V}}(v)$ is a well defined map, for all $v \in \mathbb{V}_j^{\mathbb{X}}$.

We claim that the maps $\varphi_j^{\mathbb{V}}$ and $\varphi_j^{\mathbb{E}}$ satisfy the consistency conditions as described in Prop. 3.4. We only need to check the consistency condition $\varphi_j^{\mathbb{V}}D_j^{\mathbb{X}} = D_j^{\mathbb{Y}}\varphi_j^{\mathbb{E}}$. The other one is true since $U_j^{\mathbb{T}w}(e) = U_j^{\mathbb{X}}(e)$ for all $e \in \mathbb{E}_j^{\mathbb{T}w-1}$. If we show that $\varphi_j^{\mathbb{V}}D_j^{\mathbb{X}} = D_j^{\mathbb{Y}}\varphi_j^{\mathbb{E}}$, then the map $\varphi : (\mathbb{X}, f) \rightarrow (\mathbb{X}, g)$ would be an actual morphism in \mathbf{Reeb} .

Indeed, fix any w . Let $e \in \mathbb{E}_j^{\mathbb{T}w}$. We have the following cases:

- **Case 1, $D_j^{\mathbb{X}}(e)$ is not a reticulation-vertex:** We compute

$$\begin{aligned}\varphi_j^{\mathbb{V}} D_j^{\mathbb{X}}(e) &= \varphi_{w,j}^{\mathbb{V}} D_j^{\mathbb{T}w}(e) \text{ (by definition of } D_j^{\mathbb{T}w} \text{ and because } \mathbb{E}_j^{\mathbb{X}} = \mathbb{E}_j^{\mathbb{T}w}\text{)} \\ &= D_j^{\mathbb{V}} \varphi_{w,j}^{\mathbb{E}}(e) \text{ (by consistency of } \varphi_{w,j}^{\mathbb{E}}\text{)} \\ &= D_j^{\mathbb{V}} \varphi_j^{\mathbb{E}}(e).\end{aligned}$$

- **Case 2, $D_j^{\mathbb{X}}(e) = r_i$ and $e = e_{w_i}^{(i)}$:** We compute

$$\begin{aligned}\varphi_j^{\mathbb{V}} D_j^{\mathbb{X}}(e_{w_i}^{(i)}) &= \varphi_j^{\mathbb{V}}(r_i) \\ &= \varphi_{w,j}^{\mathbb{V}}(r_i) \\ &= \varphi_{w,j}^{\mathbb{V}}(D_j^{\mathbb{T}w}(e_{w_i}^{(i)})) \text{ (by definition of } D_j^{\mathbb{T}w}\text{)} \\ &= D_j^{\mathbb{V}} \varphi_{w,j}^{\mathbb{E}}(e_{w_i}^{(i)}) \text{ (by consistency of } \varphi_{w,j}^{\mathbb{E}}\text{)} \\ &= D_j^{\mathbb{V}} \varphi_j^{\mathbb{E}}(e_{w_i}^{(i)}).\end{aligned}$$

- **Case 3, $D_j^{\mathbb{X}}(e) = r_i$ and $e \neq e_{w_i}^{(i)}$:** We compute

$$\begin{aligned}\varphi_j^{\mathbb{V}} D_j^{\mathbb{X}}(e) &= \varphi_j^{\mathbb{V}}(r_i) \\ &= \varphi_{w,j}^{\mathbb{V}}(r_i) \\ &= \varphi_{w,j}^{\mathbb{V}}(e) \text{ (by Eqn. (2))} \\ &= \varphi_{w,j}^{\mathbb{V}} D_j^{\mathbb{T}w}(e) \text{ (by definition of } D_j^{\mathbb{T}w}\text{)} \\ &= D_j^{\mathbb{V}} \varphi_{w,j}^{\mathbb{E}}(e) \text{ (by consistency of } \varphi_{w,j}^{\mathbb{E}}\text{)} \\ &= D_j^{\mathbb{V}} \varphi_j^{\mathbb{E}}(e).\end{aligned}$$

Hence there exists a map $\varphi : (\mathbb{X}, f) \rightarrow (\mathbb{X}, g)$ in **Reeb**. Note that since $\mathbb{X} = (\mathbb{X}, f)$ is trivially partially ordered, then φ_w is trivially order preserving. Finally, by definition, we have $\varphi_j^{\mathbb{E}} = (\mu^{\mathbb{V}})^{-1} \circ \mu$. Hence φ is \mathbb{E} -edge labelling preserving. Therefore φ is a morphism in **Reeb**^{ord}[\mathbb{E}].

