The Ehrenfeucht-Fraïssé Game for Paths and Cycles

Jason Brown and Richard Hoshino
Department of Mathematics and Statistics
Dalhousie University
Halifax, Nova Scotia, Canada B3H 3J5

Abstract

Let L and R be two graphs. For any positive integer n, the Ehrenfeucht-Fraïssé game $G_n(L, R)$ is played as follows: on the i^{th} move, with $1 \leq i \leq n$, the first player chooses a vertex on either L or R, and the second player responds by choosing a vertex on the other graph. Let l_i be the vertex of L chosen on the i^{th} move, and let r_i be the vertex of R chosen on the i^{th} move. The second player wins the game iff the induced subgraphs $L[\{l_1, l_2, \ldots, l_n\}]$ and $R[\{r_1, r_2, \ldots, r_n\}]$ are isomorphic under the mapping sending l_i to r_i. It is known that the second player has a winning strategy if and only if the two graphs, viewed as first-order logical structures (with a binary predicate E), are indistinguishable (in the corresponding first-order theory) by sentences of quantifier depth at most n. In this paper we will give the first complete description of when the second player has a winning strategy for L and R being both paths or both cycles. The results significantly improve previous partial results.

Keywords: Ehrenfeucht, Fraïssé, game, graph, logic

1 Introduction

Let $L = (X_L, \{\mathcal{L}_1, \ldots, \mathcal{L}_k\})$ and $R = (X_R, \{\mathcal{R}_1, \ldots, \mathcal{R}_k\})$ be two discrete structures of the same type (that is, for each i, the relations $\mathcal{L}_i$ and $\mathcal{R}_i$, on sets X_L and X_R respectively, have the same arity). For any integer n, the Ehrenfeucht-Fraïssé game $G_n(L, R)$ is played as follows: on the i^{th} move, with $1 \leq i \leq n$, the first player (*The Spoiler*) chooses an element from either L or R, and the second player (*The Duplicator*) responds by choosing an element from the other structure. Let l_i be the element of L chosen on the i^{th} move, and let r_i be the element of R chosen on the i^{th} move. The Duplicator wins the game iff the induced substructures $L[\{l_1, l_2, \ldots, l_n\}]$ and $R[\{r_1, r_2, \ldots, r_n\}]$ are isomorphic under the mapping sending l_i to r_i. The Ehrenfeucht-Fraïssé game was introduced by Ehrenfeucht in the context of linear orderings [4] in order to study the logical properties of ordinal numbers. A similar notion was formulated independently by Fraïssé in [7], which did not use the game-theoretic terminology. The main theorem [4, 7] states that the Duplicator has a winning strategy in the game $G_n(L, R)$ if and only if the two discrete structures are indistinguishable by sentences of quantifier depth at most n. The study of Ehrenfeucht-Fraïssé games has significant applications to finite model theory and computational complexity theory [3, 12].

In this paper, we shall investigate Ehrenfeucht-Fraïssé games on (simple) graphs. We will provide a full analysis of the Ehrenfeucht-Fraïssé game when L and R are both paths, and when L and R are both cycles. These problems have been examined in past papers [2, 8], and partial results have been found. The first complete solutions appear here.

2 The Ehrenfeucht-Fraïssé Game for Graphs

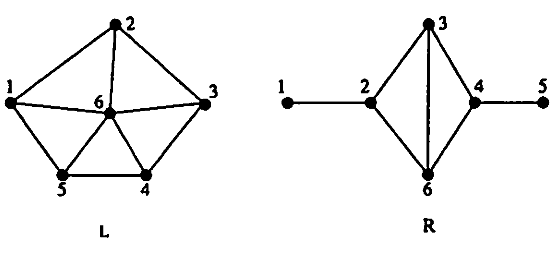
For two graphs L and R, the Duplicator wins the Ehrenfeucht-Fraïssé game $G_n(L, R)$ precisely when we have the following situation at the end of the game: $l_i l_j \in E(L)$ iff $r_i r_j \in E(R)$, for all $1 \leq i < j \leq n$. Now, we will allow for preassigned initial sequences, where each player has already made h moves in the game, for some integer h. Let $l_1, \ldots, l_h$ be the vertices chosen on L, and let $r_1, \ldots, r_h$ be the vertices chosen on R. Then the game $G_n((L; l_1, \ldots, l_h), (R; r_1, \ldots, r_h))$ consists of n moves: $l_{h+1}, \ldots, l_{h+n}$ are the moves made on L, and $r_{h+1}, \ldots, r_{h+n}$ are the moves made on R. The Duplicator wins precisely when we have the following situation at the end of the game: $l_i l_j \in E(L)$ iff $r_i r_j \in E(R)$, for $1 \leq i < j \leq h + n$.

Note that if for some i, j with $1 \leq i < j \leq h$, we have $l_i l_j \in E(L)$ and $r_i r_j \notin E(R)$ (or $l_i l_j \notin E(L)$ and $r_i r_j \in E(R)$), then for all n, the Spoiler automatically has a winning strategy in $G_n((L; l_1, \ldots, l_h), (R; r_1, \ldots, r_h))$.

We write $L \nsim_n R$ if the Spoiler has a winning strategy in $G_n(L, R)$ and $L \sim_n R$ if the Duplicator has a winning strategy in $G_n(L, R)$. Note that one player must have a winning strategy in $G_n(L, R)$, because this is a finite game with perfect information, and ties are not possible.

If $L \sim_n R$ for all integers n, we write $L \sim R$. Otherwise, we write $L \nsim R$.

Let us look at a specific example. Let L and R be the two graphs below, and consider the game $G_2(L, R)$. We prove that $L \nsim_2 R$.



The Spoiler's first move is $l_1 = 6$. Then the Duplicator must choose some vertex r_1 in R. Now, for each vertex $r_1 \in R$, there exists at least one vertex v such that r_1 is not adjacent to v. The Spoiler then chooses $r_2 = v$. Since any vertex the Duplicator picks as l_2 will be adjacent to $l_1 = 6$, we have $l_1 l_2 \in E(L)$ and $r_1 r_2 \notin E(R)$. Thus, the Duplicator has lost. Clearly, this argument can be generalized: if L has a universal vertex and R does not, then $L \nsim_2 R$.

Graphs can be viewed as a first-order theory (with equality) with a single binary predicate, E. We write $E(x, y)$ when vertices x and y are adjacent.

For any first-order sentence φ, its *quantifer depth* is denoted by $qd(\varphi)$, and is defined recursively by the following rules:

(i) If φ is atomic (i.e., φ has no quantifiers), then $qd(\varphi) = 0$.

(ii) If φ is $\neg\psi$, then $qd(\varphi) = qd(\psi)$.

(iii) If φ is $\psi \vee \eta$, $\psi \wedge \eta$, $\psi \to \eta$, or $\psi \leftrightarrow \eta$, then $qd(\varphi) = \max(qd(\psi), qd(\eta))$.

(iv) If φ is $(\forall v)\psi$ or $(\exists v)\psi$, then $qd(\varphi) = qd(\psi) + 1$.

In our graph-theoretic context, there are only two types of atomic first-order sentences: $x = y$ and $E(x, y)$.

Let S be the set of first-order sentences, and define

$$S_n = \{\varphi \in S : \; qd(\varphi) \leq n\}.$$

For graphs L and R, we write $L \cong_n R$ if L and R satisfy the same sentences in S_n. We say that L and R are *elementarily equivalent* (denoted by $L \cong R$) if for every sentence φ in S, φ is true in L iff φ is true in R.

Theorem 2.1 *Let L and R be graphs. Then, $L \sim_n R$ iff $L \cong_n R$.*

This theorem is a specific case of a more general result [4] which relates the theory of first-order logic to the Ehrenfeucht-Fraïssé game in the much more general context of finitary structures.

Theorem 2.1 can be applied in the following way: if we know that $L \sim_n R$, then every sentence φ with $qd(\varphi) \leq n$ must be true for both graphs, or for neither graph. Also, if we have a sentence φ with $qd(\varphi) = n$ that is true for only one of the two graphs, then we can immediately conclude that $L \nsim_n R$.

For example, the sentence $\varphi = (\exists w)\ (\forall x)\ ((w \neq x) \to E(w, x))$ models the property that a graph has a universal vertex. The quantifier depth of φ is 2. And this is consistent with the observation that we made earlier that if L has a universal vertex and R does not, then $L \nsim_2 R$. Using Ehrenfeucht-Fraïssé games, we can construct proofs that certain properties of finite graphs cannot be modelled by a first-order sentence. For example, the property of being bipartite is not expressible as a sentence in first-order logic as $C_{2^{n-1}+4}$ is bipartite, $C_{2^{n-1}+3}$ is not, and we shall see (Theorem 4.9) that $C_{2^{n-1}+4} \sim_n C_{2^{n-1}+3}$. So there is no first-order sentence that describes the property that a graph is bipartite.

