Extension preservation in the finite and prefix classes of first order logic

Abhisekh Sankaran
University of Cambridge

Joint work with Anuj Dawar

CSL '21
Jan 25, 2021

Introduction

- The Łoś-Tarski theorem ('54 – '55) from model theory characterizes FO definable extension preserved properties of arbitrary structures in terms of existential sentences.
- Historically significant: among the earliest applications of Gödel's Compactness theorem and opened the area of preservation theorems in model theory.
- Fails in the finite: there is an extension preserved FO sentence that is not equivalent to any existential sentence over all finite structures (Tait, '59).
- Rosen and Weinstein observed ('94) that Tait's sentence is expressible in $\texttt{Datalog}(\neg)$, and asked if FO $\cap$ $\texttt{Datalog}(\neg)$ is contained in some prefix class of FO beyond the existential.

Main results

$$\Sigma_n := \exists\bar{x}_1\forall\bar{x}_2\exists\bar{x}_3\ldots Q\bar{x}_n\alpha(\bar{x}_1,\ldots,\bar{x}_n)$$ where α is quantifier-free
$$\Pi_n := \forall\bar{x}_1\exists\bar{x}_2\forall\bar{x}_3\ldots Q^c\bar{x}_n\alpha(\bar{x}_1,\ldots,\bar{x}_n)$$ and $Q = \forall$ iff n is even.

Theorem

Tait's sentence is a $\Sigma_3 \cap \texttt{Datalog}(\neg)$ sentence that is extension preserved over all finite structures, but is not equivalent over this class to any Π_3 sentence.

Theorem

For every n, there is a vocabulary σ_n and a $\Sigma_{2n+1} \cap \texttt{Datalog}(\neg)$ sentence of FO(σ_n) that is extension preserved over all finite structures, but is not equivalent over this class to any Π_{2n+1} sentence.

Main results

$$\Sigma_n := \exists\bar{x}_1\forall\bar{x}_2\exists\bar{x}_3\ldots Q\bar{x}_n\alpha(\bar{x}_1,\ldots,\bar{x}_n)$$ where α is quantifier-free
$$\Pi_n := \forall\bar{x}_1\exists\bar{x}_2\forall\bar{x}_3\ldots Q^c\bar{x}_n\alpha(\bar{x}_1,\ldots,\bar{x}_n)$$ and $Q = \forall$ iff n is even.

Theorem

Tait's sentence is a $\Sigma_3 \cap \texttt{Datalog}(\neg)$ sentence that is extension preserved over all finite structures, but is not equivalent over this class to any Π_3 sentence.

Theorem

No prefix class of FO is expressive enough to capture:

- Extension preserved FO properties in the finite, and even
- FO $\cap$ $\texttt{Datalog}(\neg)$ queries in the finite[a]

[a] Resolves an open problem posed by Rosen and Weinstein in '94

The sentence SomeTotalR_n

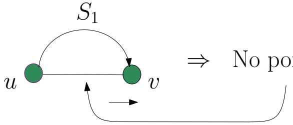
Construction of the sentence

$$\text{SomeTotalR}_1 := (\text{LO} \wedge \text{PartialSucc}_1) \rightarrow \exists u \exists v \ \text{RTotal}_1(u, v)$$

$$(\in \text{FO}(\sigma_1) \text{ where } \sigma_1 = \{\leq, R_1, S_1\})$$

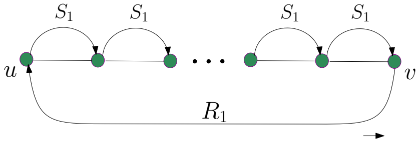
LO := " $\leq$ is a linear order"

$\text{PartialSucc}_1 := \forall u \forall v$



$\Rightarrow$ No points between u and v

$\text{RTotal}_1(u, v) :=$



Construction of the sentence

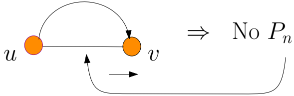
$$\text{SomeTotalR}_n := (\text{LO} \wedge \text{PartialSucc}_n) \rightarrow \exists u \exists v \ \text{RTotal}_n(u, v)$$

$$(\in \text{FO}(\sigma_n) \text{ where } \sigma_n = \sigma_{n-1} \cup \{P_n, R_n, S_n\})$$