Remark 4.10. By definition of \mathbb{T}_w , each of the leaves of \mathbb{T}_w that have been created after the w -cut is identified with an edge that is incident from above of a reticulation-vertex but is not equal to $e_{w_i}^{(i)}$ for all $i = 1, \dots, m$. Therefore the cardinality of the set of the leaves that are additional to the n -leaves of \mathbb{X} is exactly

$$\sum_{i=1}^m (d_i - 1),$$

which is equal to the first Betti number $B_1 = s$ of \mathbb{X} , because of Prop. 3.2. Hence, each tree \mathbb{T}_w has $(n + s)$ -leaves. \square

Example 4.11. Consider a Reeb graph $\mathbb{X} = (\mathbb{X}, f)$ with four leaves and one cycle, and edges labelled by e_1, e_2, \dots, e_8 , as shown in Fig. 2. By applying Thm. 4.7 we get two trees with 5-leaves in the tree-decomposition of \mathbb{X} . For each of the trees, the one additional leaf to the four leaves of \mathbb{X} corresponds to one of the two edges e_5 and e_6 , respectively. Thus there are exactly two such trees with 5-leaves.

Example 4.12. Consider a Reeb graph $\mathbb{X} = (\mathbb{X}, f)$ with a single reticulation-vertex—where in this particular example we also think of it as a leaf—three edges, and two cycles, and edges labelled by e_1, e_2, e_3 , as shown in Fig. 3. By applying Thm. 4.7 we get three trees with 4-leaves in the tree-decomposition of \mathbb{X} . Note that the trees are in one to one correspondence with all possibilities of isolating an edge (which is identified with the reticulation-vertex) from the other edges, given that they are incident from above some reticulation vertex. Thus there are exactly three such trees with 4-leaves.

5 Isomorphism complexity of Reeb graphs

Thm. 4.7 can help improve our understanding of the isomorphism complexity of Reeb graphs under some fixed parameter. Here we will refer to *isomorphism complexity* simply by *complexity*.

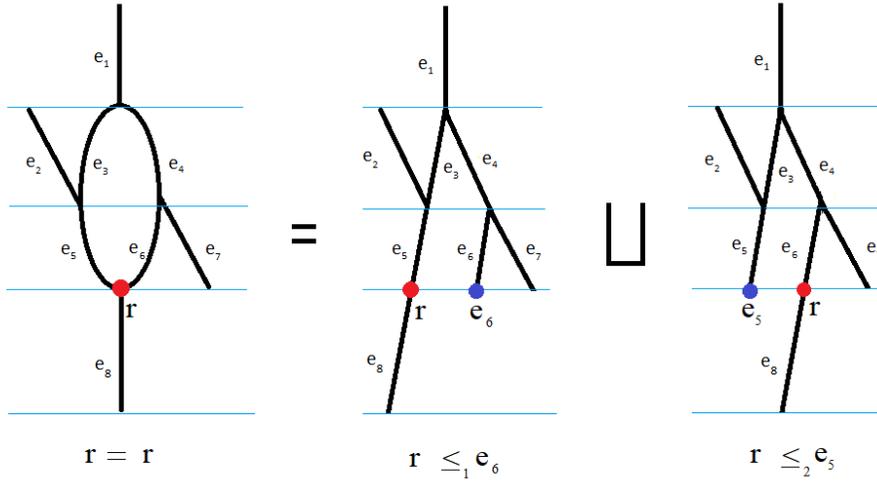


Figure 2: An example describing the tree-decomposition for $n = 4$ and $s = 1$.

