An Exposition of Ehrenfeucht–Fraïssé Games

Arpon Basu
Undergraduate, Computer Science and Engineering
Indian Institute of Technology Bombay

December 2021

Abstract

We start with an introduction to Ehrenfeucht–Fraïssé Games, establish the Fundamental theorem of EF games, the Methodology theorem, Hanf's theorem, and finally prove the inexpressibility of TWO-COLORABILITY and ACYCLICITY in first-order logic.

Contents

1 Introduction

In this report, we shall seek to study and reason about Ehrenfeucht–Fraïssé Games, which provide us with a game theoretic setting for analyzing logic, which, although not very difficult, is a subtle shift in perspective which makes apparent many theorems in logic that would otherwise be hard to reason about using the semantics of propositional logic alone.
Thus, to that extent, we shall first define Ehrenfeucht–Fraïssé Games, establish its grammar and methodology and slowly build an arsenal of tools, to finally deploy those tools to prove some pretty strong and interesting theorems delineating the expressibility of First Order logic.
Our presentation style will go as follows: First, there will be definitions, followed by lemmata and theorems elucidating the structure defined by those definitions, and finally, a few examples and/or counterexamples and/or applications of those theorems.
The text has been mainly kept self-contained: The only background expected of the reader is that she should be aware of the basic notions in propositional logic (such as quantifiers, predicates, truth tables, etc.), relations and functions (injectivity, surjectivity, bijectivity, equivalence relations) and set theory (power sets and the like). Even these are expanded upon in detail whenever required, so the reader can understand this report with minimal background.

2 Preliminaries

Before beginning our study of the Ehrenfeucht–Fraïssé games, some definitions need to be put for clarity down the road. So there we go:
A **Structure** $\mathcal{A}$ can be defined as a triple $(|\mathcal{A}|, \tau, \mathcal{I})$, where $|\mathcal{A}|$ is the **domain** or **universe** of the structure, τ is it's **vocabulary** and $\mathcal{I}$ is a collection of **Interpretation functions** which "actualize" the vocabulary over the universe.
A **Universe** $|\mathcal{A}|$ in the context of Model theory refers to a **set** over which logical operations can be performed according to the relation and function symbols that accompany it in the vocabulary of the structure it belongs to. We often use the terms (and symbols) "structure" ($\mathcal{A}$) and it's "universe" ($|\mathcal{A}|$) interchangeably. We also sometimes prohibit the inclusion of the empty set as the universe of some structure in our investigations in Model theory, especially while studying first-order logic. However, other than this soft constraint, no other special property is expected in general of the set that represents the universe of some model. In particular, we don't enforce any **field** axioms upon the set.
A **Signature** σ is the triple $(\mathrm{S_f}, \mathrm{S_{rel}}, \mathrm{ar})$ where $\mathrm{S_f}$ and $\mathrm{S_{rel}}$ denote function and relation symbols respectively, such that $\mathrm{S_f} \cap \mathrm{S_{rel}} = \emptyset$. Also, $\mathrm{ar} : \mathrm{S_f} \cup \mathrm{S_{rel}} \longrightarrow \mathbb{N} = \{1, 2, 3, \ldots\}$ is the **arity** function which assigns a non zero natural number to each function and relation symbol, ie:- a function f with arity l is a mapping from $|\mathcal{A}|^l \longrightarrow |\mathcal{A}|$ while a relation R with arity l is a subset of $|\mathcal{A}|^l$.
A **Vocabulary** τ is similar to a signature except that its arity function's co-domain is $\mathbb{N} \cup \{0\}$, ie:- it allows **nullary** functions (functions with arity 0) a.k.a **constant symbols** within its definition.
With these definitions in the bag, we can actually give an example of a structure $\mathcal{A} = (\mathbb{Q}, \{+, \cdot, x \mapsto x^2, \mathbf{0}, \mathbf{1}, \leq\}, \{\mathcal{I}_+, \mathcal{I}_\cdot, \mathcal{I}_\leq\})$, whose signature is $\sigma = \{+, \cdot, x \mapsto x^2, \leq\}$, whose vocabulary is $\tau = \{+, \cdot, x \mapsto x^2, \mathbf{0}, \mathbf{1}, \leq\}$, and whose **interpretation functions** are $\mathcal{I}_+ : \mathbb{Q}^2 \mapsto \mathbb{Q}; +(x, y) = x + y$, $\mathcal{I}_\cdot : \mathbb{Q}^2 \mapsto \mathbb{Q}; \cdot(x, y) = xy$ and $\mathcal{I}_\leq : \mathbb{Q}^2 \mapsto \{\text{True}, \text{False}\}; \leq (x, y) = \text{True}$ iff $x \leq y$ according to the usual order on $\mathbb{Q} \subset \mathbb{R}$. We choose not to define interpretation functions for constant symbols as a matter of convention, although there is no harm in doing so either.
An important note:- When we say that two structures share a common signature and hence inherit the same function symbols, it may not be necessary that their interpretation functions on both universes are identical. Indeed, let $\mathcal{A} := (\mathbb{R}^2, \|\cdot\|, \mathcal{I}_{\|\cdot\|})$ and let $\mathcal{B} := (\mathbb{R}^2, \|\cdot\|, \mathcal{I}_{\|\cdot\|})$ be two structures with identical signatures. However, it could very well be the case that $\mathcal{I}^A_{\|\cdot\|} := \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ while $\mathcal{I}^B_{\|\cdot\|} := |x_1 - x_2| + |y_1 - y_2|$, ie:- the same norm function was actualized differently on the two structures.
A **homomorphism** $h : A \mapsto B$ is a mapping between two sets A and B which have a common relation symbol μ defined

on them such that $\mu_A(a_1, a_2, ..., a_l) \implies \mu_B(h(a_1), h(a_2), ..., h(a_l)) \; \forall (a_1, a_2, ..., a_l) \in A^l$, where l is the arity of μ.
An **embedding** $e : A \mapsto B$ is an injective mapping between two sets A and B which have a common relation symbol μ defined on them such that $\mu_A(a_1, a_2, ..., a_l) \Leftrightarrow \mu_B(e(a_1), e(a_2), ..., e(a_l)) \; \forall (a_1, a_2, ..., a_l) \in A^l$, where l is the arity of μ.
Once again, note that for both homomorphisms and embeddings, the relation μ may be actualized differently on $\mathcal{A}$ and $\mathcal{B}$, hence the different subscripts μ_A and μ_B.
An **isomorphism**, denoted by the symbol $\cong$, is a surjective embedding between two sets.
The **closure of a set** S **w.r.t to a function** $f : S \mapsto S$**, denoted as** $\langle S \rangle^f$ is defined as $\bigcup_{s \in S} \bigcup_{i=0}^{\infty} \{f^i(s)\}$, where $f^0(s) := s$, and $f^{k+1}(s) := f(f^k(s)) \; \forall k \geq 0$.
The definition of closure may also be extended for multi-arity functions $f : S^l \mapsto S^l$ as follows: Let $(s_1, s_2, ..., s_l) \ni S^l \mapsto \mathcal{F}((s_1, s_2, ..., s_l)) = \{s_1, s_2, ..., s_l\} \in \mathcal{P}(S)$ be the "flattening" function, ie:- a function which flattens tuples to sets (while removing repeated elements). Then the closure of S under f is $\bigcup_{s \in S^l} \bigcup_{i=0}^{\infty} \mathcal{F}(f^i(s))$.
A structure $\mathcal{A}$ is an **induced substructure** of $\mathcal{B}$ if $\sigma(\mathcal{A}) = \sigma(\mathcal{B})$, $|\mathcal{A}| \subseteq |\mathcal{B}|$, and the interpretation functions of $\mathcal{A}$ and $\mathcal{B}$ agree on $|\mathcal{A}|$. Furthermore, the closure of $|\mathcal{A}|$ under the function symbols of $\sigma(\mathcal{A})$ yields a **closed induced substructure** of $\mathcal{B}$.

