A Proofs

Lemma 21 *If $R \subset S_1 \times S_2$ is a non-empty bisimulation between graphs $\mathcal{S}_1 = (S_1, E_1, g_2, b_1)$ and $\mathcal{S}_2 = (S_2, E_2, g_2, b_2)$, then R is total.*

Proof. There is $(s, s') \in R$ by non-triviality. Suppose $s_1 \in S_1$. Since $\mathcal{S}_1$ is connected, there is a path $(t_0, n_0, \ldots, n_{k-1}, t_k)$ with $t_0 = s$ and $t_k = s_1$. We will construct a sequence of vertices $t'_0, t'_1, \ldots, t'_k$ in S_2 and show that for all $j \leq k$ we have that $(t_j, t'_j) \in R$. As a consequence, it will follow that $(t_k, t'_k) = (s_1, t'_k) \in R$, so we have satisfied one direction of totality. The other direction follows symmetrically with 3(2)(Zag). First, let $t'_0 = s'$. Clearly $(t_0, t'_0) \in R$. By induction assume that we have found t'_j for some $j < k$ such that $(t_j, t'_j) \in R$. Let $\lambda_j = g(t_j, t_{j+1})$. By 3(1)(Zig), there exists some $t'_{j+1} \in S_2$ such that $t'_j \xrightarrow{\lambda_j} t'_{j+1}$ and $(t_{j+1}, t'_{j+1}) \in R$ as required.

Lemma 22 *The unraveling is a tree, infinite if the graph has a proper cycle.*

Proof. Let $\mathcal{S} = (S, E, g, b)$ be a graph and $T_{s_0}(\mathcal{S})$ its unraveling at $s_0 \in S$. First, let us verify that $T_{s_0}(\mathcal{S})$ is a graph according to Definition 1. Conditions 1(1) and 1(2) follow directly from the definition. Note that there are no loops in $T_{s_0}(\mathcal{S})$ because there is an edge between vertices in $T_{s_0}(\mathcal{S})$ only if they represent paths of different lengths. Condition 1(3) is also obvious from the above definition. Finally connectedness (condition 1(4)) follows from symmetry and because any path can be traced back to the root vertex (s_0).

$T_{s_0}(\mathcal{S})$ is infinite if there exists a proper cycle because then there are proper paths of any finite length from s_0. It remains to show that $T_{s_0}(\mathcal{S})$ has no proper cycles. Suppose $p = (\bar{s}_0, n_0, \ldots, n_{k-1}, \bar{s}_k)$ is a proper path in $T_{s_0}(\mathcal{S})$. We will show that it is not a cycle, i.e. $\bar{s}_0 \neq \bar{s}_k$. From the definition of unraveling it is clear that every vertex $\bar{s} \in S$ has at most one $\bar{E}$-neighbor which represents a shorter path than $\bar{s}$, so since p is proper, there is some $j \leq k$ such that

$$\ell(\bar{s}_0) \geq \cdots \geq \ell(\bar{s}_j) \leq \ell(\bar{s}_{j+1}) \leq \cdots \leq \ell(\bar{s}_n).$$

If $j = 0$ or $j = n$, then it is clear that the path cannot be a cycle (unless $n = 0$, but then the path is trivial and hence not proper). Otherwise, $\bar{s}_0$ and $\bar{s}_k$ are end extentions of $\bar{s}_j$. If $\bar{s}_k = \bar{s}_0$ then there must be backtracking in p (the formal proof of this is left to the reader). Thus, $\bar{s}_0 \neq \bar{s}_k$ as required.

A.1 Correctness of the graph learning algorithm

As mentioned in Section 5, we can view the algorithm as creating an increasing sequence of finite trees $T_0 \subset T_1 \subset T_2$ and equivalence relations on those trees $E_0 \subset E_1 \subset E_2$. Once we identify some points according to one of the equivalence relations, it forces potentially an infinite part of the tree to wrap around a cycle or collection of cycles (when the first Betti number of the quotient is increased). Formally, we will define this as a type of closure and show that the closure of a

sufficiently large finite part is the whole tree with the appropriate equivalence relation to give an isomorphic copy of the original graph.

Let $\mathcal{T}$ be a tree. Let $\mathfrak{T}$ be the set of pairs (T^*, E^*) such that T^* is a subset of $\mathcal{T}$ and E^* is an equivalence relation on T^*. Technically, T^*/E^* is what is called a *hypothesis graph* in the algorithm. We say that E^* is p-coherent if each E^*-equivalence class is a subset of $p^{-1}(s)$ for some s. We say that a pair (T_1^*, E_1^*) is an *expansion* of (T_0^*, E_0^*), if $T_0^* \subset T_1^*$ and $E_0^* \subset E_1^*$ and we denote it by $(T_0^*, E_0^*) \subset (T_1^*, E_1^*)$. We will now define a closure operator for such pairs. Let $(T_0^*, E_0^*) \in \mathfrak{T}$. Let $(T_0, E_0) = (T_0^*, E_0^*)$. Suppose (T_n, E_n) has been defined. Let $(t_0, \lambda_0, t_0'), (t_1, \lambda_1, t_1'), \dots$ be the (possibly infinite) list of all elements (t_i, λ_i, t_i') of $T_n \times \Lambda \times (\mathcal{T} \setminus T_n)$ such that for each i there are $u_i, u_i' \in T_n$ with $u_i \xrightarrow{\lambda} u_i'$ and $(u_i, t_i) \in E_1^*$. Then define $T_{n+1} = T_n \cup \{t_1', t_2', \dots\}$ and $E_{n+1} = \langle E_n \cup \{(u_i', t_i')\}\rangle$ where $\langle A\rangle$ is the smallest equivalence relation containing a given set A of pairs. Then the closure of (T_0^*, E_0^*) is the pair