The following corollaries are immediate.

Corollary 2.2 *Let L and R be two (finite) graphs. Then $L \sim R$ iff $L \cong R$.*

Corollary 2.3 *Both $\sim_n$ and $\sim$ are equivalence relations on the class of all graphs.*

One can quickly show that for each n, $\sim_n$ has only finitely many equivalence classes. Proofs of this result can be found in [3] and [10].

In [10], Rosenstein analyzes the Ehrenfeucht-Fraïssé game for linear orderings. In this game, the Duplicator wins iff the elements $l_1, l_2, \ldots, l_n$ are in the same order in L as the elements $r_1, r_2, \ldots, r_n$ are in R. In other words, the Duplicator wins precisely when we have the following situation at the end of the game: $l_i <_L l_j$ iff $r_i <_R r_j$.

Since a path is analogous to a linear ordering (the Hasse diagram of the latter being the former), it would appear that the Ehrenfeucht-Fraïssé game on linear orderings is equivalent to the Ehrenfeucht-Fraïssé game on paths. However, the latter is much more difficult to analyze due to the following "splitting lemma" which is true for linear orderings, but not true for paths:

Lemma 2.4 *Let L and R be linear orderings, and n be a positive integer. For each $a \in L$, define $L_{<a} = \{v \in L | v <_L a\}$ and $L_{>a} = \{v \in L | v >_L a\}$, and for each $b \in R$, define $R_{<b} = \{v \in R | v <_R b\}$, and $R_{>b} = \{v \in R | v >_R b\}$.*

Then the Duplicator has a winning strategy in the game $G_{n+1}(L, R)$ iff the following conditions are true:

1. *For every $a \in L$, there is an element $b \in R$ for which the Duplicator has a winning strategy in both the games $G_n(L_{<a}, R_{<b})$ and $G_n(L_{>a}, R_{>b})$.*

2. *For every $b \in R$, there is an element $a \in L$ for which the Duplicator has a winning strategy in both the games $G_n(L_{<a}, R_{<b})$ and $G_n(L_{>a}, R_{>b})$.*

Using Lemma 2.4, it is a straightforward induction exercise to prove that:

Theorem 2.5 *Let L and R be linear orderings. Then the Duplicator has a winning strategy in the Ehrenfeucht-Fraïssé game $G_n(L, R)$ iff $|L| = |R|$, or $|L|, |R| \geq 2^n - 1$.*

With this theorem [10], we have a complete analysis of the Ehrenfeucht-Fraïssé game on linear orderings. In the following section, we give a complete analysis of the Ehrenfeucht-Fraïssé game on paths. Since this splitting lemma fails for paths (we shall explain why in the next section), we will need to develop more sophisticated techniques to solve this problem.

3 The Ehrenfeucht-Fraïssé Game on Paths

Let P_n denote the path on n vertices, where the vertices are labelled $1, \dots, n$.

Given two positive integers m and n, we shall determine all the values of k for which $P_m \sim_k P_n$. Also, for any given k, we shall find all ordered pairs (m, n) for which $P_m \sim_k P_n$. This will give us a complete solution to the Ehrenfeucht-Fraïssé game on paths.

Note that if $m = n$, then $P_m \sim_k P_n$ for all k. So let us assume that $m \neq n$. In [2], Brown and Woodrow define $f(k)$ to be the smallest integer t such that $P_m \sim_k P_n$ if $m = n$ or $m, n \geq t$.

If we can find t such that $P_t \sim_k P_n$ for all $n \geq t$, then by transitivity, we have $P_m \sim_k P_n$ for all $m, n \geq t$. Hence, it follows that $f(k) \leq t$. Brown and Woodrow show that $f(k)$ is well-defined and prove that

$$\frac{1}{2} \cdot 2^k + 1 \leq f(k) \leq \frac{3}{2} \cdot 2^k + 1.$$

We will prove the following explicit formula for $f(k)$:

Theorem 3.1 *$f(1) = 1$, $f(2) = 4$, $f(3) = 7$, and $f(k) = 2^k$ for $k \geq 4$.*

Clearly, $f(1) = 1$. It is straightforward to verify that $P_4 \sim_2 P_n$, for all $n \geq 4$. Thus, $f(2) \leq 4$. To conclude that $f(2) = 4$, it suffices to show that $P_3 \nsim_2 P_4$. And this is immediate because P_3 has a universal vertex, but P_4 does not.

We now prove that $f(3) = 7$. First, let us show that $P_6 \nsim_3 P_7$. Consider the following first-order sentence of quantifier depth 3:

$$\varphi = (\exists v)(\exists w)(\forall x)(((v \neq x) \wedge (w \neq x)) \rightarrow (E(v, x) \vee E(w, x))).$$

In other words, the sentence φ models the property that there is a *dominating set* of size two. Since P_6 satisfies φ and P_7 does not, we have $P_6 \not\equiv_3 P_7$. By Theorem 2.1, we have $P_6 \nsim_3 P_7$. Therefore, we conclude that $f(3) > 6$.

Now we prove that $f(3) \leq 7$ by showing that $P_7 \sim_3 P_n$, for any $n \geq 7$. We first describe the Duplicator's response to the Spoiler's first move:

If the Spoiler plays $l_1 = 1, 2, 3, 4$, then the Duplicator plays $r_1 = l_1$.

If the Spoiler plays $l_1 = 5, 6, 7$, then the Duplicator plays $r_1 = n - (7 - l_1)$.

If the Spoiler plays $r_1 = 1, 2, 3$, then the Duplicator plays $l_1 = r_1$.

If the Spoiler plays $r_1 = 4, 5, \dots, n - 3$, then the Duplicator plays $l_1 = 4$.

If the Spoiler plays $r_1 = n - 2, n - 1, n$, then the Duplicator plays $l_1 = n - (7 - r_1)$.

Now, if the Spoiler selects r_2 so that its distance from r_1 is 1 or 2, then the Duplicator selects l_2 so that $l_2 - l_1 = r_2 - r_1$. Similarly, if the Spoiler selects

l_2 so that its distance from l_1 is 1 or 2, then the Duplicator selects r_2 so that $r_2 - r_1 = l_2 - l_1$.

If the Spoiler selects r_2 so that its distance from r_1 exceeds 2, then the Duplicator selects any l_2 such that $|l_2 - l_1| > 2$. Similarly, if the Spoiler selects l_2 so that its distance from l_1 exceeds 2, then the Duplicator selects any r_2 such that $|r_2 - r_1| > 2$.

The Spoiler makes one final move. If she plays r_3 so that it is adjacent to r_1 and/or r_2, then the Duplicator can play l_3 appropriately and win the game. If the Spoiler chooses r_3 so that it is not adjacent to either r_1 or r_2, then the Duplicator can choose any l_3 so that is is not adjacent to either l_1 or l_2, and the Duplicator wins in this case as well. This is possible because P_n has no dominating set of size 2, for any $n \geq 7$. So in all cases, the Duplicator wins the game. The Duplicator's play is analogous if the Spoiler plays l_3, as the Duplicator will select r_3 appropriately. Hence, we have shown that $P_7 \sim_3 P_n$, for all $n \geq 7$. We conclude that $f(3) = 7$.

Let us briefly explain why the Ehrenfeucht-Fraïssé game on paths is so difficult to analyze. Suppose that in the game $G_3(P_7, P_8)$, we have $l_1 = r_1 = 4$ and $l_2 = r_2 = 6$. Suppose the Spoiler selects $r_3 = 8$ on her final move. Then the Duplicator can only win by "jumping" to the other side of P_7, i.e., playing $l_3 = 1$ or $l_3 = 2$. These jump moves are perfectly legitimate; in fact, it is the only way that the Duplicator can win. Thus, the splitting lemma (Lemma 2.4) that worked for linear orderings fails for paths, due to the jump moves that are possible. Since we must consider these jump moves in our analysis of the Ehrenfeucht-Fraïssé game on paths, our analysis becomes much more complex.

To prove that $f(k) = 2^k$ for all $k \geq 4$, we must prove that $f(k) \geq 2^k$ and $f(k) \leq 2^k$. We prove the lower bound first.

3.1 Lower Bound

Let P be a path of finite length, and let a and b denote two vertices of P. Define $\rho_0(a,b) = (a = b)$ and $\rho_1(a,b) = E(a,b)$. For $i \geq 2$, recursively define

$$\rho_i(a,b) = (\exists c)\ (\rho_{2^k}(a,c) \wedge \rho_{i-2^k}(b,c)) \wedge (\neg \rho_{2^{k+1}-i}(a,b)),$$

where k is the unique integer for which $2^k + 1 \leq i \leq 2^{k+1}$.