3 An Introduction to Ehrenfeucht–Fraïssé Games

3.1 A Description of the Game

Ehrenfeucht–Fraïssé games are back-and-forth games played between two players "Spoiler" Samson and "Duplicator" Delilah on two universes $\mathcal{A}$ and $\mathcal{B}$ each equipped with the *same* **vocabulary** τ.
The game begins with the corresponding "initial" *configurations* α_0 and β_0 mapping const(τ) to the constant symbols in $|\mathcal{A}|$ and $|\mathcal{B}|$ respectively. At every move in the game, Samson (the Spoiler, responsible for highlighting the difference between $\mathcal{A}, \mathcal{B}$ in terms of their predicate structure) places a pebble, ie:- chooses some element in either $|\mathcal{A}|$ or $|\mathcal{B}|$ following which Delilah (the Duplicator) places her pebble on the other universe, thus extending our *configuration* to the partial functions:

$$\alpha_r : (\mathrm{const}(\tau) \cup \{x_1, x_2, ..., x_r\}) \longrightarrow |\mathcal{A}|$$

$$\beta_r : (\mathrm{const}(\tau) \cup \{x_1, x_2, ..., x_r\}) \longrightarrow |\mathcal{B}|$$

At every stage/move of the game, Delilah has to ensure that $\beta_r \circ \alpha_r^{-1}$ is an isomorphism between $|\mathcal{A}|$ and $|\mathcal{B}|$, ie:- an onto embedding between the sub-structures of $\mathcal{A}$ and $\mathcal{B}$ *induced* as the ranges of the partial functions α_r and β_r, ie:- $\forall$ $0 \leq r \leq m$, Delilah must ensure that $\langle \mathrm{rng}(\alpha_r) \rangle^{\mathcal{A}} \cong \langle \mathrm{rng}(\beta_r) \rangle^{\mathcal{B}}$.
Now, if it so happens that for a k-pebble m-move game (note that k is not necessarily equal to m as we may have pebbles on the board before the game started: Indeed, one may treat constant symbols as pre-placed pebbles on their respective universe) $\mathcal{G}_m^k(\mathcal{A}, \alpha_0, \mathcal{B}, \beta_0)$ Delilah indeed has a winning strategy, we say that $(\mathcal{A}, \alpha_0) \sim_m^k (\mathcal{B}, \beta_0)$. We shall prove below that the relation $\sim_m^k$ is an equivalence relation.
Note that the EF game is a game of perfect information (both Samson and Delilah have complete information about both the structures $\mathcal{A}$ and $\mathcal{B}$), ie:- some player has a winning strategy $\forall 0 \leq k, m$. Thus we can always talk about either Samson or Delilah having a winning strategy. Also note that the external referee judging the game has no information about either structure, and thus it's up-to Samson to highlight the difference between the two. Both players always play optimally.
All this notation for explaining the semantics of the game may be quite intimidating, and thus to clear matters up, we shall provide some examples below.

3.2 A Few Examples

Actually coming to explicit examples, consider the game $\mathcal{G}_3^3$ between $(\mathbb{Z}, <, =)$ and $(\mathbb{Q}, <, =)$. Also assume that there are no constant symbols. We show that Delilah has no winning strategy. Indeed, Samson can first choose $0 \in \mathbb{Z}$, to which say Delilah chooses $q_1 \in \mathbb{Q}$. Then Samson chooses $1 \in \mathbb{Z}$, to which Delilah chooses $q_2 \in \mathbb{Q}$, where $q_2 \neq q_1$ otherwise the equality predicate would be violated. Then Samson chooses $\frac{q_1+q_2}{2} \in \mathbb{Q}$, to which Delilah can't respond since $\nexists z \in \mathbb{Z}$ such that $0 < z < 1$, thus demonstrating to the external referee the difference in the two structures.
Another game could be between $(2\mathbb{Z}, |, =)$ and $(3\mathbb{Z}, |, =)$. Note that for this game, Delilah has a winning strategy $\forall\, 0 \leq k, m$ because no matter what Samson chooses, Delilah can multiply $\frac{2}{3}$ (if Samson chose in $3\mathbb{Z}$) or $\frac{3}{2}$ (if Samson chose in $2\mathbb{Z}$) to Samson's choice and always maintain an isomorphism.
Finally, note that the EF game between $(\mathbb{Q}, <, =)$ and $(\mathbb{R}, <, =)$ can also be won by Delilah $\forall\, 0 \leq k, m$, because both rationals and reals are *dense*, ie:- using only the *less-than* predicate along with the default equality predicate, one can't establish any difference between $\mathbb{Q}$ and $\mathbb{R}$: Indeed, from analysis one knows that at an axiomatic level, what separates $\mathbb{R}$ from $\mathbb{Q}$ is the fact that $\mathbb{R}$ is complete: ie:- every Cauchy sequence in $\mathbb{R}$ converges in $\mathbb{R}$, while the same is not true for $\mathbb{Q}$. Unfortunately, any definition of *completeness* must necessarily involve second order logic, and the fact that the above EF game was not able to separate the two universes is an important foreshadowing to a cornerstone result in the theory of EF games: The equivalence of two structures in terms of satisfying all FO logic statements with k variables and m (nested) quantifiers can be directly determined by the EF game $\mathcal{G}_m^k(\mathcal{A}, \alpha_0, \mathcal{B}, \beta_0)$!

3.3 Some Commentary

First of all, note that we can translate EF games for a large class of universes and their vocabularies into a common problem much more commonly accessed: Suppose that the only symbol in the vocabulary of our universes $\mathcal{A}$ and $\mathcal{B}$ is a binary relation symbol. Then note that the predicate structure of the universe can be expressed as a directed graph, and the EF game basically reduces to (Delilah) finding out the (largest) common induced sub-graphs of the two graphs representing the universe and their predicate structure.
In fact, the entire EF game (on a relational vocabulary) itself can be reduced to finding largest common induced subgraphs of "hypergraph stacks" (hypergraphs are necessary for accommodating relations with arity ≥ 3, and stacks represent the multiple possible relation symbols that can be present in the vocabulary, each warranting its own hypergraph). However, these are harder to visualize about than the vanilla directed graph case.

3.4 $\sim_m^k$ is an equivalence relation

Note that since the initial configuration $u_0 : \text{const}(\tau) \longrightarrow |\mathcal{U}|$ for any universe $\mathcal{U}$ is pre-determined and can't change during the course of the game, we shall simplify our notation slightly and say $\mathcal{A} \sim_m^k \mathcal{B}$ when we want to mean $(\mathcal{A}, \alpha_0) \sim_m^k (\mathcal{B}, \beta_0)$.
Now, we know that a relation is an equivalence relation if and only if it's reflexive, symmetric, and transitive. We shall establish each of this one by one.

3.4.1 Reflexivity of $\sim_m^k$

Note that for any universe $\mathcal{U}$, $\mathcal{U} \sim_m^k \mathcal{U}$ because no matter what element Samson chooses, Delilah can choose the same element from the other copy of the universe and the sub-structures formed thus, are identical and hence isomorphic, at every move.