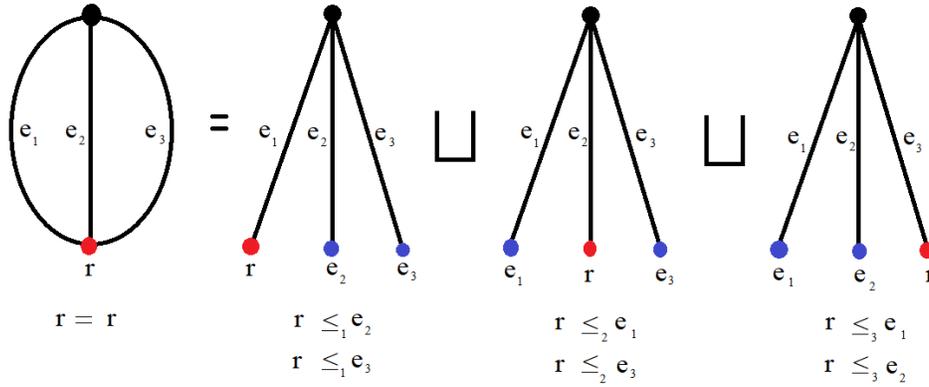


Figure 3: An example describing the tree-decomposition for $n = 1$ and $s = 2$.

5.1 Upper bound on the complexity

Let $\mathbb{X} = (\mathbb{X}, f)$ be a Reeb graph with edge-sequence \mathbb{E} . Let $\mathbb{Y} = (\mathbb{Y}, g)$ any Reeb graph. Checking whether \mathbb{Y} has same edge-structure as \mathbb{X} is equivalent to checking whether \mathbb{Y} has equivalent edge-sequence with \mathbb{E} . This takes time $O(|\mathbb{E}|)$, where $|\mathbb{E}|$ is the cardinality of \mathbb{E} . Now, by Euler's characteristic formula we have $s = |\mathbb{E}| - |\mathbb{V}| + 1$, and so $|\mathbb{E}| = |\mathbb{V}| + s - 1$. By definition, if two Reeb graphs $(\mathbb{X}, f), (\mathbb{Y}, g)$ are isomorphic, then they have the same edge structure. Therefore:

$$\left(\text{Complexity of Reeb graphs} \right) = \left(\text{Complexity of Reeb graphs of same edge structure} \right) + O(|\mathbb{V}| + s).$$

By Thm. 4.7, two Reeb graphs of same edge structure, with n leaves and Betti number $B_1(\mathbb{X}) = s$, are isomorphic if and only if they have the same tree-decomposition. To check for isomorphism it takes at worst, as many number of steps as the cardinality of the set $\mathcal{T}(\mathbb{X})$ times the complexity of checking whether two ordered trees with $(n+s)$ -leaves are isomorphic in $\mathbf{Reeb}^{ord}[\mathbb{E}_\bullet]$, where \mathbb{E}_\bullet is the edge structure of \mathbb{X} . Namely:

$$\left(\text{Complexity of Reeb graphs of same edge structure} \right) = \left(\text{Complexity of ordered trees} \right) \cdot O(|\mathcal{T}(\mathbb{X})|).$$

Now checking if two trees with $(n+s)$ -leaves are isomorphic takes time $O(n+s)$.⁵ By construction, the cardinality of $\mathcal{T}(\mathbb{X})$ of a Reeb graph \mathbb{X} is equal to the product of indegrees of the reticulation-vertices, i.e.

$$|\mathcal{T}(\mathbb{X})| = \prod_{i=1}^m d_i$$

Therefore we get

$$\left(\text{Complexity of Reeb graphs} \right) = O((n+s) \prod_{i=1}^m d_i) + O(|\mathbb{V}| + s).$$

Now because $L(\mathbb{X}) \subset \mathbb{V}$ and $|L(\mathbb{X})| = n$, we get $n \leq |\mathbb{V}|$. Also since $s \geq 0$ ($s = 0$ if \mathbb{X} is a tree) and $|\mathbb{V}| \geq 2$, we have the obvious bounds $n+s \leq |\mathbb{V}| + s \leq |\mathbb{V}|(1+s)$. Thus:

$$\left(\text{Complexity of Reeb graphs} \right) = O(|\mathbb{V}|(1+s) \prod_{i=1}^m d_i) + O(|\mathbb{V}|(1+s)) = O(|\mathbb{V}|(1+s) \prod_{i=1}^m d_i).$$

5.2 Parametrized complexity

Although this is a good bound, we would like to consider an upper bound on the isomorphism complexity that does not depend on the indegrees of reticulation-vertices, but only depends on the number of vertices $|\mathbb{V}|$ and the Betti number $B_1(\mathbb{X}) = s$. Since each of the indegrees of the reticulation nodes is $d_i > 1$ we have the bound $d_i \leq 2^{d_i-1}$. Taking the product over all these inequalities and since

$$\sum_{i=1}^m (d_i - 1) = s,$$

we get

$$\left(\text{Complexity of Reeb graphs with } B_1(\mathbb{X}) = s \right) = O(|\mathbb{V}|(1+s)2^s).$$

Hence, the Reeb graph isomorphism problem is fixed parameter tractable when the parameter is the first Betti number of the graph. Finally, note that the bound $d_i \leq 2^{d_i-1}$ is tight: indeed, if we consider Reeb graphs with n -leaves and $B_1(\mathbb{X}) = s$, such that the indegree of each of its reticulation-vertices is 2, then the product of all the indegrees is exactly 2^s .