Lemma 3.2 *Let a and b be two distinct vertices of P, and let n be an integer, with $n \geq 0$. If $1 \leq i \leq 2^n$, then $qd(\rho_i(a,b)) \leq n$. Furthermore, $\rho_i(a,b)$ is a sentence that holds if and only if* $\text{dist}(a,b) = i$, *where* $\text{dist}(a,b)$ *denotes the distance between vertices a and b.*

Proof We proceed by induction on n. The claim is trivial for $n = 0$.

Now suppose that we have proven the claim for $n = k$, i.e., for $1 \leq i \leq 2^k$. We now prove that the result holds for $n = k+1$. Let $2^k + 1 \leq i \leq 2^{k+1}$.

We recursively defined the sentence

$$\rho_i(a,b) = (\exists c)\ (\rho_{2^k}(a,c) \wedge \rho_{i-2^k}(b,c)) \wedge (\neg\rho_{2^{k+1}-i}(a,b)).$$

First we show that this sentence is true iff $\mathrm{dist}(a,b) = i$.

Note that $i-2^k \leq 2^k$ and $2^{k+1}-i \leq 2^k$. By the induction hypothesis, $\rho_i(a,b)$ holds iff $\mathrm{dist}(a,b) \neq 2^{k+1}-i$, and there exists a vertex c such that $\mathrm{dist}(a,c) = 2^k$ and $\mathrm{dist}(b,c) = i - 2^k$.

If a is between b and c, then $2^k = \mathrm{dist}(a,c) < \mathrm{dist}(b,c) = i - 2^k \leq 2^k$, a contradiction. If b is between a and c, then $\mathrm{dist}(a,b) = \mathrm{dist}(a,c) - \mathrm{dist}(b,c) = 2^{k+1}-i$, a contradiction. Thus, these conditions force c to be the middle vertex. In this case, we have $\mathrm{dist}(a,b) = \mathrm{dist}(a,c)+\mathrm{dist}(b,c) = i$. The converse is clearly true: if $\mathrm{dist}(a,b) = i$, then there must exist a vertex c between a and b such that $\mathrm{dist}(a,c) = 2^k$ and $\mathrm{dist}(b,c) = i - 2^k$. Furthermore, $\mathrm{dist}(a,b) \neq 2^{k+1} - i$, since $i > 2^k$. Thus, the sentence $\rho_i(a,b)$ holds iff $\mathrm{dist}(a,b) = i$.

By the induction hypothesis, $qd(\rho_{2^k}(a,c))$, $qd(\rho_{i-2^k}(b,c))$, $qd(\rho_{2^{k+1}-i}(a,b))$ are each at most k. Therefore, we have proven that

$$qd(\rho_i(a,b)) \leq \max(1+\max(k,k),k) = k+1.$$

□

We now define $\phi_j(b)$, for each $j \geq 0$. First, we define:

$$\begin{aligned}
\phi_0(b) &= (\neg\exists c)\ (\neg\exists d)\ (c \neq d) \wedge (E(c,b) \wedge E(d,b)).\\
\phi_1(b) &= (\exists c)\ (\forall d)\ (d \neq b \rightarrow\ \neg E(d,c)).\\
\phi_2(b) &= (\exists c)\ (\forall d)\ (E(d,c) \rightarrow E(d,b)).
\end{aligned}$$

For $j \geq 3$, recursively define

$$\phi_j(b) = (\exists c)\ (\phi_{2^k-2}(c) \wedge \rho_{j-2^k+2}(b,c)) \wedge \neg(\exists d)\left(\phi_{2^k-2}(d) \wedge \left(\bigvee_{i=1}^{j-2^k+1} \rho_i(b,d)\right)\right),$$

where k is the unique integer for which $2^k - 1 \leq j \leq 2^{k+1} - 2$.

This defines $\phi_j(b)$ for each $j \geq 0$.

Lemma 3.3 *Let b be a vertex of P, and let n be an integer, with $n \geq 2$. If $0 \leq j \leq 2^n - 2$, then $qd(\phi_j(b)) \leq n$. Furthermore, $\phi_j(b)$ is a sentence that holds if and only if j is the shortest distance from b to an endpoint of P.*

Proof We proceed by induction on n. The claim holds for $n = 2$, as $qd(\phi_j(b)) \leq 2$ for each of the sentences $\phi_0(b)$, $\phi_1(b)$, and $\phi_2(b)$, and for each $j = 0,1,2$, $\phi_j(b)$ is true iff j is the shortest distance from b to an endpoint of P. (Note that we cannot start the induction at $n = 1$ as $qd(\phi_0(b)) = 2$.)

Now suppose that we have proven the claim for $n = k$, i.e., for $0 \le j \le 2^k - 2$. We now prove that the result holds for $n = k + 1$. Let $2^k - 1 \le j \le 2^{k+1} - 2$.

We recursively defined

$$\phi_j(b) = (\exists c)\ (\phi_{2^k-2}(c) \wedge \rho_{j-2^k+2}(b,c)) \wedge \neg(\exists d) \left(\phi_{2^k-2}(d) \wedge \left(\bigvee_{i=1}^{j-2^k+1} \rho_i(b,d) \right) \right).$$

For convenience, define enddist(b) to be the shortest distance from b to an endpoint of P. We show that $\phi_j(b)$ holds iff enddist(b) $= j$.

Suppose that enddist(b) $= j$. Then there exists a vertex c with enddist(c) $= 2^k - 2$ and dist(b, c) $= j - 2^k + 2$. By the induction hypothesis and Lemma 3.2, c satisfies $\phi_{2^k-2}(c)$ and $\rho_{j-2^k+2}(b,c)$. Furthermore, there cannot exist d with $\left(\phi_{2^k-2}(d) \wedge \left(\bigvee_{i=1}^{j-2^k+1} \rho_i(b,d)\right)\right)$, as otherwise, enddist($b$) $\le (2^k - 2) + (j - 2^k + 1) = j - 1 < j$, which is a contradiction. Hence, $\phi_j(b)$ holds.

Now suppose that $\phi_j(b)$ holds. Then by Lemma 3.2 and the induction hypothesis, there exists a vertex c such that enddist(c) $= 2^k - 2$ and dist(b, c) $= j - 2^k + 2$. Also, there does not exist any vertex d with enddist(d) $= 2^k - 2$ and dist(b, d) $\le j - 2^k + 1$. This latter condition proves that enddist(b) $> (2^k - 2) + (j - 2^k + 1) = j - 1$. We wish to prove that enddist(b) $= j$.

Let v_1 and v_2 be the endpoints of P, with c between b and v_2. If dist(c, v_1) $= 2^k - 2$, then dist(b, v_2) $\ge$ dist(b, v_1) $=$ dist(c, v_1) $-$ dist(b, c) $= (2^k - 2) - (j - 2^k + 2) = 2^{k+1} - j - 4$. But $2^{k+1} - j - 4 = 2(2^k - j - 2) + j < j$, which contradicts enddist(b) $\ge j$. So we must have dist(c, v_2) $= 2^k - 2$, from which it follows that dist(b, v_2) $= (j - 2^k + 2) + (2^k - 2) = j$. Thus, enddist($b$) $= j$ and this proves our claim.

By the induction hypothesis and Lemma 3.2, $qd(\phi_{2^k-2}(c))$, $qd(\rho_{j-2^k+2}(b,c))$, and $qd(\phi_{2^{k+1}-j-4}(b))$ are at most k. And so we have proven that

$$qd(\phi_j(b)) = \max(1 + \max(k,k), k) = k + 1.$$

□

Lemma 3.4 *For each $k \ge 4$, define*

$$\varphi_k = (\exists a)\ (\forall b)\ \left((b \ne a) \to \left(\bigvee_{i=1}^{2^{k-2}} \rho_i(a,b) \right) \vee \left(\bigvee_{j=0}^{2^{k-2}-2} \phi_j(b) \right) \right).$$

Then, for $k \ge 4$, P_{2^k-1} satisfies φ_k, and P_{2^k} does not satisfy φ_k.

Proof First we show that φ_k is true for P_{2^k-1}. Let a be the vertex in the centre of the path. Now select any b, with $b \ne a$. We will show that either $\rho_i(a,b)$ or $\phi_j(b)$ is true, for some $1 \le i \le 2^{k-2}$ or $0 \le j \le 2^{k-2} - 2$.