3.4.2 Symmetry of $\sim_m^k$

If $\mathcal{A} \sim_m^k \mathcal{B}$, then we have that at every step, no matter which element Samson chooses from $|\mathcal{A}| \cup |\mathcal{B}|$, Delilah has a winning move until k pebbles and/or m moves are exhausted, and thus reversing the order of $\mathcal{A}$ and $\mathcal{B}$ makes no difference to the game whatsoever (indeed, no semantic of the game treats the first structure in the relation differently from the second. The winning strategy of Delilah doesn't change, and if $\beta_r \circ \alpha_r^{-1}$ is an isomorphism, then its inverse $\alpha_r \circ \beta_r^{-1}$ is an isomorphism too).

3.4.3 Transitivity of $\sim_m^k$

Let $\mathcal{A} \sim_m^k \mathcal{B}$, and $\mathcal{B} \sim_m^k \mathcal{C}$. We have to show that $\mathcal{A} \sim_m^k \mathcal{C}$.
To that end, we have to demonstrate a winning strategy for Delilah in the k-pebble m-move game between $\mathcal{A}$ and $\mathcal{C}$.
Now, note that in the $(\mathcal{A}, \mathcal{C})$ EF game, Samson can either choose his element either from $|\mathcal{A}|$ or from $|\mathcal{C}|$. If he chooses an element from $|\mathcal{A}|$, then Delilah chooses an element in $|\mathcal{B}|$ according to her winning strategy (which exists as per our hypothesis), and following that she *pretends* that her chosen element in $|\mathcal{B}|$ was actually a Samson move, and finally she chooses another element in $|\mathcal{C}|$ in accordance with her winning strategy for the $(\mathcal{B}, \mathcal{C})$ game, and that element in $|\mathcal{C}|$ is what she plays against Samson in their actual game.
On the other hand if Samson chooses an element in $|\mathcal{C}|$, then Delilah does the exact opposite of what was described above (note that EF games are symmetric as proved in the earlier subsection): She first chooses an element in $|\mathcal{B}|$, pretends that that was a Samson move in the $(\mathcal{B}, \mathcal{A})$ EF game, and then plays her hand by choosing an element in $|\mathcal{A}|$, which is her move for the actual $(\mathcal{A}, \mathcal{C})$ game going on.
Finally, it only remains to prove that the strategy delineated above is indeed valid for the $\mathcal{G}_m^k(\mathcal{A}, \alpha_0, \mathcal{C}, \gamma_0)$ game. To that extent, note that all the while that Delilah has been playing the $(\mathcal{A}, \mathcal{C})$ game, she has also been constructing an auxiliary sub-structure in $\mathcal{B}$ corresponding to the partial functions $\{\beta_r\}_{r\geq 0}$ (along with the already present $\{\alpha_r\}_{r\geq 0}$ and $\{\gamma_r\}_{r\geq 0}$ series) such that $\beta_r \circ \alpha_r^{-1}$ and $\gamma_r \circ \beta_r^{-1}$ are isomorphisms *sharing the same domain*. Since compositions of isomorphisms are also isomorphisms, we have that $(\gamma_r \circ \beta_r^{-1}) \circ (\beta_r \circ \alpha_r^{-1}) \equiv \gamma_r \circ \alpha_r^{-1}$ is an isomorphism too ($\forall\ 0 \leq r \leq m$), thus finishing our proof.

4 The Fundamental Theorem of Ehrenfeucht–Fraïssé games

Before stating the fundamental theorem of Ehrenfeucht–Fraïssé games, we once again need to brush up on some definitions, so here we go.

4.1 Some More Definitions

A **sentence** in FO logic is a logical statement in which all variables are bound to a quantifier, ie:- there are no "free" variables.
A **theory** $\Gamma \subseteq \mathcal{L}(\tau)$ is a set of sentences.
The **quantifier rank** of a sentence is defined to be the *maximum nesting level of quantifiers* in that sentence.
For example, $\forall x \forall c \forall \epsilon \exists \delta\ (\epsilon > 0) \wedge (\delta > 0) \wedge (0 < |x - c| < \delta) \implies (|f(x) - f(c)| < \epsilon)$ is a sentence of quantifier rank 4, stating that the function f is continuous everywhere, with the domain of discourse being $\mathbb{R}$.
Let $\mathcal{L}_m^k(\tau)$ be a language in first-order logic with exactly k variables and quantifier rank at most m. Also, let $\mathcal{L}_\omega^\omega(\tau)$ denote the set of *all* **finite** first order sentences on τ.
Two structures $\mathcal{A}$ and $\mathcal{B}$ with the same vocabulary τ are called $k - m$ equivalent if they agree and disagree upon the same sentences in $\mathcal{L}_m^k$ constructed with symbols from τ, ie:- $\forall \varphi \in \mathcal{L}_m^k(\tau)\ \mathcal{A} \models \varphi \Leftrightarrow \mathcal{B} \models \varphi$, and this equivalence is represented as $\mathcal{A} \equiv_m^k \mathcal{B}$.

4.2 A Crucial Lemma

Having completed the above definitions, we now proceed to prove a crucial lemma, mainly to establish that for "nice" enough structures, the number of formulae it can generate is finite. This may not seem like much, but it's necessary to ensure that we can take conjunctions and disjunctions over the entire set of sentences later on, which would be ill-defined if the underlying set weren't finite.
The method of proof is more or less enumerative in the sense that it explicitly shows which predicates *can* be generated using the predicates in our vocabulary.

Lemma 1. *Let τ be a finite relational vocabulary of a finite structure, and let $\mathcal{L}_r$ be the set of all FO sentences with symbols from τ. Then the number of inequivalent formulae in $\mathcal{L}_r$ is finite.*

The inductive proof. We first show that the number of inequivalent formulae in $\mathcal{L}_0$ (the set of formulae with no quantifiers, only variables and predicates) is finite. How do we do that? For any l-ary relation $R \in \mathrm{S}_{\mathrm{rel}}$, consider the outputs of all possible tuples in $|\mathcal{A}|^l$ when evaluated by R. If k is the cardinality of $|\mathcal{A}|$ (ie:- we are working in $\mathcal{L}_0^k$), then we have k^l such **propositions**, each of which is either true or false. Let $n = \sum_{R \in \mathrm{S}_{\mathrm{rel}}} k^{\mathrm{ar}(R)}$. Since we have n propositions with us, the truth value of each of which is independent of others, we have 2^n possible truth-value "states", ie:- if our propositions were, say p_1, p_2, p_3 and p_4, then the truth state 1011 would be represented by $p_1 \wedge \neg p_2 \wedge p_3 \wedge p_4$ (This sentence is also known as a **complete sentence** since it is satisfied for a unique assignment of truth values). Now, any formula in $\mathcal{L}_0^k$ can be uniquely expressed as a **Disjunctive Normal Form (DNF)**, ie:- an OR of ANDs, where the "AND"s in turn refer to any one of our truth-states. Since a DNF takes an OR of only a subset of the 2^n truth-states, the total number of DNFs is atmost the number of subsets of the set of our truth-states, ie:- the total number of formulae (which obviously upper bound the number of inequivalent formulae) is upper bounded by 2^{2^n}, where $n = \sum_{R \in \mathrm{S}_{\mathrm{rel}}} k^{\mathrm{ar}(R)}$, and thus the total number of inequivalent formulae in $\mathcal{L}_0^k$ is finite $\forall k \in \mathbb{N}$.
Now, for the inductive step, assume that the number of inequivalent formulae in $\mathcal{L}_m^k$ is finite for some $m \geq 0$. Consider $\mathcal{L}_{m+1}^k$, ie:- an additional quantifier is nested upon the sentences in $\mathcal{L}_m^k$. Consider any of the propositions in the sentences in $\mathcal{L}_m^k$, as described earlier. Attach the extra quantifier in $\mathcal{L}_{m+1}^k$ to this proposition to make a new one. Since there are finitely many propositions in $\mathcal{L}_m^k$, we can only generate finitely many new ones out of them (to be precise, if there are P propositions, then we can generate $2P$ propositions out of them by attaching the quantifiers $\forall$ and $\exists$), and since there are finitely many inequivalent formulae that can be generated from finitely many propositions, we are done, ie:- $\mathcal{L}_{m+1}^k$ has finitely many inequivalent formulae.
Thus, by induction, $\mathcal{L}_r$, actualized over a finite relational vocabulary, has finitely many inequivalent formulae $\forall r \geq 0$. Hence proved. □