6 Phylogenetic networks viewed as ordered Reeb graphs

In computational phylogenetics a *phylogenetic network* is any rooted DAG, G , that can be used to represent the evolutionary relationships among biological organisms, e.g. genes, often called *taxa*. These taxa are represented by an ordering on the leaves of the rooted DAG.

Remark 6.1. Quite often in practice, the taxa are represented by the totally ordered set $\{1 < 2 < \dots < n\}$. However, this is very restrictive for general phylogenetic networks. Namely, one can have a phylogenetic network where some of its leaves are ordered in many different ways, and some of them might not be labelled at all. So it is better to consider a partial order on the leaves.

Naturally one considers a pair of phylogenetic networks to be *isomorphic* if there is an isomorphism between their underlying DAGs that also preserves the *order* of their corresponding taxa.

⁵It is known from basic graph theory that checking if two trees with k -leaves are isomorphic takes $O(k)$ steps

6.1 Time consistent phylogenetic networks

In many applications, phylogenetic networks are equipped with a *time assignment* on the nodes or vertices, which we can think of as a height function $f : G \rightarrow \mathbb{R}$ on the graph G . The time assignment map f is often required to be *time consistent* which essentially means that the time stamp of any ‘non-root node’ should be strictly larger than the time stamp of its ‘parent node’. See [25], pg 167. This is exactly what Reeb graphs satisfy as we emphasized in Ex 3.8. Moreover, if we restrict to ordered Reeb graphs, then these structures can be thought of as vertex weighted DAGs (see Sec. 3.2) such that their edge sets and vertex sets are partially ordered in a compative way. So we believe that ordered rooted Reeb graphs would serve as a natural mathematical model for time consistent phylogenetic networks. By thinking of a time consistent phylogenetic network as an ordered Reeb graph, we can also consider its tree decomposition, given by Thm. 4.7. Since the phylogenetic network is rooted and has its leaves partially ordered, by construction, each of the trees \mathbb{T}_w , yielded by the network’s tree decomposition, would have their leaves partially ordered. Therefore, each tree \mathbb{T}_w is a phylogenetic tree.

6.2 Hausdorff distance

Among many distance metrics, we can utilize the ℓ^p -cophenetic metrics, d_p , for $1 \leq p \leq \infty$, for comparing a pair of phylogenetic trees [9, 29]. Lets focus on d_∞ (for $p = \infty$). The idea is that, via the cophenetic map [34], any phylogenetic tree with n ordered leaves, $\{1 < \dots < n\}$, can be identified with a single point in the $[n(n+1)/2]$ -dimensional Euclidean space. The (i, j) -th coordinate of the point, where $i \leq j$, corresponds to the time-stamp of the *least common ancestor* of the leaves i, j in the phylogenetic tree [29]. Then, we simply consider the ℓ^∞ -norm for comparing a pair of phylogenetic trees with n ordered leaves. By the tree decomposition of Thm. 4.7, we can thus identify any time consistent phylogenetic network with n ordered leaves, and first Betti number s , as a finite subset of the the $[(n+s)(n+s+1)/2]$ -dimensional Euclidean space. Hence, we can utilize the Hausdorff distance for comparing a pair of phylogenetic networks.

Definition 6.2. Let (X, d_X) be any metric space, and let A, B be any subsets of X . The *Hausdorff distance* of A, B is

$$d_H(A, B) = \max\{\max_{a \in A} \min_{b \in B} d_X(a, b), \max_{b \in B} \min_{a \in A} d_X(a, b)\}.$$

Example 6.3. Consider the pair N_1, N_2 of time consistent phylogenetic networks as in Fig. 4, modelled as ordered Reeb graphs, and having the non-trivial partial order on their leaves, $1 < 2$.