Suppose on the contrary that neither condition holds. Then, we must have $\text{dist}(a,b) \geq 2^{k-2}+1$. Also, if v_1 and v_2 are the two endpoints of P_{2^k-1}, then we must have $\text{dist}(v_1,b) \geq 2^{k-2}-1$ and $\text{dist}(v_2,b) \geq 2^{k-2}-1$. If $b < a$, then we have $2^{k-1}-1 = \text{dist}(v_1,a) = \text{dist}(v_1,b)+\text{dist}(b,a) \geq (2^{k-2}+1)+(2^{k-2}-1) = 2^{k-1}$, a contradiction. We get the same contradiction if $b > a$. So for each b with $b \neq a$, either $\rho_i(a,b)$ is true for some $1 \leq i \leq 2^{k-2}$ or $\phi_j(b)$ is true for some $0 \leq j \leq 2^{k-2}-2$. Thus, φ_k is true for P_{2^k-1}.

Now we show that φ_k is not true for P_{2^k}. On the contrary, suppose such a vertex a exists. Then for each of the other 2^k-1 vertices, either $\rho_i(a,b)$ is true for some $1 \leq i \leq 2^{k-2}$ or $\phi_j(b)$ is true for some $0 \leq j \leq 2^{k-2}-2$. Since we are dealing with a path, $\rho_i(a,b)$ can be true for at most two vertices, for any i. Similarly, $\phi_j(b)$ can be true for at most two vertices, for any j. Since we have 2^{k-2} different possibilities for i and $2^{k-2}-1$ different possibilities for j, there are only $2(2^{k-2})+2(2^{k-2}-1) = 2^k-2$ possible choices for b. Hence, our path P_{2^k} contains at most $1+(2^k-2) = 2^k-1$ vertices, a contradiction. Hence, we conclude that φ_k is not true for P_{2^k}. □

Theorem 3.5 *For $k \geq 4$, we have $f(k) \geq 2^k$.*

Proof By Lemma 3.2 and Lemma 3.3, $qd(\rho_i(a,b)) \leq k-2$ for each $1 \leq i \leq 2^{k-2}$, and $qd(\phi_j(b)) \leq k-2$ for each $0 \leq j \leq 2^{k-2}-2$. Therefore, $qd(\varphi_k) \leq 2+\max(k-2,k-2) = k$. By Lemma 3.4, φ_k is true for P_{2^k-1} but not true for P_{2^k}. This proves that $P_{2^k-1} \not\equiv_k P_{2^k}$. By Theorem 2.1, $P_{2^k-1} \not\sim_k P_{2^k}$. Hence, we have proven that for all $k \geq 4$, we have $f(k) \geq 2^k$. □

Note that this proof fails for $k=3$, due to Lemma 3.3 (since $k=3$, we have $n=1$, and Lemma 3.3 requires $n \geq 2$). That is why we have $f(3)=7$, and not $f(3)=8$.

3.2 Upper Bound

To complete the proof, we must prove the upper bound, that $f(k) \leq 2^k$ for all $k \geq 4$. To do this, we describe a winning strategy for the Duplicator in the game $G_k(P_a,P_b)$, where $a,b \geq 2^k$.

If L and R are paths, we define l_i and l_j to be *chosen neighbours* in the game $G_n((L;l_1,l_2,\ldots,l_h),(R;r_1,r_2,\ldots,r_h))$ if $1 \leq i,j \leq h$, $i \neq j$, and there does not exist an index p with $\min(l_i,l_j) < l_p < \max(l_i,l_j)$. We have an analogous definition for r_i and r_j being chosen neighbours.

Let $L = P_a$ and $R = P_b$ for some positive integers a and b. If $h \geq 1$, define L_{lgap} to be the shortest distance between the left endpoint of L and a chosen vertex of L (i.e., the number of vertices to the left of the first l_i). Hence, $L_{lgap} = \min(l_1,l_2,\ldots,l_h)-1$. Also, define L_{rgap} to be the shortest distance between the right endpoint of L and a chosen vertex of L. Hence, $L_{lgap} = a-\max(l_1,l_2,\ldots,l_h)$. Similarly, we define $R_{lgap} = \min(r_1,r_2,\ldots,r_h)-1$, and $R_{rgap} = b-\max(r_1,r_2,\ldots,r_h)$.

Lemma 3.6 *Let $n \geq 1$ be a positive integer. Let $L = P_a$, $R = P_b$.*

Consider the game $G_n((L; l_1, l_2, \ldots, l_h), (R; r_1, r_2, \ldots, r_h))$, where we assume without loss that all of the l_i's are distinct and all of the r_i's are distinct. Then, for all $h \geq 1$, the Duplicator has a winning strategy in this game if all the following conditions are satisfied:

1. *If l_i and l_j are chosen neighbours with $|l_i - l_j| \leq 2^n$, then r_i and r_j are chosen neighbours with $|r_i - r_j| = |l_i - l_j|$. Also, if r_i and r_j are chosen neighbours with $|r_i - r_j| \leq 2^n$, then l_i and l_j are chosen neighbours with $|l_i - l_j| = |r_i - r_j|$.*
2. *$L_{lgap}, R_{lgap} \geq 2^n - 1$, or there exists an index m with $l_m = \min(l_1, \ldots, l_h)$, $r_m = \min(r_1, \ldots, r_h)$, and $L_{lgap} = R_{lgap}$.*
3. *$L_{rgap}, R_{rgap} \geq 2^n - 1$, or there exists an index m with $l_m = \max(l_1, \ldots, l_h)$, $r_m = \max(r_1, \ldots, r_h)$, and $L_{rgap} = R_{rgap}$.*
4. *Either there exist chosen neighbours l_i and l_j with $|l_i - l_j| \geq 2^{n+1}$, or we have $\max(L_{lgap}, L_{rgap}) \geq 2^{n+1} - 1$.*
5. *Either there exist chosen neighbours r_i and r_j with $|r_i - r_j| \geq 2^{n+1}$, or we have $\max(R_{lgap}, R_{rgap}) \geq 2^{n+1} - 1$.*

Proof First, note that condition 1 implies that

$$(L; l_1, l_2, \ldots, l_h) \sim_0 (R; r_1, r_2, \ldots, r_h).$$

We proceed by induction on n. Consider the case $n = 1$. Without loss, assume the Spoiler chooses l_{h+1} in L, and l_{h+1} is different from all the previous l_i's. There are three possible moves by the Spoiler: she can select l_{h+1} so that it is adjacent to two, one, or none of the vertices in $\{l_1, \ldots, l_h\}$. We shall prove that the Duplicator can always respond so that the two resulting subgraphs are isomorphic after $h+1$ vertices are chosen on each graph.

Suppose the Spoiler selects l_{h+1} so that it is adjacent to two chosen vertices l_i and l_j (l_i and l_j must be chosen neighbours). Then, $|l_i - l_j| = 2$. From condition 1, $|r_i - r_j| = 2$. Hence, the Duplicator can select r_{h+1} so that it is adjacent to both r_i and r_j.

Suppose the Spoiler selects l_{h+1} so that it is only adjacent to some l_i. Then the Duplicator must select r_{h+1} so that is is only adjacent to r_i. Assume this is not possible. There are three possible cases to consider, depending on the location of r_i:

Case 1: There exist distinct indices j, k with $r_i - r_j \leq 2$ and $r_k - r_i \leq 2$.

Case 2: There exists j with $r_j - r_i \leq 2$, $r_i = \min(r_1, r_2, \ldots, r_h)$ and $R_{lgap} = 0$.

Case 3: There exists j with $r_i - r_j \leq 2$, $r_i = \max(r_1, r_2, \ldots, r_h)$ and $R_{rgap} = 0$.

In the first case, condition 1 implies that $l_i - l_j \leq 2$ and $l_k - l_i \leq 2$, so the Spoiler could not have chosen l_{h+1} so that it is only adjacent to l_i. In the second case, the first two conditions imply that $l_j - l_i \leq 2$ and $L_{lgap} = 0$, and so we obtain our contradiction here as well. The third case is identical to the second. Thus, if the Spoiler selects l_{h+1} so that it is adjacent to only l_i, then the Duplicator can select r_{h+1} so that it is adjacent to only r_i.

Finally, we consider the case when l_{h+1} is adjacent to none of the chosen vertices. Then by condition 5, either there exist chosen neighbours r_i and r_j with $|r_i - r_j| \geq 4$, or $\max(R_{lgap}, R_{rgap}) \geq 3$. And in both situations, we see that the Duplicator can play r_{h+1} so that it is adjacent to none of $\{r_1, r_2, \ldots, r_h\}$.