4.3 The Fundamental theorem of Ehrenfeucht–Fraïssé games

Theorem 1. *Let $\mathcal{A}$ and $\mathcal{B}$ be two structures with a common, finite, relational vocabulary τ. Then*

$$(\mathcal{A} \sim_m^k \mathcal{B}) \Leftrightarrow (\mathcal{A} \equiv_m^k \mathcal{B})$$

The proof given below establishes the theorem by cleverly exploiting what is known as *quantifier elimination* to reduce complicated sentences down inductively, thus helping show a winning strategy for one of the players.

A brief sketch of the proof. For both parts of the iff condition we proceed by induction.
For the first part of the iff condition $((\mathcal{A} \sim_m^k \mathcal{B} \implies \mathcal{A} \equiv_m^k \mathcal{B}) \Leftrightarrow (\mathcal{A} \not\equiv_m^k \mathcal{B} \implies \mathcal{A} \nsim_m^k \mathcal{B}))$, suppose $\mathcal{A} \not\equiv_{m+1}^k \mathcal{B}$. Then note that if φ is the sentence in $\mathcal{L}_{m+1}^k$ over which $\mathcal{A}$ and $\mathcal{B}$ disagree, then run the following algorithm on φ: Replace all $\forall p$ quantifiers in φ by $\neg\exists\neg p$, and all $p \vee q$ disjunctions by $\neg(\neg p \wedge \neg q)$. Then φ will reduce down to $\alpha \wedge \beta$ and/or $\neg\alpha$. If

$\mathcal{A}$ and $\mathcal{B}$ disagree on $\alpha \wedge \beta$, then they must disagree on either α or β, and if $\mathcal{A}$ and $\mathcal{B}$ disagree on $\neg\alpha$, then they disagree on α too. In essence we stripped down the "parse tree" of φ (the maximum depth of which was $m+1$, but it could have had branches of other, smaller depths too) to a single branch emanating from the root, and moreover the root contains the $\exists$ quantifier now by our algorithm. Thus, α, the pruned version of φ, is in fact $\exists x_i \delta$, where $\delta \in \mathcal{L}_m^k$. Thus $\mathcal{A}$ and $\mathcal{B}$ disagree on $\delta \in \mathcal{L}_m^k$ for some value z of x_i for which δ is true in $\exists x_i \delta$. Also, by the induction hypothesis, Samson has a winning strategy for the $\mathcal{G}_m^k$ game. Now, just extend that $\mathcal{G}_m^k$ game, by taking Samson's first move to be z. Then WLOG $(\mathcal{A}, z) \models \delta$ while $(\mathcal{B}, b) \not\models \delta \; \forall b$ in the choosing space of Delilah. Samson can now finish the game by his earlier strategy since $\mathcal{A} \not\equiv_m^k \mathcal{B}$ as they disagree on δ (Note that we have been slightly loose with the notation in $(\mathcal{A}, z)$, whereas it actually should be $(\mathcal{A}, \alpha_1)$, where α_1 is the *configuration* generated by the selection of z). As for the reverse direction, let the $k-(m+1)$ game begin, and let Samson choose anything to define the first *configuration* α_1. Let ϕ be the set of all inequivalent sentences satisfied by $(\mathcal{A}, \alpha_1)$ and let $\Phi := \wedge_{\varphi \in \phi} \varphi$ (this conjunction relied on the fact that $|\phi| < \infty$, as proved earlier, because infinite conjunctions are ill-defined). Since $(\mathcal{A}, \alpha_0) \models (\exists x_i \Phi) \in \mathcal{L}_{m+1}^k$ (with the witness being the pebble Samson chose) and since $\mathcal{A} \equiv_{m+1}^k \mathcal{B}$, we have $(\mathcal{B}, \beta_0) \models (\exists x_i \Phi)$. Choose the witness x_i for Φ in the Delilah choosing space, and let Delilah play the move of the witness. Then the (α_1, β_1) configuration is in the $\mathcal{L}_m^k$ domain, and by induction, Delilah has a winning strategy for the $\mathcal{G}_m^k$ game, thus proving the reverse direction too.
Hence proved. □

Phew! So now we have a very nice looking theorem which says that the equivalence of two structures over a set of sentences, very hard to check or reason about, can be suitably simulated by a game! However, what looks too good to be true probably is, and we shall take some wind out of our sails by following up on the proof of this theorem by an example where the the theorem fails :)!

4.3.1 Counterexample for an infinite relational vocabulary

Let $|\mathcal{A}| = \mathcal{P}(\mathbb{N})$, and let $|\mathcal{B}|$ be the set of all finite subsets of $\mathbb{N}$. Let $\tau = \langle R_1^1, R_2^1, \ldots \rangle$ where $R_i^1(S) :=$ True iff $i \in S$.
Then note that $\mathcal{A} \nsim_k^k \mathcal{B} \; \forall k \in \mathbb{N}$ (note that since there are no constant symbols, $k = m$) as Samson can simply choose $\mathbb{N} \in |\mathcal{A}|$ in his first move, and keep choosing arbitrary members of $|\mathcal{A}|$ in his subsequent moves, forcing Delilah to choose finite subsets of $\mathbb{N}$ as her moves. Finally, at the end of k rounds, since all of Delilah's choices are finite sets, $\exists z \in \mathbb{N}$ such that z doesn't belong to any set that Delilah has chosen. Thus while one member ($\mathbb{N}$) of Samson's sub-structure satisfies the R_z^1 predicate, no member of Delilah's sub-structure does so, thus forbidding any isomorphism, leading to a defeat for Delilah.
However, $\mathcal{A} \equiv_m^k \mathcal{B} \; \forall k, m \geq 0$. Assume for the sake of contradiction that doesn't happen, ie:- for some k_0, m_0 $\mathcal{A} \not\equiv_{m_0}^{k_0} \mathcal{B}$. Then by the algorithm described in the previous proof, $\mathcal{A}$ and $\mathcal{B}$ disagree upon some sentence φ of the format $\exists S_i \delta$. Now, note that all our predicates are "membership" predicates, ie:- their conjunction and disjunction describe sets with certain properties. Therefore, it can't happen that $\mathcal{A}$ is unable to satisfy some sentence while $\mathcal{B}$ is, since $|\mathcal{B}| \subseteq |\mathcal{A}|$. Thus, we have $\mathcal{A} \models \delta$, while $\mathcal{B} \not\models \delta$. Strip away all quantifiers from φ like this, until we have a $k_1 \leq k_0$ variable sentence with no quantifiers. Since that sentence has a finite number of predicates, $\exists z \in \mathbb{N}$ such that R_z^1 is the highest predicate (predicate with highest subscript) present in the sentence. Now, since $\mathcal{A}$ and $\mathcal{B}$ disagree over it, it must follow that this sentence describes atleast one infinite set $S \subseteq \mathbb{N}$ (because if all sets described were finite, then $\mathcal{B}$ would've been indistinguishable from $\mathcal{A}$ w.r.t that sentence). Since S is infinite, $\exists z_0 \in S$ such that $z_0 > z$, but there was no predicate in our sentence which talked about the inclusion or exclusion of z_0, leading to a contradiction.