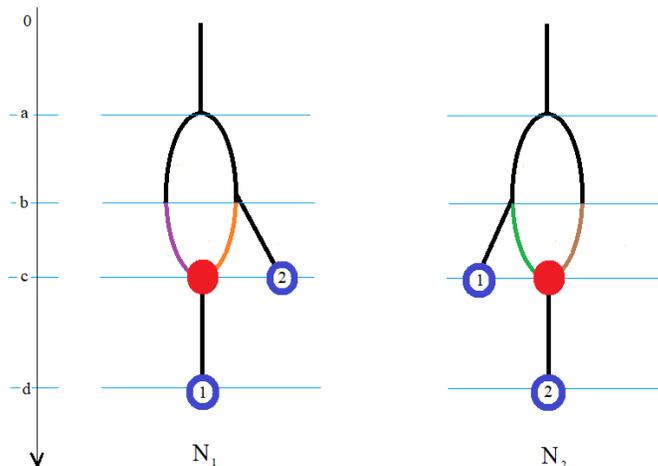


Figure 4: A pair of time consistent phylogenetic networks, each having 2 leaves and 1 cycle.

Also assume $0 < a < b < c < d$ and $c - b < b - a < d - c$ as in Fig. 4. It is easy to check that, if we forget the non-trivial order on the leaves of N_1, N_2 , then these networks are isomorphic as ordinary Reeb graphs. However they are not order-preserving isomorphic. We do this by showing that the Hausdorff distance—with respect to d_∞ —of the

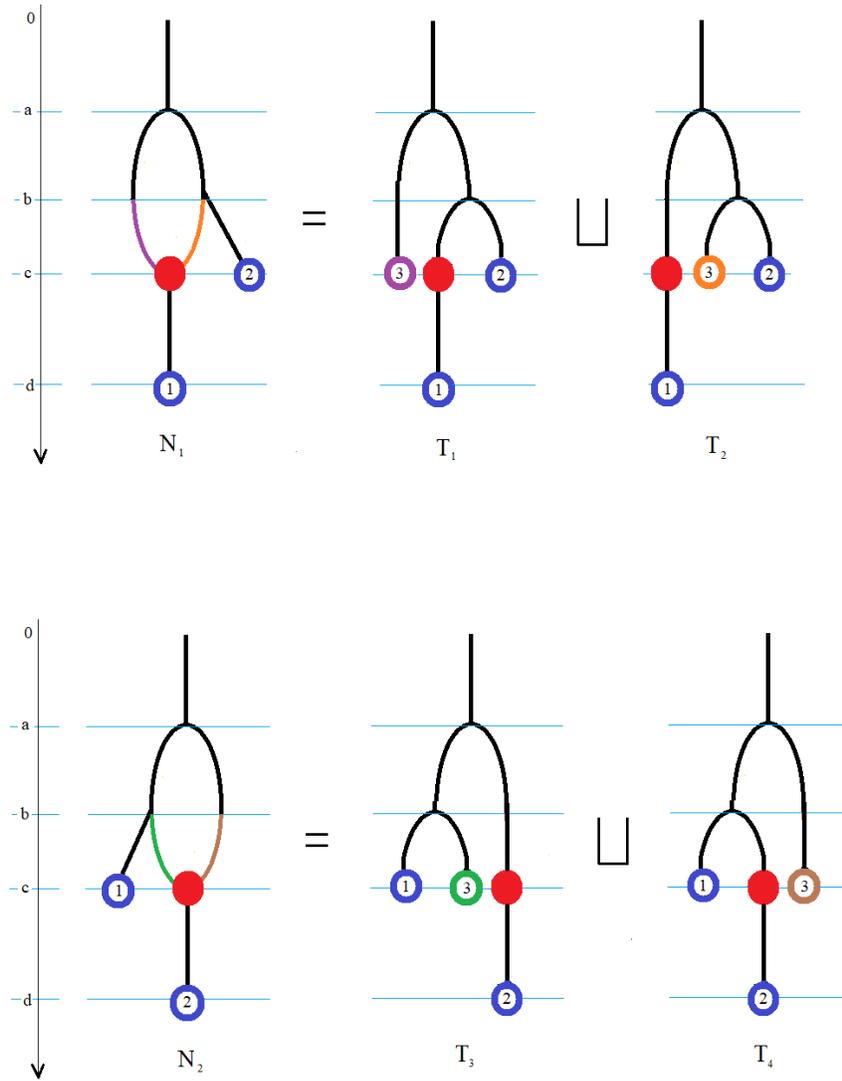


Figure 5: The additional leaf in each tree corresponds to an edge incident from above the reticulation vertex.