Therefore, we have verified the lemma for the base case $n = 1$. Let us suppose the lemma is true for $n = k$. If all five conditions of the lemma hold for the game $G_{k+1}((L; l_1, \ldots, l_h), (R; r_1, \ldots, r_h))$, we shall prove that the Duplicator can respond to the Spoiler's next move to create a game $G' = G_k((L; l_1, \ldots, l_{h+1}), (R; r_1, \ldots, r_{h+1}))$ which also satisfies the five conditions of the lemma. Then by the induction hypothesis, the Duplicator has a winning strategy in the original game.

Without loss, assume the Spoiler selects a vertex l_{h+1} in L. There are two cases to consider. Either the Spoiler plays l_{h+1} between some two chosen vertices l_p and l_q, or she plays l_{h+1} so that it is between a chosen vertex and an endpoint of P_a.

Consider the first case. From the given conditions, $|l_p - l_q| = |r_p - r_q|$ or $|l_p - l_q|, |r_p - r_q| \geq 2^{k+1} + 1$. Hence, the Duplicator can respond to the Spoiler so that if $|l_t - l_{h+1}| \leq 2^k$ or $|r_t - r_{h+1}| \leq 2^k$ (for $t = p, q$), then $|l_t - l_{h+1}| = |r_t - r_{h+1}|$. The interesting case occurs when the Spoiler selects l_{h+1} between two chosen vertices l_p and l_q such that $|l_{h+1} - l_p|, |l_{h+1} - l_q| \geq 2^k + 1$. Then we employ condition 5. If there exist chosen neighbours r_i and r_j with $|r_i - r_j| \geq 2^{k+2}$, then the Duplicator selects r_{h+1} between r_i and r_j so that $r_{h+1} = \min(r_i, r_j) + (2^k + 1)$. If $R_{lgap} \geq 2^{k+2} - 1$, then the Duplicator selects $r_{h+1} = \min(r_1, r_2, \ldots, r_h) - (2^k + 1)$. Finally, if $R_{rgap} \geq 2^{k+2} - 1$, then the Duplicator selects $r_{h+1} = \max(r_1, r_2, \ldots, r_h) + (2^k + 1)$. In all scenarios, we see that the first three conditions hold in G'.

Let us now consider the case when the Spoiler selects l_{h+1} so that it is between a chosen vertex and an endpoint of P_a. Let us just examine the case where $l_{h+1} < \min(l_1, \ldots, l_h) = l_m$, as the other case follows by symmetry. If $l_m = r_m$, then the Duplicator selects $r_{h+1} = l_{h+1}$. So suppose $L_{lgap} = l_m - 1 \geq 2^{k+1} - 1$ and $R_{lgap} \geq 2^{k+1} - 1$. If $l_{h+1} \leq 2^k - 1$, the Duplicator makes $r_{h+1} = l_{h+1}$, verifying condition 2. If $l_m - l_{h+1} \leq 2^k$, then the Duplicator makes $r_m - r_{h+1} = l_m - l_{h+1}$, verifying condition 1. Otherwise, the Duplicator uses condition 5 and follows the strategy described at the end of the previous paragraph. In all scenarios, we see that the first three conditions hold in G'.

To conclude the proof, we prove that conditions 4 and 5 must hold for G'. By our assumption, condition 4 holds in the original game. The first case to consider is that there exist chosen neigbours l_i and l_j with $|l_i - l_j| \geq 2^{k+2}$. If l_{h+1} is chosen between l_i and l_j, then we have $|l_i - l_{h+1}| \geq 2^{k+1}$ or $|l_j - l_{h+1}| \geq 2^{k+1}$.

Otherwise, l_i and l_j are chosen neighbours on G' with $|l_i - l_j| \geq 2^{k+2} > 2^{k+1}$. The other case to consider is $\max(L_{lgap}, L_{rgap}) \geq 2^{k+2} - 1$. Without loss, suppose $L_{lgap} \geq 2^{k+2} - 1$, and let $l_m = \min(l_1, l_2, \ldots, l_h)$. Then for any l_{h+1}, we must have $l_{h+1} - 1 \geq 2^{k+1} - 1$ or $|l_{h+1} - l_m| \geq 2^{k+1}$. Hence, condition 4 must hold in G', regardless of the choice for l_{h+1}. Similarly, condition 5 must hold in G', regardless of the choice for r_{h+1}.

Therefore, we have proven that for any move l_{h+1} that the Spoiler makes, the Duplicator can respond so that the five conditions of the lemma are verified for G'. And so, by the induction hypothesis, we conclude that the Duplicator has a winning strategy in the game $G_{k+1}((L; l_1, \ldots, l_h), (R; r_1, \ldots, r_h))$. This completes the induction. □

Theorem 3.7 *$f(k) \leq 2^k$ for all $k \geq 4$.*

Proof Fix $k \geq 4$. Let $a = 2^k$ and $b = 2^k + h$ for some arbitrary $h \geq 1$. We shall prove that $P_a \sim_k P_b$. By transitivity, this will imply that $P_m \sim_k P_n$ for $m, n \geq 2^k$, and hence, it follows by definition that $f(k) \leq 2^k$.

By symmetry, assume that on the Spoiler's first move, she will play either $l_1 \leq 2^{k-1}$ or $r_1 \leq 2^{k-1} + \lfloor \frac{h+1}{2} \rfloor$. There are two cases to consider.

Case 1: The Spoiler selects $l_1 \leq 2^{k-1} - 1$ or $r_1 \leq 2^{k-1} - 1$.

The Duplicator responds by making $r_1 = l_1$. If the Spoiler plays l_2 or r_2 with $l_2 < l_1$ or $r_2 < r_1$, then the Duplicator will make $l_2 = r_2$. If the Spoiler makes $l_2 - l_1 \leq 2^{k-2}$ or $r_2 - r_1 \leq 2^{k-2}$, then the Duplicator will play so that $l_2 - l_1 = r_2 - r_1$. If the Spoiler makes $a - l_2 \leq 2^{k-2} - 2$ or $b - r_2 \leq 2^{k-2} - 2$, then the Duplicator will play so that $a - l_2 = b - r_2$. In all other cases, the Duplicator will select $l_2 = l_1 + (2^{k-2} + 1)$ or $r_2 = r_1 + (2^{k-2} + 1)$.

We shall prove that regardless of what the Spoiler does on her third move, the Duplicator can respond in such a way that all five conditions of Lemma 3.6 are satisfied in the game $G' = G_{k-3}((L; l_1, l_2, l_3), (R; r_1, r_2, r_3))$. This will prove that the Duplicator has a winning strategy in $G_k(L, R)$, since $k \geq 4$.

First, we show that the last two conditions of the lemma must hold in the game G'. Suppose this is not true. Then, $L_{lgap} < 2^{k-2} - 1$, $L_{rgap} < 2^{k-2} - 1$, and for all pairs of chosen neighbours (l_i, l_j), we have $|l_i - l_j| < 2^{k-2}$. Then, $2^k - 1 = a - 1 < (2^{k-2} - 1) + 2^{k-2} + 2^{k-2} + (2^{k-2} - 1) = 2^k - 2$, a contradiction. Therefore, condition 4 must be satisfied in G', and similarly, so must condition 5. To complete the proof, we just need to verify that the Duplicator can play his third move so that the first three conditions are satisfied in G'.

The cases when $l_2 = r_2$ and $l_2 - l_1 = r_2 - r_1$ are easily dealt with, so let us examine the last two cases. Consider the case when $a - l_2 = b - r_2$. If the Spoiler selects l_3 such that $l_3 < l_1$ or $l_3 > l_2$ (or r_3 such that $r_3 < r_1$ or $r_3 > r_2$), then the Duplicator copies the move on the other graph (i.e. he will make $r_3 = l_3$ or $b - r_3 = a - l_3$), and this will satisfy the given conditions. Thus, assume the Spoiler will play l_3 between l_1 and l_2, or r_3 between r_1 and r_2. We have

$r_2 - r_1 \geq l_2 - l_1 = (a-1)-(a-l_2)-(l_1-1) \geq (2^k-1)-(2^{k-2}-2)-(2^{k-1}-2) = 2^{k-2}+3$. Thus, the Duplicator can respond in such a way that $|l_3 - l_i| \leq 2^{k-3}$ iff $|r_3 - r_i| \leq 2^{k-3}$, for $i = 1, 2$. We quickly see that the first three conditions of the lemma are satisfied by this strategy.

Finally, we consider the case when the Duplicator selects $l_2 = l_1 + (2^{k-2}+1)$ or $r_2 = r_1 + (2^{k-2}+1)$ on his second move. Suppose that the Duplicator selects l_2, as the other case follows similarly. Let us assume that $r_2 - r_1 > 2^{k-2}+1$, otherwise $|l_2 - l_1| = |r_2 - r_1|$, and this case is easily dealt with. We have $a - l_2 = a-(l_2-l_1)-l_1 \geq 2^k-(2^{k-2}+1)-(2^{k-1}-1) = 2^{k-2}$, and $b - r_2 \geq 2^{k-2}-1$.