$\text{LO} :=$ " $\leq$ is a linear order"

$\text{PartialSucc}_n := \forall u \forall v$

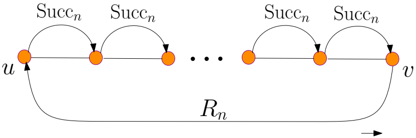
$$\text{Succ}_n := P_n(u) \wedge P_n(v) \wedge S_n(u, v) \wedge \text{SomeTotalR}_{n-1}^{[u,v]}$$



$\Rightarrow$ No P_n points between u and v

● $-$ P_n

$\text{RTotal}_n(u, v) :=$



Construction of the sentence

$$\mathrm{SomeTotalR}_n := (\mathrm{LO} \wedge \mathrm{PartialSucc}_n) \rightarrow \exists u \exists v \ \mathrm{RTotal}_n(u, v)$$

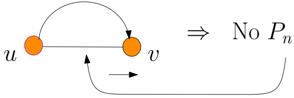
$$(\in \mathrm{FO}(\sigma_n) \text{ where } \sigma_n = \sigma_{n-1} \cup \{P_n, R_n, S_n\})$$

$\mathrm{LO} :=$ " $\leq$ is a linear order"

$\mathrm{PartialSucc}_n := \forall u \forall v$

$$\mathrm{SomeTotalR}_n \in \Sigma_{2n+1}$$

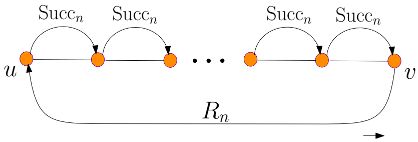
$$\mathrm{Succ}_n := P_n(u) \wedge P_n(v) \wedge S_n(u, v) \wedge \mathrm{SomeTotalR}_{n-1}^{[u,v]}$$



$\Rightarrow$ No P_n points between u and v

● $- P_n$

$\mathrm{RTotal}_n(u, v) :=$



SomeTotalR_n is $\texttt{Datalog}(\neg)$ expressible

- A $\texttt{Datalog}(\neg)$ rule is of one of the foll. forms:

$$\begin{array}{lcl} R(\bar{x}) & \longleftarrow & A(\bar{x}_1) \\ R(\bar{x}) & \longleftarrow & R_1(\bar{x}_1), \ldots, R_n(\bar{x}_n) \end{array}$$

  where A is an atom (equality included) or its negation, but each non-atom predicate R_i (which could be R) must appear un-negated. In both rules, LHS variables $\subseteq$ RHS variables.
- A $\texttt{Datalog}(\neg)$ program is a finite set of $\texttt{Datalog}(\neg)$ rules.
- Every $\texttt{Datalog}(\neg)$ program is extension closed.
- $\neg$ LO and $\neg$ Succ_n are easily expressible in $\texttt{Datalog}(\neg)$ via induction; RTotal_n is essentially a reachability condition and hence $\texttt{Datalog}(\neg)$ expressible. Then so is SomeTotalR_n.

SomeTotalR$_n$ is not Π_{2n+1} expressible

Theorem

The Σ_{2n+1} sentence SomeTotalR$_n$ is not equivalent over all finite σ_n-structures to any Π_{2n+1} sentence.

- Let $\Sigma_{n,k}$ = all Σ_n sentences in which each quantifier block has size $\leq k$; so $\Sigma_n = \bigcup_{k \geq 0} \Sigma_{n,k}$. Define $\Pi_{n,k}$ analogously.
- Let $\mathcal{A} \Rrightarrow_{n,k} \mathcal{B}$ = for each $\Sigma_{n,k}$ sentence θ, it holds that $\mathcal{A} \models \theta \rightarrow \mathcal{B} \models \theta$.
- $\mathcal{A} \Rrightarrow_{n,k} \mathcal{B}$ is equivalent to: for each $\Pi_{n,k}$ sentence γ, it holds that $\mathcal{B} \models \gamma \rightarrow \mathcal{A} \models \gamma$.
- For each n, k, we construct a model $\mathcal{M}_{n,k}$ and a non-model $\mathcal{N}_{n,k}$ of SomeTotalR$_n$ such that $\mathcal{N}_{n,k} \Rrightarrow_{2n+1,k} \mathcal{M}_{n,k}$.