tree decompositions of N_1, N_2 is non-zero. Let $\mathcal{T}(N_1) = \{T_1, T_2\}$ and $\mathcal{T}(N_2) = \{T_3, T_4\}$ be the tree decompositions

of N_1 and N_2 as in Fig. 5. We have

$$\begin{aligned}
d_\infty(T_1, T_3) &= \left\| \begin{pmatrix} d & b & a \\ \cdot & c & a \\ \cdot & \cdot & c \end{pmatrix} - \begin{pmatrix} c & a & b \\ \cdot & d & a \\ \cdot & \cdot & c \end{pmatrix} \right\|_\infty = d - c, \\
d_\infty(T_1, T_4) &= \left\| \begin{pmatrix} d & b & a \\ \cdot & c & a \\ \cdot & \cdot & c \end{pmatrix} - \begin{pmatrix} c & b & a \\ \cdot & d & a \\ \cdot & \cdot & c \end{pmatrix} \right\|_\infty = d - c, \\
d_\infty(T_2, T_3) &= \left\| \begin{pmatrix} d & a & a \\ \cdot & c & b \\ \cdot & \cdot & c \end{pmatrix} - \begin{pmatrix} c & a & b \\ \cdot & d & a \\ \cdot & \cdot & c \end{pmatrix} \right\|_\infty = d - c, \\
d_\infty(T_2, T_4) &= \left\| \begin{pmatrix} d & a & a \\ \cdot & c & b \\ \cdot & \cdot & c \end{pmatrix} - \begin{pmatrix} c & b & a \\ \cdot & d & a \\ \cdot & \cdot & c \end{pmatrix} \right\|_\infty = d - c,
\end{aligned}$$

since $c - b < b - a < d - c$. Hence, we obtain

$$\begin{aligned}
d_H(\mathcal{T}(N_1), \mathcal{T}(N_2)) &= \max \left\{ \max_{1 \leq i \leq 2} \left\{ \min_{3 \leq j \leq 4} d_\infty(T_i, T_j) \right\}, \max_{3 \leq j \leq 4} \left\{ \min_{1 \leq i \leq 2} d_\infty(T_i, T_j) \right\} \right\} \\
&= d - c > 0.
\end{aligned}$$

Remark 6.4. Assume A, B are finite. By definition, the Hausdorff distance between A, B can be computed in at most $|A| \cdot |B|$ -many steps. That means, in particular, that computing the Hausdorff distance of a pair of phylogenetic networks with n leaves and first Betti number s , takes $O(|\mathbb{V}|^2(1+s)^{24s})$ time. Namely, computing the Hausdorff distance is fixed parameter tractable when the parameter is the first Betti number.

7 Concluding remarks

In this manuscript we showed that Reeb graphs are classified up to isomorphism by their tree-decomposition, we constructed upper bounds for their isomorphism complexity, and showed that the isomorphism problem for Reeb graphs is fixed parameter tractable when the parameter is the first Betti number. We proposed **Reeb**^{ord} as model for phylogenetic networks. Moreover, we proposed the use of Hausdorff distance as a metric for phylogenetic networks with n leaves and first Betti number s . We speculate that our results, on one hand would further our understanding of both the structure and isomorphism complexity of Reeb graphs, and on the other hand they will enhance the existing methods on phylogenetic networks by providing new insights on how to do statistics and data analysis on these structures.

Future work: It is in the author's interests to apply properly the notion of Hausdorff distance to define a metric on time consistent phylogenetic networks. In order to do that some restrictions on the ordered Reeb graphs (phylogenetic networks) may need to be considered in addition to requiring a fixed number of leaves and cycles, e.g. we may need to assume in particular a total order on the leaves and reticulation vertices. Also a natural question to ask is whether there exist other tree decompositions for general ordered Reeb graphs, perhaps with fewer tree factors, and how do they look like.