If the Spoiler selects l_3 or r_3 with $a - l_3 \leq 2^{k-3}-2$ or $b - r_3 \leq 2^{k-3}-2$, then the Duplicator will make $a - l_3 = b - r_3$. If the Spoiler makes $|l_3 - l_i| \leq 2^{k-3}$ or $|r_3 - r_i| \leq 2^{k-3}$ for $i = 1, 2$, then the Duplicator will respond so that $|l_3 - l_i| = |r_3 - r_i|$. Otherwise, the Duplicator will play $l_3 = l_2 + (2^{k-3}+1)$ if the Spoiler plays r_3, or he will play $r_3 = r_2 - (2^{k-3}+1)$ if the Spoiler plays l_3. In all situations, the first three conditions of Lemma 3.6 are satisfied in G'. Therefore, we conclude that regardless of what the Spoiler does, the Duplicator can play his first three moves so that he has a winning strategy in G'.

Case 2: The Spoiler selects $l_1 = 2^{k-1}$ or $r_1 = 2^{k-1}+m$, for some m with $0 \leq m \leq \lfloor \frac{h+1}{2} \rfloor$.

The Duplicator responds by selecting $l_1 = 2^{k-1}$ if the Spoiler chose r_1, and $r_1 = 2^{k-1}$ if the Spoiler chose l_1.

If the Spoiler selects $l_2 \leq 2^{k-2}-1$ or $r_2 \leq 2^{k-2}-1$, then the Duplicator makes $l_2 = r_2$. If she plays $a - l_2 \leq 2^{k-2}-2$ or $b - r_2 \leq 2^{k-2}-2$, then the Duplicator makes $a - l_2 = b - r_2$. If she plays $|l_2 - l_1| \leq 2^{k-2}$ or $|r_2 - r_1| \leq 2^{k-2}$, then the Duplicator makes $|l_2 - l_1| = |r_2 - r_1|$. In all other cases, the Duplicator will play $l_2 = l_1 + (2^{k-2}+1)$ or $r_2 = r_1 + (2^{k-2}+1)$.

We quickly see that this strategy ensures that the first three conditions of the lemma hold for $G_{k-2}((L; l_1, l_2), (R; r_1, r_2))$. Condition 4 is satisfied because $l_2 < l_1$ implies that $L_{rgap} = a - l_1 = 2^{k-1} > 2^{k-1}-1$ and $l_2 > l_1$ implies that $L_{lgap} = l_1 - 1 = 2^{k-1}-1$. Similarly, condition 5 is satisfied. Therefore, by Lemma 3.6, the Duplicator has a winning strategy in $G_{k-2}((L; l_1, l_2), (R; r_1, r_2))$, and so he has a winning strategy in $G_k(L, R)$. □

3.3 Main Result

Earlier, we showed that $f(1) = 1, f(2) = 4, f(3) = 7$. By Theorem 3.5 and Theorem 3.7, we have shown that $f(k) = 2^k$ for all $k \geq 4$. In other words, we have proven that $t = f(k)$ is the smallest positive integer for which $P_m \sim_k P_n$ if $m = n$ or $m, n \geq t$. In this section, we shall complete our analysis of the Ehrenfeucht-Fraïssé game on paths by showing that if $m \neq n$ and $\min(m, n) < f(k)$, then $P_m \nsim_k P_n$. Thus, for any given m, n, k, we can determine who has the winning strategy in the game $G_k(P_m, P_n)$.

For each positive integer t, we define a first-order sentence ω_t:

For $t \leq 6$, we define the following:

$$\begin{aligned}
\omega_1 &= (\forall a)(\forall b)(a = b). \\
\omega_2 &= (\forall a)(\forall b)(a \neq b \rightarrow E(a,b)). \\
\omega_3 &= (\exists a)(\forall b)(b \neq a \rightarrow E(a,b)). \\
\omega_4 &= (\forall a)(\exists b)(\forall c)((E(a,c) \rightarrow c = b) \vee (E(b,c) \rightarrow c = a)). \\
\omega_5 &= (\exists a)(\forall b)(a \neq b \rightarrow E(a,b) \vee ((\exists c)(E(a,c) \wedge E(b,c)))). \\
\omega_6 &= (\exists a)(\exists b)(\forall c)(((c \neq a) \wedge (c \neq b)) \rightarrow E(a,c) \vee E(b,c)).
\end{aligned}$$

And for $t \geq 7$, define:

$$\omega_t = (\forall a)\ (\exists b) \left(\left(\bigvee_{i=1}^{\lfloor \frac{t+2}{4} \rfloor} \rho_i(a,b) \right) \wedge \left(\bigvee_{j=0}^{\lfloor \frac{t-4}{4} \rfloor} \phi_j(b) \right) \right) \quad \text{if } t \text{ is even, and}$$

$$\omega_t = (\exists a)\ (\forall b) \left(b \neq a \rightarrow \left(\bigvee_{i=1}^{\lfloor \frac{t+1}{4} \rfloor} \rho_i(a,b) \right) \vee \left(\bigvee_{j=0}^{\lfloor \frac{t-5}{4} \rfloor} \phi_j(b) \right) \right) \quad \text{if } t \text{ is odd.}$$

Note that $\omega_{2k-1} = \varphi_k$, which we defined in Lemma 3.4.

Lemma 3.8 *Let t be a positive integer. Then P_t satisfies ω_t, and P_n does not satisfy ω_t, for all $n > t$.*

Proof The proof of this lemma follows the same spirit as Lemma 3.4. We easily verify that the lemma is true for $t \leq 6$, so consider the case $t \geq 7$. First, we examine the case when t is odd.

Let a be the vertex in the centre of the path P_t. Now select any b, with $b \neq a$. We will show that either $\rho_i(a,b)$ or $\phi_j(b)$ is true, for some $1 \leq i \leq \lfloor \frac{t+1}{4} \rfloor$ or $0 \leq j \leq \lfloor \frac{t-5}{4} \rfloor$.

Suppose on the contrary that neither condition holds. Then, we must have $\text{dist}(a,b) \geq \lfloor \frac{t+1}{4} \rfloor + 1$. Also, if v_1 and v_2 are the two endpoints of P_t, then we must have $\text{dist}(v_1,b) \geq \lfloor \frac{t-5}{4} \rfloor + 1$ and $\text{dist}(v_2,b) \geq \lfloor \frac{t-5}{4} \rfloor + 1$. If $b < a$, then we have $\frac{t-1}{2} = \text{dist}(v_1,a) = \text{dist}(v_1,b) + \text{dist}(b,a) \geq (\lfloor \frac{t-5}{4} \rfloor + 1) + (\lfloor \frac{t+1}{4} \rfloor + 1) = \frac{t+1}{2}$, which is a contradiction. We get the same contradiction if $b > a$. This proves that P_t satisfies ω_t. Now we show that ω_t is not true for P_n, for any $n > t$. On the contrary, suppose that ω_t is true for P_n. Since we are dealing with a path, $\rho_i(a,b)$ can be true for at most two vertices, for any i. Similarly, $\phi_j(b)$ can be true for at most two vertices, for any j. Therefore,

$$n \leq 1 + 2 \cdot \left(\left\lfloor \frac{t+1}{4} \right\rfloor + \left(\left\lfloor \frac{t-5}{4} \right\rfloor + 1 \right) \right) = 1 + 2 \cdot \left(\frac{t-1}{2} \right) = t.$$

That gives us our desired contradiction.

Now we examine the case when t is even. Without loss, assume $a \le \frac{t}{2}$. If $1 \le a \le \lfloor \frac{t}{4} \rfloor$, then select $b = 1$. Then $\phi_j(b)$ is true for $j = 0$ and $\rho_i(a, b)$ is true for $i = a - b \le \lfloor \frac{t}{4} \rfloor - 1 < \lfloor \frac{t+2}{4} \rfloor$. Otherwise, select $b = \lfloor \frac{t}{4} \rfloor$. Then $\phi_j(b)$ is true for $j = \lfloor \frac{t-4}{4} \rfloor$ and $\rho_i(a, b)$ is true for $i = a - b \le \frac{t}{2} - \lfloor \frac{t}{4} \rfloor = \lfloor \frac{t+2}{4} \rfloor$. This proves that P_t satisfies ω_t.