4.3.2 A saving grace

We saw above how easily the delicate chain of arguments needed to establish the fundamental theorem of Ehrenfeucht–Fraïssé games unravels with even a slight relaxation on the types of vocabularies permitted. However, while infinitely many

relations are quite hard to circumvent, functional symbols may in fact be introduced in the vocabulary, subjected to some constraints inspired by the counterexample above. Note that in both of the counterexamples above, the the fundamental theorem of Ehrenfeucht–Fraïssé games fails due to the failure of the crucial lemma which posits that sets of sentences we are talking about must have finitely many inequivalent sentences.
Thus, the condition we impose is: Let τ be a finite vocabulary which may contain function symbols and let $\forall$ finite $S \subseteq |\mathcal{A}|$, it must be the case that $|\langle \mathrm{S} \rangle^{\sigma(\mathcal{A})}| < \infty$.
Then the number of inequivalent formulae in $\mathcal{L}_m^k(\tau)$ is finite, $\forall k, m \geq 0$.
How do we go about showing this? Consider any $f \in \mathrm{S_f} \in \sigma(\mathcal{A})$, where the f is assumed to be unary (otherwise composing f with itself would be an issue since the co-domain of f is S. However, our proof goes through in the same way if f is a l-ary function that outputs an element in S^l itself). Then note that the number of distinct (inequivalent) functions in $\{f, f^2, f^3, \ldots\}$ can be at most the number of distinct mappings from S to S, which is N^N, where $N = |\langle \mathrm{S} \rangle^{\sigma(\mathcal{A})}|$ (if f is l-ary, then the upper bound is $(N^l)^{N^l}$). Thus $\exists K \in \mathbb{N}$ such that $\forall k \geq K$ $f^k \cong f^{k_0}$ for some $k_0 < K$.
Now that we have bounded the maximum composition of f in our sentences, for each distinct $g \equiv f^k$, **Skolemize** the function g in the following way: If $g : M \mapsto M$ (where $M = S^l$ for some $l \in \mathbb{N}$), then construct the predicate $R_g \subseteq M \times M$ such that $R_g(m_1, m_2) :=$ True iff $m_2 = g(m_1)$.
Thus, since we have finitely many different functions, we have finitely many (generated) relations too, and thus our finite functional vocabulary τ is *transformed* into an equivalent finite relational vocabulary $\tilde{\tau}$, for which the crucial lemma (that the number of inequivalent formulae in $\mathcal{L}_m^k(\tau) \equiv \mathcal{L}_m^k(\tilde{\tau})$ is finite) and consequently, the fundamental theorem of Ehrenfeucht–Fraïssé games, does hold.

5 Methodology theorem, Hanf's Theorem

Note that when we established the equivalence between EF games and the expressibility powers of finite FO theories, we intended to use it as an abstraction between proving the equivalence of two theories and simulating a game, because the former was deemed to be hard to reason about in comparison to the latter.
But also note that the fundamental theorem of EF games still requires us to explicitly establish the existence of winning strategies for Delilah over all possible "trajectories" the game could take, which while in some sense easier than reasoning about first-order sentences, is still a difficult task in many cases.
We shall thus seek to establish some more mathematical machinery to further abstract away Delilah's winning strategies into something equivalent, but easier to reason about the entity. To that extent, we first state the Methodology theorem, build up some more definitions, and then come to Hanf's theorem, which quantitatively pins down the expressive power of first-order logic.
So without further ado, let's begin.

5.1 The Methodology Theorem

Theorem 2. *Let C be a (possibly infinite) set of structures sharing a common finite relational vocabulary τ. A **boolean query** $\mathcal{S} \subseteq C$ can NOT be described with $\mathcal{L}_r$ (which is the set of sentences with a nest of at most r quantifiers) iff $\exists$ a sequence of structures $\{\mathcal{A}_r\}_{r \geq 1}$, $\{\mathcal{B}_r\}_{r \geq 1} \in C$ such that $\mathcal{A}_r \in \mathcal{S} \not\ni \mathcal{B}_r$ but $\mathcal{A}_r \sim_r \mathcal{B}_r$ $\forall r \geq 1$.*

A brief sketch of the proof. If there indeed exists a sequence of structures $\{\mathcal{A}_r\}_{r \geq 1}$, $\{\mathcal{B}_r\}_{r \geq 1} \in C$ such that $\mathcal{A}_r \in \mathcal{S} \not\ni \mathcal{B}_r$ but $\mathcal{A}_r \sim_r \mathcal{B}_r$ $\forall r \geq 1$, then we have that $\mathcal{A}_r$ and $\mathcal{B}_r$ agree on every statement in $\mathcal{L}_r$ yet disagree vis-a-vis their belonging in $\mathcal{S}$, thus showing that the set $\mathcal{S}$ can't be constructed using statements in first-order logic (with at most r nested quantifiers) alone.
As for the reverse direction, we recall the notion of a complete sentence from Section 4.2 and let our propositions be the truth values of relations (in $\mathrm{S_r}(\tau)$) at every point in their domain (as we did while proving the finiteness of $\mathcal{L}_0^k$). Let

Φ be the set of all complete sentences which can be constructed from the above propositions. Then note that for any $\mathcal{D} \in C$ there exists a unique sentence $\phi \in \Phi$ such that $\mathcal{D} \models \phi$ because if (for the sake of contradiction) there existed ϕ, ϕ' such that $(\mathcal{D} \models \phi) \wedge (\mathcal{D} \models \phi')$ then we could simply choose an arbitrary formula φ which assumes the value "True" on the truth state represented by ϕ and "False" on the truth state represented by ϕ', leading to the conclusion $(\mathcal{D} \models \phi \vdash \varphi) \wedge (\mathcal{D} \models \phi' \vdash \neg\varphi) \implies (\mathcal{D} \models \varphi) \wedge (\mathcal{D} \models \neg\varphi)$, which is absurd.
Now, if for any $\mathcal{A}, \mathcal{B} \in C$ such that $\mathcal{A} \in \mathcal{S} \not\ni \mathcal{B}$ it so happens that $\mathcal{A} \models \varphi$ and $\mathcal{B} \models \varphi$ for some complete sentence φ, then it essentially implies $\mathcal{A} = \mathcal{B}$ because a complete sentence encodes within itself complete information regarding every relation in τ, and for two structures to have identical "marker" complete sentences would imply that they are equal, which again leads to a contradiction since $\mathcal{A}$ and $\mathcal{B}$ belong to different (disjoint) sets.
Thus, we have that every structure in $\mathcal{S}$ satisfies some sentence in a certain subset $\{\varphi_1, \varphi_2, ..., \varphi_s\} \subseteq \Phi$, while every structure in $C \setminus \mathcal{S}$ satisfies some sentence in $\Phi \setminus \{\varphi_1, \varphi_2, \ldots, \varphi_s\}$, ie:- $\mathcal{A}_r \nsim_r \mathcal{B}_r \ \forall r \geq 1$ since by the fundamental theorem of EF games

$$\mathcal{A}_r \sim_r \mathcal{B}_r \Leftrightarrow \mathcal{A}_r \equiv_r \mathcal{B}_r \implies \exists \varphi \in \mathcal{L}_r (\mathcal{A}_r \models \varphi) \wedge (\mathcal{B}_r \models \varphi)$$