Furthermore, just like Reeb graphs can be identified with nice enough cosheaves on \mathbb{R} valued in the category **Set** of all sets [16], it seems that ordered Reeb graphs can be identified with constructible cosheaves $\mathbf{F} : \mathbf{Int} \rightarrow \mathbf{Pos}$ on \mathbb{R} valued in the category **Pos** of all posets. If this speculation is true then one can define a notion of *interleaving distance* for ordered Reeb graphs—in the sense of Bubenik et al. [8]—and thus phylogenetic networks in particular. The interleaving distance will thus be a metric on the collection of arbitrary phylogenetic networks, i.e. not just the ones with fixed number of leaves and first Betti number. This would give an advantage for one to use the interleaving metric over the Hausdorff distance as a metric for phylogenetic networks. By Rem. 6.4, computing the Hausdorff distance of a pair of phylogenetic networks is fixed parameter tractable when the parameter is the first Betti number, but we do not know whether this is the case also for the interleaving distance. In the near future the author wishes to work towards the sheaf-theoretic aspects of ordered Reeb graphs and their implications to phylogenetics, e.g. defining an interleaving metric for the comparison of arbitrary time consistent phylogenetic networks.

8 Conflict of Interest Statement

The author states that there is no conflict of interest.

References

- [1] JIRI Adámek. Herrlich and h., strecker, ge, abstract and concrete categories. the joy of cats. *Pure and Applied Mathematics, A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, xiv*, 1990.
- [2] Pankaj K Agarwal, Herbert Edelsbrunner, John Harer, and Yusu Wang. Extreme elevation on a 2-manifold. *Discrete & Computational Geometry*, 36(4):553–572, 2006.
- [3] Ulrich Bauer, Elizabeth Munch, and Yusu Wang. Strong Equivalence of the Interleaving and Functional Distortion Metrics for Reeb Graphs. In Lars Arge and János Pach, editors, *31st International Symposium on Computational Geometry (SoCG 2015)*, volume 34 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 461–475, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- [4] Silvia Biasotti, Daniela Giorgi, Michela Spagnuolo, and Bianca Falcidieno. Reeb graphs for shape analysis and applications. *Theoretical computer science*, 392(1-3):5–22, 2008.
- [5] Louis J Billera, Susan P Holmes, and Karen Vogtmann. Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics*, 27(4):733–767, 2001.
- [6] Håvard Bakke Bjerkevik and Magnus Bakke Botnan. Computational Complexity of the Interleaving Distance. In Bettina Speckmann and Csaba D. Tóth, editors, *34th International Symposium on Computational Geometry (SoCG 2018)*, volume 99 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 13:1–13:15, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- [7] Adam Bouland, Anuj Dawar, and Eryk Kopczyński. On tractable parameterizations of graph isomorphism. In *International Symposium on Parameterized and Exact Computation*, pages 218–230. Springer, 2012.
- [8] Peter Bubenik, Vin De Silva, and Jonathan Scott. Metrics for generalized persistence modules. *Foundations of Computational Mathematics*, 15(6):1501–1531, 2015.
- [9] Gabriel Cardona, Arnau Mir, Francesc Rosselló, Lucia Rotger, and David Sánchez. Cophenetic metrics for phylogenetic trees, after Sokal and Rohlf. *BMC bioinformatics*, 14(1):3, 2013.
- [10] Gabriel Cardona, Francesc Rosselló, and Gabriel Valiente. Extended newick: it is time for a standard representation of phylogenetic networks. *BMC bioinformatics*, 9(1):532, 2008.
- [11] Gunnar Carlsson, Afra Zomorodian, Anne Collins, and Leonidas J Guibas. Persistence barcodes for shapes. *International Journal of Shape Modeling*, 11(02):149–187, 2005.
- [12] Frédéric Chazal and Jian Sun. Gromov-hausdorff approximation of metric spaces with linear structure. *arXiv preprint arXiv:1305.1172*, 2013.
- [13] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Extending persistence using poincaré and lefschetz duality. *Foundations of Computational Mathematics*, 9(1):79–103, 2009.
- [14] Kree Cole-McLaughlin, Herbert Edelsbrunner, John Harer, Vijay Natarajan, and Valerio Pascucci. Loops in reeb graphs of 2-manifolds. *Discrete & Computational Geometry*, 32(2):231–244, 2004.
- [15] William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *Journal of Algebra and its Applications*, 14(05):1550066, 2015.
- [16] Vin De Silva, Elizabeth Munch, and Amit Patel. Categorized reeb graphs. *Discrete & Computational Geometry*, 55(4):854–906, 2016.
- [17] Tamal K Dey, Fengtao Fan, and Yusu Wang. An efficient computation of handle and tunnel loops via reeb graphs. *ACM Transactions on Graphics (TOG)*, 32(4):32, 2013.
- [18] Barbara Di Fabio and Claudia Landi. The edit distance for reeb graphs of surfaces. *Discrete & Computational Geometry*, 55(2):423–461, 2016.
- [19] Andreas Dress. The category of x-nets. In *Networks: from biology to theory*, pages 3–22. Springer, 2007.
- [20] Herbert Edelsbrunner, John Harer, and Amit K Patel. Reeb spaces of piecewise linear mappings. In *Symposium on Computational Geometry*, pages 242–250, 2008.