Ehrenfeucht-Fraïssé game argument

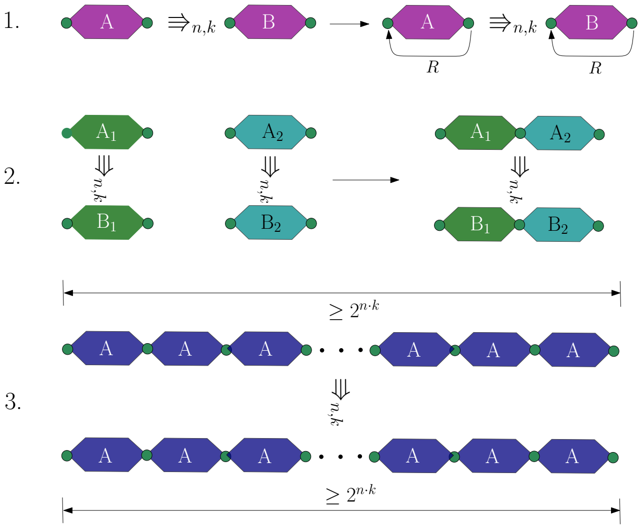
Ehrenfeucht-Fraïssé (EF) game for $\Rrightarrow_{n,k}$

- Two players: Spoiler and Duplicator; Game arena: a pair $(\mathcal{A}, \mathcal{B})$ of structures; Rounds: n.
- In odd rounds i, Spoiler chooses a k-tuple $\bar{a}_i$ from $\mathcal{A}$ and in even rounds i, he chooses a k-tuple $\bar{b}_i$ from $\mathcal{B}$.
- Duplicator responds with k-tuples $\bar{b}_i$ from $\mathcal{B}$ in odd rounds and with k-tuples $\bar{a}_i$ from $\mathcal{A}$ in even rounds.
- Duplicator wins the play of the game if $(\bar{a}_i \mapsto \bar{b}_i)_{1 \leq i \leq n}$ is a partial isomorphism between $\mathcal{A}$ and $\mathcal{B}$. She has a winning strategy if she wins every play of the game.

Theorem

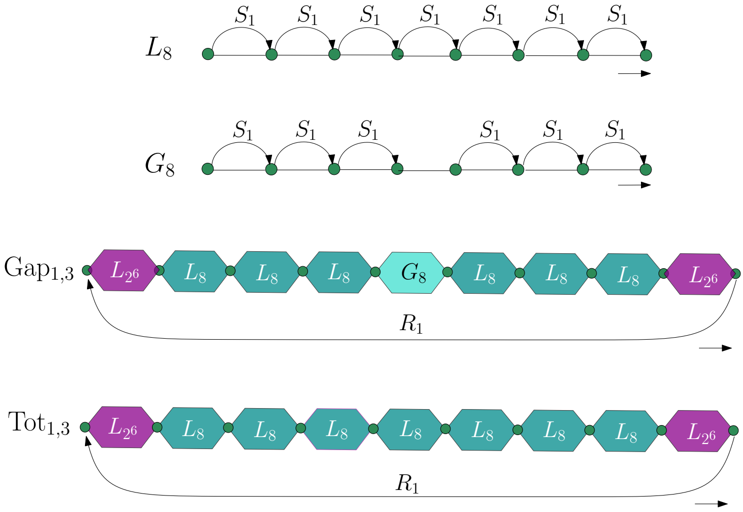
Duplicator has a winning strategy in the above game iff $\mathcal{A} \Rrightarrow_{n,k} \mathcal{B}$.