Finally, we show that ω_t is not true for any P_n, with $n > t$. Suppose ω_t is true for P_n, with $n > t$. Select $a = \frac{t}{2} + 1$, and let v be either endpoint of P_{t+1}. Then, $\text{dist}(a, v) \ge a - 1 = \frac{t}{2}$. For any $b \ne a$, if $\text{dist}(a, b) \le \lfloor \frac{t+2}{4} \rfloor$ and $\text{dist}(b, v) \le \lfloor \frac{t-4}{4} \rfloor$, then $\frac{t}{2} \le \text{dist}(a, v) = \text{dist}(a, b) + \text{dist}(b, v) \le \lfloor \frac{t+2}{4} \rfloor + \lfloor \frac{t-4}{4} \rfloor = \frac{t}{2} - 1$, a contradiction.

So in both cases (t odd and t even), we have proven that P_t satisfies ω_t, and P_n does not satisfy ω_t, for all $n > t$. □

Theorem 3.9 *Let $k \ge 2$. Then $P_m \not\sim_k P_n$ if $m \ne n$ and* $\min(m, n) < f(k)$.

Proof From Theorem 2.1 and Lemma 3.8, we have $P_t \not\sim_r P_n$ for all $n > t$, where $r = qd(w_t)$. Note that $P_t \not\sim_r P_n$ implies that $P_t \not\sim_{r'} P_n$, for all $r' > r$. So many of our proofs to show that $P_m \not\sim_k P_n$ will follow from previous cases. (For example, $P_3 \not\sim_2 P_n$ for $n > 3$ implies that for any $k > 2$, we have $P_3 \not\sim_k P_n$).

If $1 \le t \le 3$, then $r = qd(w_t) = 2$, and so $P_t \not\sim_2 P_n$ for all $n > t$. This proves the theorem for $k = 2$, since $f(2) = 4$. If $4 \le t \le 6$, then $r = qd(w_t) = 3$, and so $P_t \not\sim_3 P_n$ for all $n > t$. This proves the theorem for $k = 3$, since $f(3) = 7$. If $t = 7$, then $r = qd(w_7) = 2 + \max\{1, 2\} = 4$. Thus, $P_7 \not\sim_4 P_n$ for all $n > 7$.

Now consider $t \ge 8$. By Lemma 3.2 and 3.3, we have $qd(\rho_i(a, b)) \le \lceil \log_2 i \rceil$ and $qd(\phi_j(b)) \le \lceil \log_2(j + 2) \rceil$.

If t is odd, then we have

$$\begin{aligned} qd(w_t) &\le 2 + \max\left\{ \left\lceil \log_2 \left\lfloor \frac{t+1}{4} \right\rfloor \right\rceil, \left\lceil \log_2 \left(\left\lfloor \frac{t-5}{4} \right\rfloor + 2 \right) \right\rceil \right\} \\ &= 2 + \left\lceil \log_2 \left\lfloor \frac{t+3}{4} \right\rfloor \right\rceil . \end{aligned}$$

And similarly, if t is even, then we have

$$qd(w_t) \le 2 + \left\lceil \log_2 \left\lfloor \frac{t+4}{4} \right\rfloor \right\rceil .$$

Fix $k \ge 4$. Then for any $2^{k-1} \le t \le 2^k - 1$, we have $qd(w_t) \le 2 + (k-2) = k$, and so it follows that $P_t \not\sim_k P_n$ for all $n > t$. Hence, $P_m \not\sim_k P_n$ if $m \ne n$ and $\min(m, n) < 2^k = f(k)$. This proves the theorem for all $k \ge 4$. □

We have proven that $P_m \sim_k P_n$ whenever $m = n$ or $m, n \ge f(k)$. And Theorem 3.9 tells us that $P_m \not\sim_k P_n$ if $m \ne n$ and $\min(m, n) < f(k)$. Hence, we have now derived a complete solution to the Ehrenfeucht-Fraïssé game, when L and R are both paths.

Theorem 3.10 *Let $f : \mathbb{N} \to \mathbb{N}$ be a function with $f(1) = 1$, $f(2) = 4$, $f(3) = 7$, and $f(k) = 2^k$ for all $k \geq 4$. Then for each positive integer k, the Duplicator wins the game $G_k(P_m, P_n)$ if and only if $m = n$ or $m, n \geq f(k)$.*

4 The Ehrenfeucht-Fraïssé Game on Cycles

We now give a complete analysis of the Ehrenfeucht-Fraïssé game, when L and R are both cycles. Let C_n be the cycle on n vertices, with vertices labelled $1, 2, \ldots, n$. Let us define $g(k)$ to be the smallest integer t such that $C_m \sim_k C_n$ if $m = n$ or $m, n \geq t$.

In [2], Brown and Woodrow show that $g(k)$ is well-defined and prove that $g(k) > 2^{k-1}$. In this paper, we will prove the following explicit result:

Theorem 4.1 *$g(1) = 1$, $g(2) = 4$, and $g(k) = 2^{k-1} + 3$, for all $k \geq 3$.*

Clearly, $g(1) = 1$. It is straightforward to verify that $C_4 \sim_2 C_n$, for all $n \geq 4$. Thus, $g(2) \leq 4$. To conclude that $g(2) = 4$, it suffices to show that $C_3 \not\sim_2 C_4$. And this is immediate because C_3 has a universal vertex, but C_4 does not.

Let us prove that $g(k) = 2^{k-1} + 3$, for all $k \geq 3$. First we prove the lower bound.

4.1 Lower Bound

Lemma 4.2 *For each $k \geq 3$, define*

$$\varphi_k = (\forall a)(\exists b)(\forall c)\left(((c \neq a) \wedge (c \neq b)) \to \left(\bigvee_{i=1}^{2^{k-3}} \rho_i(a,c)\right) \vee \left(\bigvee_{i=1}^{2^{k-3}} \rho_i(b,c)\right)\right).$$

Then, for $k \geq 3$, $C_{2^{k-1}+2}$ satisfies φ_k, and $C_{2^{k-1}+3}$ does not satisfy φ_k.

Proof This proof is similar to the proof of Lemma 3.4, and so we omit the details. □

Theorem 4.3 *For $k \geq 3$, we have $g(k) \geq 2^{k-1} + 3$.*

Proof Let $k \geq 3$. By Lemma 3.2, if $1 \leq i \leq 2^{k-3}$, then $qd(\rho_i(a,b)) \leq k - 3$. Therefore, we have $qd(\varphi_k) \leq 3 + (k-3) = k$. By Lemma 4.2, φ_k is true for $C_{2^{k-1}+2}$ and not true for $C_{2^{k-1}+3}$. This proves that $C_{2^{k-1}+2} \not\equiv_k C_{2^{k-1}+3}$. By Theorem 2.1, we have $C_{2^{k-1}+2} \not\sim_k C_{2^{k-1}+3}$. Hence, $g(k) \geq 2^{k-1} + 3$. □

4.2 Upper Bound

Lemma 4.4 *Let $L = P_a$ and $R = P_b$.*

Consider the game $G_n((L; l_1, l_2), (R; r_1, r_2))$, where $l_1 = 1, l_2 = a, r_1 = 1, r_2 = b$. Then the Duplicator has a winning strategy if $|l_2 - l_1| \geq 2^n + 3$ and $|r_2 - r_1| \geq 2^n + 3$.

Proof First note that the result is trivial if $n = 1$ or $n = 2$, so assume $n \geq 3$.

If the Spoiler plays so that $|l_3 - l_i| \leq 2^{n-1}$ or $|r_3 - r_i| \leq 2^{n-1}$ for $i = 1, 2$, then the Duplicator responds by making $r_3 - r_i = l_3 - l_i$. Otherwise, the Spoiler will play so that $|l_3 - l_i| \geq 2^{n-1} + 1$ or $|r_3 - r_i| \geq 2^{n-1} + 1$ for $i = 1, 2$. In this case, the Duplicator plays l_3 so that $l_3 = l_2 - (2^{n-1} + 1)$ or r_3 so that $r_3 = r_2 - (2^{n-1} + 1)$. Thus, all four of the distances $|l_3 - l_1|, |l_3 - l_2|, |r_3 - r_1|, |r_3 - r_2|$ are at least $2^{n-1} + 1$. Also, $\max(|l_3 - l_1|, |l_3 - l_2|) \geq 2^{n-1} + 2$ and $\max(|r_3 - r_1|, |r_3 - r_2|) \geq 2^{n-1} + 2$. Without loss, assume that $|l_3 - l_1|, |r_3 - r_1| \geq 2^{n-1} + 2$.

If the Spoiler plays her next move such that $|l_4 - l_i| \leq 2^{n-2}$ or $|r_4 - r_i| \leq 2^{n-2}$ (for some $i = 1, 2, 3$), then the Duplicator responds by making $l_4 - l_i = r_4 - r_i$. Otherwise, the Duplicator plays $l_4 = l_3 - (2^{n-2} + 1)$ or $r_4 = r_3 - (2^{n-2} + 1)$.