Thus the condition $\mathcal{A}_r \sim_r \mathcal{B}_r$ is violated. But also note that $\Theta := \vee_{i=1}^{s} \varphi_i$ (this disjunction is well defined because it's finite, and it's finite because $|\Phi| < \infty$, as proved in Section 4.2) can describe $\mathcal{S}$ (as every structure in $\mathcal{S}$ has a unique "marker" complete sentence which is separate from the markers of structures in $C \setminus \mathcal{S}$), contradicting the fact that $\mathcal{S}$ was FO indescribable.
Hence proved. □

5.2 Gaifman Graphs and Local Isomorphisms

So now that we have established the Methodology theorem, we can simply establish a series of structures in which a strategy for Delilah has to be demonstrated to capture FO indescribability. However, the task of abstracting away even the winning strategies remains. To that extent, follow some definitions which will help us state **Hanf's theorem** in the next subsection.
Let τ be a finite vocabulary consisting only of related symbols and constants. The **Gaifman** graph of a structure $\mathcal{A}$ on the vocabulary τ is defined as $G_{\mathcal{A}}(|\mathcal{A}|, E_{\mathcal{A}})$, where

$$E_{\mathcal{A}} := \{\{a, b\} : a, b \in |\mathcal{A}|, \exists R \in \mathrm{S_r}(\tau), \exists 1 \leq i \neq j \leq \mathrm{ar}(R), R(c_1, c_2, ...) = \mathrm{True}, c_i = a; c_j = b\}$$

Note that the Gaifman graph is simple and undirected.
A structure $\mathcal{A}$ is termed **bounded-degree** if the degree of each vertex in its Gaifman graph is finite. Note that the Gaifman graph itself may be infinite, though.
Let $\mathbb{N}^* := \mathbb{N} \cup \{0, \omega\}$, where ω is the **ordinal** which is greater than every natural number, ie:- $\omega > n \forall n \in \mathbb{N}$. In simple words, "$\omega = \infty$".
On the Gaifman graph, there exists a natural notion of **distance** between two elements of $|\mathcal{A}|$ as being the **length of the shortest path** between those elements traversing along the edges of the Gaifman graph. To that extent, we define the function **dist**: $|\mathcal{A}|^2 \mapsto \mathbb{N}^*$ which signifies the distance between two elements of $|\mathcal{A}|$, and satisfies all the usual distance properties of being symmetric, positive definite and satisfying triangle's inequality. We assume the convention that if two elements a, b aren't connected in G then $\mathrm{dist}(a, b) = \omega$.
Now, for any *configuration* $(\mathcal{A}, \alpha_r)$, consider the closed induced substructure $\mathcal{A}_{\mathrm{sub}}$ of $\mathcal{A}$ in the EF game. Then $\forall a \in |\mathcal{A}|$, define

$$N(a, d) := \{b : b \in \mathcal{A}; \mathrm{dist}(a, b) \leq d\} \cap |\mathcal{A}_{\mathrm{sub}}|$$

Thus, for any given $d \in \mathbb{N}^*$ and for all $a \in |\mathcal{A}|$, we have the structure $N(a, d)$ associated with a. Since all of these structures have the same vocabulary (which is the vocabulary of $\mathcal{A}$ itself), it makes sense to talk about **isomorphisms** in between these structures, ie:- we can consider isomorphisms b/w the "d-balls" of every element in $|\mathcal{A}|$, **and those isomorphisms partition $|\mathcal{A}|$ into equivalence classes**. To simplify matters, here is what is going on: For every element,

we consider the ball of radius d around it (which is basically comprised of all elements close enough to a on G), **filter out elements** (ie:- remove those elements) from that ball **not belonging to** $\mathcal{A}_{\text{sub}}$, and consider isomorphisms (note that these isomorphisms can be thought of as "partial automorphisms") between different balls with different "centers". Since isomorphism is an equivalence relation, we end up partitioning $|\mathcal{A}|$, and each equivalence class in this partition denotes those set of elements whose **filtered** d-balls are isomorphic to each other.
Thus, for any element in $|\mathcal{A}|$ and any given $d \in \mathbb{N}^*$, we can define it's *d-type* to be the equivalence class of the partition of $|\mathcal{A}|$ it belongs to.
We thus call the partition of $|\mathcal{A}|$ w.r.t filtered d-balls, and w.r.t the configuration $(\mathcal{A}, \alpha_r)$, to be the d-partition of $\mathcal{A}$.
Two sets X and Y are called **equinumerous** if there exists a bijection between the two sets.
Two non-empty sets X and Y are called **t-equipollent** if either they are equinumerous, or if the cardinalities of both the sets is greater than $t \in \mathbb{N}$, ie:- there exists a non-bijective surjection from both X and Y to the set $\{1, 2, ..., t\}$. We also extend the definition of equipollence to $t = 0$ by defining any two non-empty sets to be 0-equipollent.
A pair of structures $(\mathcal{A}, \mathcal{B})$ is called *d-nice* (w.r.t configurations $(\mathcal{A}, \alpha_r)$ and $(\mathcal{B}, \beta_r)$) if for every equivalence class in the d-partition of $\mathcal{A}$, there exists a unique equinumerous equivalence class in the d-partition of $\mathcal{B}$.
A pair of structures $(\mathcal{A}, \mathcal{B})$ is called (d, s)*-nice* (w.r.t configurations $(\mathcal{A}, \alpha_r)$ and $(\mathcal{B}, \beta_r)$) if for every equivalence class in the d-partition of $\mathcal{A}$, there exists a unique s-equipollent equivalence class in the d-partition of $\mathcal{B}$.

5.3 Hanf's theorem

There are two versions of Hanf's theorem. We present both of them below.

Theorem 3. *If $(\mathcal{A}, \mathcal{B})$ is a* 2^r-nice *pair of structures (w.r.t the configurations α_0, β_0) sharing a common vocabulary, then $\mathcal{A} \equiv_r \mathcal{B}$.*

The second version is also known as the bounded degree Hanf's theorem.

Theorem 4. *If $(\mathcal{A}, \mathcal{B})$ is a pair of bounded degree structures sharing a common vocabulary, then $\exists s > 0$ such that if $(\mathcal{A}, \mathcal{B})$ are a* $(2^r, s)$-nice *pair (w.r.t the configurations α_0, β_0), then $\mathcal{A} \equiv_r \mathcal{B}$.*

An outline of the Proof. For any element v which Samson chooses, let V be it's equivalence class. Since Delilah has an isomorphism h for configuration "m" (by an inductive hypothesis mentioned ahead), let $V^h := h(V)$ be the counterpart of V in the other universe. Choose any element $v' \in V^h$, set $s_m := 1 + (r - m)d^{2^r}$ and then it can be shown that for the extended configuration "$m + 1$", $(\mathcal{A}, \mathcal{B})$ remain $(2^{r-(m+1)}, s_{m+1})$-*nice* structures (so the inductive hypothesis basically was that Delilah's choices maintain the invariant that at the m^{th} configuration, $(\mathcal{A}, \mathcal{B})$ remain $(2^{r-m}, s_m)$-*nice* structures w.r.t configurations α_m, β_m. The base case is the premise of the theorem itself), and thus at the end of r moves, $(\mathcal{A}, \mathcal{B})$ are $(1, 1)$-*nice* structures. But note that $(1, 1)$-*nice*ness is basically the definition of isomorphism itself, and thus $\exists$ an isomorphism $(\mathcal{A}, \alpha_r) \cong (\mathcal{B}, \beta_r)$ which implies that Delilah has won (ie:- her strategy works), because the isomorphism of $(\mathcal{A}, \alpha_r)$ and $(\mathcal{B}, \beta_r)$ implies that all the previous configurations had an isomorphism too, as the previous configurations were all subsets of $(\mathcal{A}, \alpha_r)$ and $(\mathcal{B}, \beta_r)$. □

However, much more important than the proof is how we can now exploit the theorem to actually do what we set out to do: Describe the FO expressibility of queries, and how do we plan to do that? **To show some query is not FO expressible, or to establish the minimum quantifier rank of the query, we *construct* a sequence of structures according to the hypothesis of the Methodology theorem and then establish $\sim_r \Leftrightarrow \equiv_r$ by Hanf's theorem so that all the hypotheses of the Methodology theorem are satisfied upon which we obtain the non-expressibility of S in $\mathcal{L}_r$.**
So let's go!