Composition properties

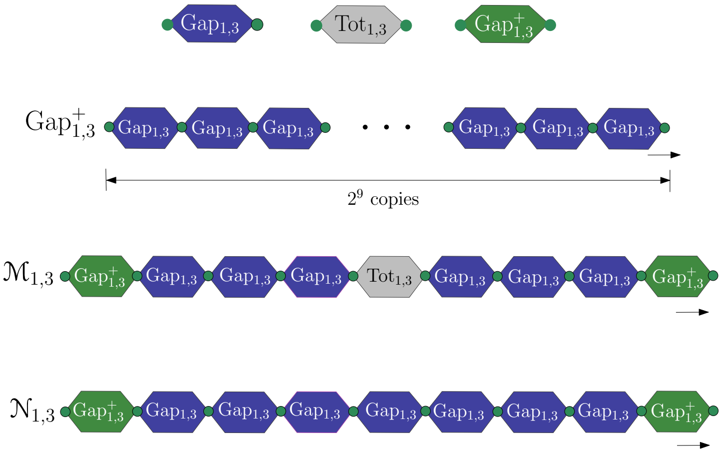


Construction of $\mathcal{M}_{n,k}$ and $\mathcal{N}_{n,k}$
(Illustrated for $k = 3$)

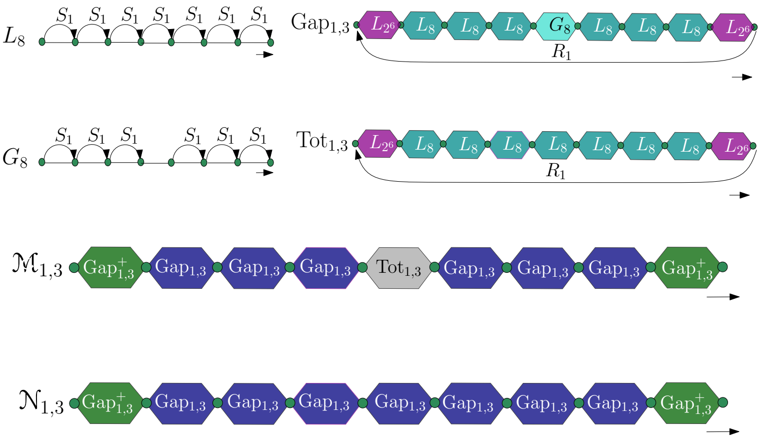
$\mathcal{M}_{1,3}$ and $\mathcal{N}_{1,3}$



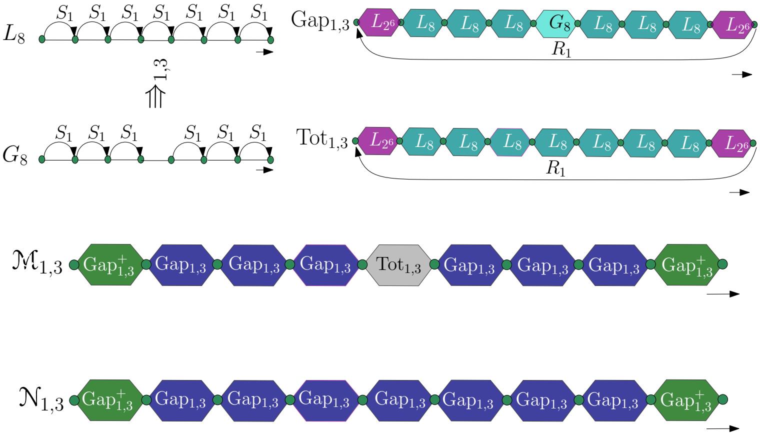
$\mathcal{M}_{1,3}$ and $\mathcal{N}_{1,3}$



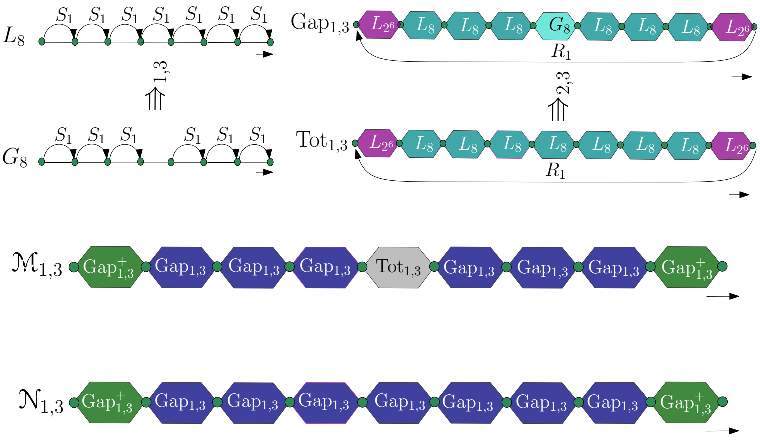
$\mathcal{M}_{1,3}$ and $\mathcal{N}_{1,3}$