- [21] Francisco Escolano, Edwin R Hancock, and Silvia Biasotti. Complexity fusion for indexing reeb digraphs. In *International Conference on Computer Analysis of Images and Patterns*, pages 120–127. Springer, 2013.
- [22] Xiaoyin Ge, Issam I Safa, Mikhail Belkin, and Yusu Wang. Data skeletonization via reeb graphs. In *Advances in Neural Information Processing Systems*, pages 837–845, 2011.
- [23] William Harvey, Yusu Wang, and Rephael Wenger. A randomized $o(m \log m)$ time algorithm for computing reeb graphs of arbitrary simplicial complexes. In *Proceedings of the twenty-sixth annual symposium on Computational geometry*, pages 267–276. ACM, 2010.
- [24] Masaki Hilaga, Yoshihisa Shinagawa, Taku Kohmura, and Toshiyasu L Kunii. Topology matching for fully automatic similarity estimation of 3d shapes. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 203–212. ACM, 2001.
- [25] Daniel H Huson, Regula Rupp, and Celine Scornavacca. *Phylogenetic networks: concepts, algorithms and applications*. Cambridge University Press, 2010.
- [26] Daniel Lokshtanov, Marcin Pilipczuk, Michał Pilipczuk, and Saket Saurabh. Fixed-parameter tractable canonization and isomorphism test for graphs of bounded treewidth. *SIAM Journal on Computing*, 46(1):161–189, 2017.
- [27] Saunders Mac Lane. *Categories for the working mathematician*, volume 5. Springer Science & Business Media, 2013.
- [28] Dmitriy Morozov, Kenes Beketayev, and Gunther Weber. Interleaving distance between merge trees. *Discrete and Computational Geometry*, 49(22-45):52, 2013.
- [29] Elizabeth Munch and Anastasios Stefanou. The ℓ^∞ -cophenetic metric for phylogenetic trees as an interleaving distance. *arXiv preprint arXiv:1803.07609*, 2018.
- [30] Monica Nicolau, Arnold J Levine, and Gunnar Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences*, 108(17):7265–7270, 2011.
- [31] Georges Reeb. Sur les points singuliers d’une forme de pfaff completement integrable ou d’une fonction numerique [on the singular points of a completely integrable pfaff form or of a numerical function]. *Comptes Rendus Acad. Sciences Paris*, 222:847–849, 1946.
- [32] Charles Semple, Mike A Steel, Richard A Caplan, Mike Steel, et al. *Phylogenetics*, volume 24. Oxford University Press on Demand, 2003.
- [33] Gurjeet Singh, Facundo Mémoli, and Gunnar E Carlsson. Topological methods for the analysis of high dimensional data sets and 3d object recognition. In *SPBG*, pages 91–100, 2007.
- [34] Robert R Sokal and F James Rohlf. The comparison of dendrograms by objective methods. *Taxon*, pages 33–40, 1962.
- [35] Zoë Wood, Hugues Hoppe, Mathieu Desbrun, and Peter Schröder. Removing excess topology from isosurfaces. *ACM Transactions on Graphics (TOG)*, 23(2):190–208, 2004.
- [36] Koichi Yamazaki, Hans L Bodlaender, Babette De Fluiter, and Dimitrios M Thilikos. Isomorphism for graphs of bounded distance width. In *Italian Conference on Algorithms and Complexity*, pages 276–287. Springer, 1997.
- [37] Yuan Yao, Jian Sun, Xuhui Huang, Gregory R Bowman, Gurjeet Singh, Michael Lesnick, Leonidas J Guibas, Vijay S Pande, and Gunnar Carlsson. Topological methods for exploring low-density states in biomolecular folding pathways. *The Journal of chemical physics*, 130(14):04B614, 2009.