$$\left(\bigcup_{n\in\mathbb{N}} T_n, \bigcup_{n\in\mathbb{N}} E_n\right).$$

Suppose $\mathcal{S}$ is a graph and $(p, q)\colon \mathcal{T} \to \mathcal{S}$ a covering map. Assuming that $\mathcal{T}$ and $\mathcal{S}$ are deterministic, we can neglect the component q from the covering map and consider only $p\colon \mathcal{T} \to \mathcal{S}$ which is onto, preserves ports and edge labels, and is a local isomorphism. The *pumping number* of $\mathcal{S}$ is the smallest $n \in \mathbb{N}$ such that all paths in $\mathcal{S}$ of length at least n must contain a cycle. The name comes from the well-known Pumping Lemma of finite automata theory. Denote by r the root of T.

Lemma 23 *Suppose $(T^*, E^*) \in \mathfrak{T}$ has the following properties*

(1) *T^* is finite*
(2) *T^* contains all elements of T of length at most $2N$ of $\mathcal{S}$ where N is the pumping number of $\mathcal{S}$*
(3) *$\{t \in T^* \mid (r, t) \in E^*\} = p^{-1}(s_0) \cap T^*$ where s_0 is the pebbled vertex in $\mathcal{S}$ or the initial vertex*
(4) *For all t_0, t_0', t_1, t_1', if $(t_0, t_0') \in E^*$, $t_0 \xrightarrow{\lambda} t_1$ and $t_1 \xrightarrow{\lambda} t_1'$, then $(t_1, t_1') \in E^*$.*
(5) *If (t, t') is not in E^* as a consequence of conditions (3) and (4), then it is not in E^*.*

Then the closure of (T^*, E^*) equals (T, E_p) where $E_p = \{(t, t') \mid p(t) = p(t')\}$ (same as B_R of Theorem 20).

Proof. Denote by $T[l]$ the restriction of $\mathcal{T}$ to sequences of length l. Recall the construction of the closure. We claim that (a) $T[2N+n] \subset T_n$, (b) $E_n \subset E_p \restriction T_n$ and (c) $E_p \restriction T[N+n] \subset E_n$. Let us prove claims (a)–(c) by induction on n. If $n = 0$, then by definition $T_0 = T^*$ and by condition (2), claim (a) follows. For (b), we need to show that the equivalence classes of E_0 are contained in the equivalence classes of E_p intersected with T^*. For the equivalence class containing the root,

it follows from condition (3). Suppose $(t_1, t_1') \in E_0$ is a consequence of condition (4). Then there are some $(t_0, t_0') \in E_0$ with $t_0 \xrightarrow{\lambda} t_1$ and $t_0' \xrightarrow{\lambda} t_1'$. If t_0 and t_0' are E_0-equivalent to the root, then the claim follows from the properties of p and otherwise by induction on the number of times condition (4) has been applied and properties of p. For (c), suppose $p(t) = p(t')$ and $t, t' \in T[N] = T[N-0]$. There are sequences of length at most N starting from both t and t' and following the same labels which ends up in two elements t_*, t_*' that are both equivalent to the root (find the pebble in $\mathcal{S}$ in less than N moves). Then by applying condition (4) backwards, yields $(t, t') \in E_n$. Suppose we have proved (a)–(c) for all $n \leq k$ and suppose $n = k+1$. Suppose t' is an element of T of length $k+1$. Then its predecessor t is of length k, so it is in T_k by the induction hypothesis. Let λ be the label such that $t \xrightarrow{\lambda} t'$. We claim that (t, λ, t') is in the list of triples which define T_{k+1} from T_k. For this we need to find elements u, u' as in the definition of that list. Let $\bar{\tau} = (\tau_0, \ldots, \tau_k, \tau_{k+1})$ be the list of all the elements of the branch leading up to t'. In particular we have $\tau_0 = r$, $\tau_k = t$, and $\tau_{k+1} = t'$. Since $k \geq 2N$, there is a cycle in the path $p(\tau_0), p(\tau_1), \ldots, p(\tau_{k+1})$ *before* the index $n - N - 1 = k + 1 - N - 1 \geq 2N - N = N$. So there are $j < j'$ such that $p(\tau_j) = p(\tau_{j'})$ and by condition (c) in the induction hypothesis, we have $(\tau_j, \tau_{j'}) \in E_n$. Suppose $\lambda_0, \ldots, \lambda_k$ are the transition labels such that $\tau_i \xrightarrow{\lambda_i} \tau_{i+1}$. Consider removing the cycle from the path and taking the inverse image of the resulting shorter path via p. The resulting path in T is shorter than $k+1$ so it is in T_n. It is not hard to see that if we let u, u' to be the last two elements of that path, they are as desired. This proves the induction step for (a). Induction steps for (b) and (c) are the same as the base case. In particular $T = \bigcup_{n \in \mathbb{N}} T_n$.

Now let $E = \bigcup E_n$. We want to show that $E = E_p$. By (b),

$$E = \bigcup_{n \in \mathbb{N}} E_n \subset \bigcup_{n \in \mathbb{N}} E_p \restriction T_n = E_p \restriction \bigcup_{n \in \mathbb{N}} T_n = E_p \restriction T = E_p.$$

Clearly $T = \bigcup_{n \in \mathbb{N}} T[N+n]$. By (c),

$$E_p = E_p \restriction T = E_p \restriction \bigcup_{n \in \mathbb{N}} T[N+n] = \bigcup_{n \in \mathbb{N}} E_p \cap T[N+n] \subset \bigcup_{n \in \mathbb{N}} E_n = E.$$

Thus, $E = E_p$. □