Now, there are $(n - 2)$ moves left in the game. We prove that the five conditions of Lemma 3.6 hold in the game $G_{n-2}((L; l_1, l_2, l_3, l_4), (R; r_1, r_2, r_3, r_4))$. The first condition is true, by Duplicator's strategy described above. The second and third conditions hold trivially since $L_{lgap} = L_{rgap} = R_{lgap} = R_{rgap} = 0$. Finally, we see that there must exist chosen neighbours l_i, l_j, r_i, r_j for which $|l_i - l_j|, |r_i - r_j| \geq 2^{n-1} + 1 > 2^{n-1}$. Therefore, all five conditions are satisfied, and so the Duplicator has a winning strategy in this reduced game. Hence, he has a winning strategy in the game $G_n((L; l_1, l_2), (R; r_1, r_2))$.

□

Consider two cycles, $L = C_a$ and $R = C_b$. In a cycle, each first move on C_a is equivalent to every other first move. The same result holds for C_b. Hence, the game $G_k(C_a, C_b)$ is equivalent to the game $G_{k-1}((C_a; l_1), (C_b; r_1))$, for any $l_1 \in L$ and $r_1 \in R$. Without loss, let us assume that $l_1 = 1$ and $r_1 = 1$.

Now, we take our cycle C_a, "cut" it at l_1, and turn it into the path P_{a+1}. We do the same for C_b. This motivates the following lemma, which is easy to verify:

Lemma 4.5 *The Duplicator has a winning strategy in $G_{k-1}((P_{a+1}; 1, a+1), (P_{b+1}; 1, b+1))$ if and only if he has a winning strategy in $G_{k-1}((C_a; l_1), (C_b; r_1))$.*

We now prove our upper bound on $g(k)$.

Theorem 4.6 $g(k) \leq 2^{k-1} + 3$, *for all* $k \geq 3$.

Proof Let $k \geq 3$. By Lemma 4.5, the Duplicator has a winning strategy in the game $G_k(C_a, C_b) = G_{k-1}((C_a; l_1), (C_b; r_1))$ iff he has a winning strategy in the game $G_{k-1}((P_{a+1};1,a+1),(P_{b+1};1,b+1))$. By Lemma 4.4, if $a=b$, or $a,b \geq 2^{k-1}+3$, then we have $(P_{a+1};1,a+1) \sim_{k-1} (P_{b+1};1,b+1)$. In other words, we have $C_a \sim_k C_b$, where $a,b \geq 2^{k-1}+3$. And so it follows that $g(k) \leq 2^{k-1}+3$. □

Combining Theorems 4.3 and 4.6, we conclude that $g(k) = 2^{k-1}+3$ for all $k \geq 3$.

Now, we complete our analysis of the Ehrenfeucht-Fraïssé game on cycles by showing that if $m \neq n$ and $\min(m,n) < g(k)$, then $C_m \nsim_k C_n$. Thus, for any given m,n,k, we can determine who has the winning strategy in the game $G_k(C_m, C_n)$.

For each $t \geq 3$, we define a first-order sentence ω_t:

$$\begin{aligned} \omega_3 &= (\forall a)(\forall b)(a \neq b \rightarrow E(a,b)). \\ \omega_4 &= (\forall a)(\exists b)(\forall c)(((c \neq a) \wedge (c \neq b)) \rightarrow E(a,c) \wedge E(b,c)). \end{aligned}$$

And for $t \geq 5$, define

$$\omega_t = (\forall a)(\exists b)(\forall c)\left(((c \neq a) \wedge (c \neq b)) \rightarrow \left(\bigvee_{i=1}^{\lfloor \frac{t+1}{4} \rfloor} \rho_i(a,c)\right) \vee \left(\bigvee_{i=1}^{\lfloor \frac{t+1}{4} \rfloor} \rho_i(b,c)\right)\right).$$

Lemma 4.7 *Let $t \geq 3$ be a positive integer. Then C_t satisfies ω_t, and C_n does not satisfy ω_t, for all $n > t$.*

Proof This proof is similar to the proof of Lemma 3.8, and so we omit the details. □

Theorem 4.8 *Let $k \geq 2$. Then $C_m \nsim_k C_n$ if $m \neq n$ and* $\min(m,n) < g(k)$.

Proof From Theorem 2.1 and Lemma 4.7, we have $C_t \nsim_r C_n$ for all $n > t$, where $r = qd(\omega_t)$. Note that $C_t \nsim_r C_n$ implies that $C_t \nsim_{r'} C_n$, for all $r' > r$.

If $t = 3$, then $r = qd(\omega_t) = 2$, and so $C_3 \nsim_2 C_n$ for all $n > t$. This proves the theorem for $k = 2$. If $t = 4$, then $r = qd(\omega_t) = 3$, and so $C_4 \nsim_3 C_n$ for all $n > 4$.

Now consider $t \geq 5$. By Lemma 3.2, we have $qd(\rho_i(a,b)) \leq \lceil \log_2 i \rceil$. We have

$$qd(\omega_t) \leq 3 + \left\lceil \log_2 \left\lfloor \frac{t+1}{4} \right\rfloor \right\rceil.$$

Fix $k \geq 3$. Then for any $2^{k-2}+3 \leq t \leq 2^{k-1}+2$, we have $qd(\omega_t) \leq 3+(k-3) = k$, and so it follows that $C_t \nsim_k C_n$ for all $n > t$.

Hence, $C_m \not\sim_k C_n$ if $m \neq n$ and $\min(m,n) < 2^{k-1} + 3 = g(k)$. This proves the theorem for all $k \geq 3$, and we are done. □

We have proven that $C_m \sim_k C_n$ whenever $m = n$ or $m, n \geq g(k)$. And Theorem 4.8 tells us that $C_m \not\sim_k C_n$ if $m \neq n$ and $\min(m,n) < g(k)$. Hence, we have now derived a complete solution to the Ehrenfeucht-Fraïssé game, when L and R are both cycles.

Theorem 4.9 *Let $g : \mathbb{N} \to \mathbb{N}$ be a function with $g(1) = 1$, $g(2) = 4$, and $g(k) = 2^{k-1} + 3$ for all $k \geq 3$. Then for each positive integer k, the Duplicator wins the game $G_k(C_m, C_n)$ if and only if $m = n$ or $m, n \geq g(k)$.*

5 Conclusions

Our analysis suggests further questions. We have now given a complete solution to the Ehrenfeucht-Fraïssé game $G_n(L, R)$, when L and R are both paths, and when L and R are both cycles. However, we have not yet examined other families of graphs, such as trees, theta graphs, k-colourable graphs, or r-regular graphs. Whether complete solutions can be provided in these cases remains open.

References

[1] S. Arora and R. Fagin, *On winning strategies in Ehrenfeucht-Fraïssé games*, Theor. Comp. Sci, 1774 (1997), 97-121.

[2] J. I. Brown and R. E. Woodrow, *Ehrenfeucht games and graphs*, University of Calgary (Technical Report), 1984.

[3] H. Ebbinghaus and J. Flum. *Finite Model Theory.* Springer, Berlin, 1995.

[4] A. Ehrenfeucht, *An application of games to the completeness problem for formalized theories*, Fundamenta Mathematicae, 49 (1961), 129-141.

[5] R. Fagin, *Easier ways to win logical games*, Proceedings of the DIMACS Workshop on Finite Models and Descriptive Complexity (1996), 3-31.

[6] R. Fagin, *Monadic generalized spectra*, Zeitschrift für Mathematische Logik und Grundlagen der Mathematik, 21 (1975), 89-96.

[7] R. Fraïssé, *Sur quelques classifications des systèmes de relations*, Publications Scientifiques de l'Université d'Alger, Ser. A, 1 (1954), 35-182.

[8] F. Harary, W. Slany, and O. Verbitsky, *On the lengths of symmetry breaking-preserving games on graphs*, Proceedings of the 2002 Algorithmic Combinatorial Game Theory Conference, Dagstuhl, to appear, Theoretical Computer Science.

[9] R. Hoshino, *An analysis of the Ehrenfeucht game on graphs*, M.Sc Thesis, Dalhousie University, 2002.

[10] J. Rosenstein, *Linear Orderings*, Academic Press, New York, 1982.

[11] T. Schwentick, *Graph connectivity and monadic NP*, Proceedings of the 35^{th} IEEE Symposium on Foundations of Computer Science (1994), 614-622.

[12] G. Turan, *On the definability of properties of finite graphs*, Discrete Mathematics, 49 (1984), 291-302.