6 Examples

We shall discuss a lot of examples below to illustrate the power of the machinery we developed above. Note that when we describe queries on graphs, it's assumed that the only relation in our finite relational vocabulary is a binary one, namely the relation which is true only when its two arguments are joined with an edge to each other. If the graph is undirected, that relation is assumed to be symmetric.
Also note that for structures that are undirected graphs, the Gaifman graph of the structure is the structure itself.
We also define $\mathcal{G}_u$ to be the set of all undirected graphs, $\mathcal{G}_d$ to be the set of all directed graphs and $\mathcal{G} := \mathcal{G}_u \cup \mathcal{G}_d$. For any set of variables $\{x_1, x_2, ..., x_k\}$, we abbreviate the fact that all the variables are distinct by the predicate distinct$(x_1, x_2, ..., x_k)$. Note that this is just the conjunction of $\binom{k}{2}$ non-equality clauses. We shall also relax the definition of $\mathcal{L}^k_m$ to include all FO statements with at most k variables and at most m nested quantifiers, as opposed to exactly k variables and m nested quantifiers.
Finally, when we say that some query requires at most k variables (or at most m quantifiers) to encode, we implicitly assume that the encoding is being done with the maximum possible efficiency, ie:- the number of variables (and/or quantifiers) in that logical statement can't be reduced further, implying that we are not "wasting" any variable/quantifier.

6.1 Finite Relational Unordered Vocabulary Queries

6.1.1 Expressibility of CLIQUE(k)

Let CLIQUE(k)$\subseteq \mathcal{G}_u$ be the set of all undirected graphs which have the complete graph K_k as their subgraph.

Proposition 1. *CLIQUE(k) $\in \mathcal{L}^k \setminus \mathcal{L}^{k-1}$, ie:- CLIQUE(k) needs atleast k variables to describe, and k variables suffice.*

Proof. Firstly, note that CLIQUE(k) is described by the following sentence

$$(\exists x_1, x_2, ..., x_k)(\text{distinct}(x_1, x_2, ..., x_k) \wedge E(x_1, x_2) \wedge E(x_1, x_3)... \wedge E(x_1, x_k)... \wedge E(x_{k-1}, x_k))$$

Thus CLIQUE(k) $\in \mathcal{L}^k$. But CLIQUE(k) $\notin \mathcal{L}^{k-1}$ since $K_k \sim^{k-1} K_{k-1}$ (indeed no matter what either player does in this game, the two sub-graphs, in the end, will always be identical if we have at most $k-1$ pairs of pebbles to spare).
Hence proved. □

6.1.2 P(n)

Let P be a boolean query on $\mathbb{N}$ such that both P and $\mathbb{N} \setminus P$ is infinite. Thus, for example, P can be the set of odd numbers, the set of primes, the set of numbers divisible by 5, and so on.
Let P(n)$\subseteq \mathcal{G}$ be the set of all graphs whose number of vertices 'n' satisfies P.

Proposition 2. $\exists\, n \in \mathbb{N}$ *P(n)* $\notin \mathcal{L}^n$.

Proof. Let G_k be the graph with k vertices, each vertex having only a self-loop. Then once again note that $G_n \sim^n G_{n+1}$ (n is an *arbitrary* natural number) as no matter what moves either player plays, with at most n pairs of pebbles, the substructures formed will be identical. Hence $G_n \equiv^n G_{n+1}$, which implies that if there did exist a sentence in $\mathcal{L}^n$ which could describe P(n), then it would have to describe P($n+1$) too. Thus, $P(n) \in \mathcal{L}^n \implies P(n+1) \in \mathcal{L}^n$. Hence if $\forall n \in \mathbb{N} P(n) \in \mathcal{L}^n$, then either $P(1) \notin \mathcal{L}^1$ (in which case we are done) or $P(n) \in \mathcal{L}^1 \forall n \in \mathbb{N}$ by induction. But note that with only one variable, we can talk about at most one vertex of our graph, and since there are arbitrarily large m, n such that $P(m)$ holds and $\neg P(n)$ holds, one variable sentence can't capture completely the properties of $P(n)$ for all natural numbers, leading to a contradiction.
Thus $\neg(\forall n \in \mathbb{N}$ P$(n) \in \mathcal{L}^n) \equiv \exists\, n \in \mathbb{N}$ P$(n) \notin \mathcal{L}^n$ holds.
Hence proved. □

6.1.3 PATH$_k(x, y)$

Let PATH$_k(x, y)$ be the set of all graphs having the (constant) vertices "x" and "y" such that $\exists$ a path of length at most $\mathbf{2^k}$ between x and y.

Proposition 3. *For all $k \geq 1$ we have* $\text{PATH}_k(x, y) \in \mathcal{L}_k^3 \setminus \mathcal{L}_{k-1}^2$.

Proof. Note that
$\text{PATH}_0(x, y) = (x = y) \vee E(x, y)$
$\text{PATH}_{k+1}(x, y) = \exists z\ (\text{PATH}_k(x, z)) \wedge (\text{PATH}_k(z, y))$
Note that even though we introduced a new variable z in our inductive step, we can still make do with exactly 3 variables x, y, and z. How? When $\text{PATH}_k(x, z)$ would be further expanded, we can re-use y as the middle variable, and similarly for $\text{PATH}_k(z, y)$ we can re-use x. This does not lead to conflicting variable descriptions because the parentheses in our expansion ensure appropriate scope resolution. Thus $\text{PATH}_k(x, y) \in \mathcal{L}_k^3$.
Also note that $\text{PATH}_k(x, y) \notin \mathcal{L}^2$. Why? Consider two graphs G and H where G is the disjoint union of two 3-cycles, while H is a 6-cycle. Furthermore consider 3 unary relations in the vocabularies of G and H (consider these as colorings) such that both the 3-cycles in G have all 3 colors, while the color in H is "RYBRYB".
Then it's easy to show that $G \sim^2 H$ (Note that $0 \leq m \leq 2$, and it's easy to see for all 3 values of m, Delilah has a winning strategy. We can just analyze all cases, and the coloring has been chosen to help Delilah decide a winning strategy), but there doesn't exist a path between the chosen vertices in G, while there does exist a path between the chosen vertices in H. Thus if there existed $k \geq 1$ such that $\text{PATH}_k(x, y) \in \mathcal{L}^2$, then G and H would differ on $\forall\, x, y\ \text{PATH}_k(x, y)$, contradicting the fundamental theorem of EF games. In fact, not only does this counter-example prove the fact that PATH is atleast a 3-variable property, it also shows that connectivity is also atleast a 3 variable property too! More on this later.
Finally, if we show that $\text{PATH}_k(x, y) \notin \mathcal{L}_{k-1}$, then we are done since $\mathcal{L}_{k-1}^2 = \mathcal{L}^2 \cap \mathcal{L}_{k-1}$. To that extent, we show a series of graphs G_k and H_k, $k \geq 1$, such that Delilah has a winning strategy in $\mathcal{G}^{k-1}(G_k, H_k)$. Assume G_k and H_k are both path graphs with *atleast* $1 + 2^k$ vertices (note that $|V(G_k)| \neq |V(H_k)|$ in general), and let the endpoints of those path graphs be marked with constants "0" and "max".
Then for $k = 1$, the initial configuration consists of a path graph with 3 vertices, and since the pebbles on the constants "0" and "max" aren't connected by an edge in both the graphs, Delilah already wins the 0-move game. Now, suppose Delilah has a strategy for the k-move game. Consider the $(k + 1)$-move game. Suppose Samson chooses some vertex. Then let Delilah choose the same vertex in the other graph. If the vertex they chose is $\leq 1 + 2^{k-1}$ from some end, then note that if Samson chooses something from the shorter piece, then the game basically reduces to two path graphs of the same length, which are identical. Delilah's winning strategy then is trivial. And if Samson chooses something from the larger part, then note that the size of the larger part is $\geq 1 + 2^{k-1}$, and thus Delilah inductively has a winning strategy. And finally, if Samson's choice was at $\geq 1 + 2^{k-1}$ distance from both ends, then after her copycat move, Delilah once again wins with her inductive winning strategy. Thus, $\text{PATH}_k(x, y) \notin \mathcal{L}_{k-1}$, and tying everything up, we get $\text{PATH}_k(x, y) \in \mathcal{L}_k^3 \setminus \mathcal{L}_{k-1}^2$.
Hence proved. □

6.1.4 CONNECTIVITY(n)

Let CONNECTIVITY(n)$\subseteq \mathcal{G}_u$ be the set of all undirected connected graphs. Let $n \geq 5$.

Proposition 4. CONNECTIVITY*(n)* $\in \mathcal{L}^3_{\lceil \log_2(n-1) \rceil + 2} \setminus \mathcal{L}^2_{\lceil \log_2(n-2) \rceil - 2}$.

Proof. Consider the sentence $\varphi := \forall\, x, y\ (\text{PATH}(x, y) \wedge (E(x, y) \Leftrightarrow E(y, x)))$.
Observe that the maximum path length between two vertices in a connected n-vertex graph is $n - 1$. From Section 6.1.3, we know that the encoding of the fact that there is a path of length at most 2^l between any two given vertices needs ex-

actly l nested quantifiers. Thus PATH(x, y) would need exactly $\lceil \log(n-1) \rceil$ nested quantifiers for the vertices furthest apart, and then we add two more quantifiers for iterating over the graph. Thus we have that:
CONNECTIVITY$(n) \in \mathcal{L}^3_{\lceil \log(n-1) \rceil + 2}$, where the 3 variable property is directly inherited from PATH.
As for the fact that CONNECTIVITY$(n) \notin \mathcal{L}^2$, that was already demonstrated in Section 6.1.3.
Finally, consider two graphs G_r and H_r where G_r is the disjoint union of two 2^{r+1}+1 cycles, while H_r is a 2^{r+2}+2 cycle. Then it's not hard to see that these two structures are 2^r-nice structures (the d-balls are all either complete path graphs or clipped path graphs, and there are equal numbers of both types in G_r and H_r) w.r.t their initial configurations, and thus by Hanf's theorem, connectivity is indescribable over $\mathcal{L}_r$. Now, set $r = \lceil \log(n-2) \rceil - 2$ so that 2^{r+2}+2 is the smallest "such number" $(2^x + 2) \geq n$.
Thus CONNECTIVITY$(n) \in \mathcal{L}^3_{\lceil \log(n-1) \rceil + 2} \setminus \mathcal{L}^2_{\lceil \log(n-2) \rceil - 2}$.
Hence proved. □

6.1.5 REACHABILITY(x, y)

Let REACHABILITY(x, y) be the set of all undirected graphs for which the vertex x is reachable from the vertex y, where x and y are two given constant vertices.

Proposition 5. *REACHABILITY(x, y) is NOT FO expressible.*

Proof. Consider the same example which we considered while proving CONNECTIVITY$(n) \notin \mathcal{L}^2_{\lceil \log(n-2) \rceil - 2}$, ie:- let G_r and H_r be two graphs where G_r is the disjoint union of two 2^{r+1}+1 cycles, while H_r is a 2^{r+2}+2 cycle. Now let $G_r' := G_{r+2}$, $H_r' := H_{r+2}$. Now, in G_r' place the two constants in two different connected components, while in H_r' we may keep the constants on any two distinct vertices of our choice. Now, note that a $r+2$ game on G_r and H_r, is a r move game on G_r' and H_r' due to the two pairs of pre-placed pebbles. Thus, directly importing results from Section 6.1.4, we obtain that $G_r' \equiv_r H_r'$, and consequently, as many times before, we establish that reachability is FO-inexpressible. □

6.1.6 TWO-COLORABILITY

Let TWO-COLORABILITY be the set of all two-colorable (bipartite) graphs in $\mathcal{G}$.

Proposition 6. *TWO-COLORABILITY is NOT FO expressible.*

Proof. Consider the same example which we considered while proving CONNECTIVITY$(n) \notin \mathcal{L}^2_{\lceil \log(n-2) \rceil - 2}$, ie:- let G_r and H_r be two graphs where G_r is the disjoint union of two 2^{r+1}+1 cycles, while H_r is a 2^{r+2}+2 cycle. Then G_r is NOT two-colorable, while H_r is. Since we already established the 2^r-niceness of these two graphs in Section 6.1.4, by the same principle we can posit that since $G_r \equiv_r H_r \ \forall\ r \geq 1$, we have that **two-colorability** is FO-inexpressible. □

6.1.7 ACYCLICITY

Let ACYCLICITY be the set of all acyclic graphs in $\mathcal{G}$.

Proposition 7. *ACYCLICITY is NOT FO expressible.*

Proof. consider two graphs G_r and H_r where G_r is the disjoint union of a 2^{r+2} cycle and a 2^{r+2} path graph, while H_r is a 2^{r+3} cycle. Then it's not hard to see that these two structures are 2^r-nice structures (the d-balls are all either complete path graphs or clipped path graphs, and there are equal numbers of both types in G_r and H_r) w.r.t their initial configurations, and thus by Hanf's theorem, $G_r \equiv_r H_r \ \forall\ r \geq 1$, and consequently acyclicity is FO inexpressible. □

7 Conclusion

Thus the reader saw over the span of a dozen pages or so how methodology for establishing the expressibility of first-order logic was carefully developed, with the final highlight being the demonstration of the fact that first-order logic is inherently **local**, a notion which was formalized with the notion of Gaifman graphs and Hanf's theorem, and it is this locality of first-order logic that prevents it from expressing "global" properties such as acyclicity and bipartite-ness.
Mathematics is a very exacting science in the sense that establishing anything string and concrete is most of the times very difficult, and while initial forays into any field often start off encouragingly, they are soon besieged by intractable problems from all directions whence progress halts.
Thus, fields that are able to logically derive their stated goals through reams of logic are often mathematical masterpieces: case in point is pre-19$^{\text{th}}$ century Number Theory, whose high noon is marked by the proof of the Quadratic Reciprocity theorem, bringing a satisfying conclusion to the works of Fermat, Euler, and Gauss, and yet opening up new avenues for research.
In not dissimilar a manner, we have now conclusively established the limits of first-order logic, what it can do, and what it can't, and while that is a magnificent ending, it is also a beginning of further study into other types of logic, such as Monadic Second Order Logic.
Perhaps not as momentous as the achievement that Quadratic Reciprocity was, the reader is still implored to look back, and appreciate the beauty of what was just done.

8 References

[1] Neil Immerman, *Descriptive Complexity*. Springer Graduate Texts in Computer Science, 1999