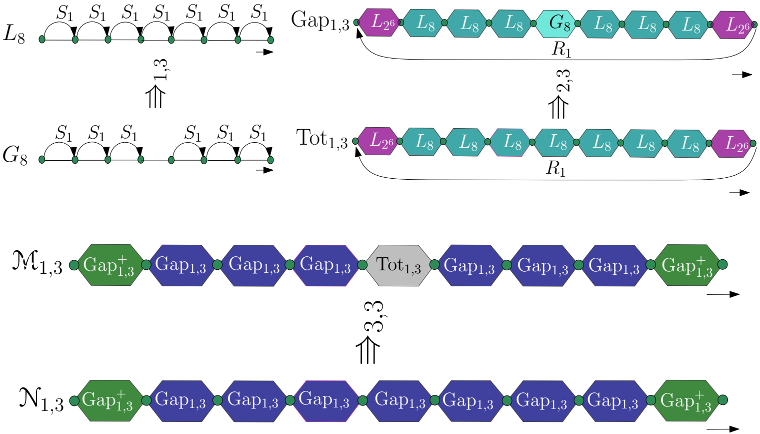
$\mathcal{M}_{1,3}$ and $\mathcal{N}_{1,3}$



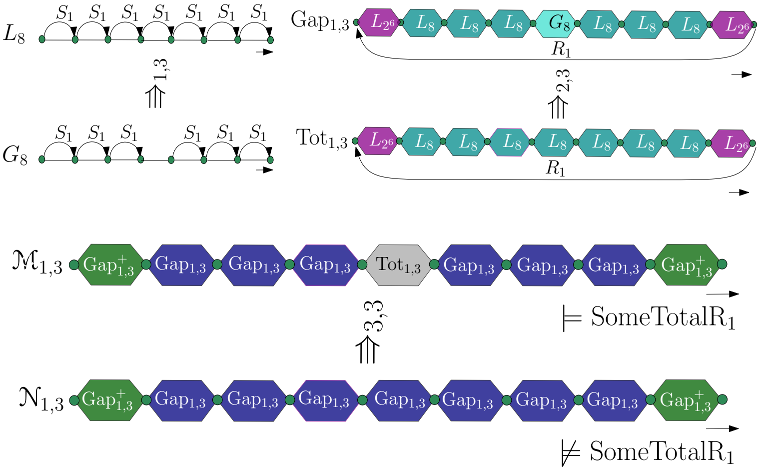
$\mathcal{M}_{1,3}$ and $\mathcal{N}_{1,3}$



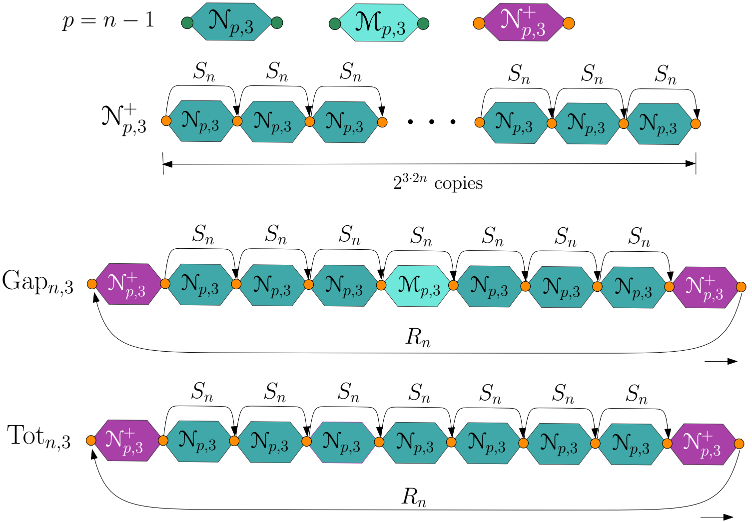
$\mathcal{M}_{1,3}$ and $\mathcal{N}_{1,3}$



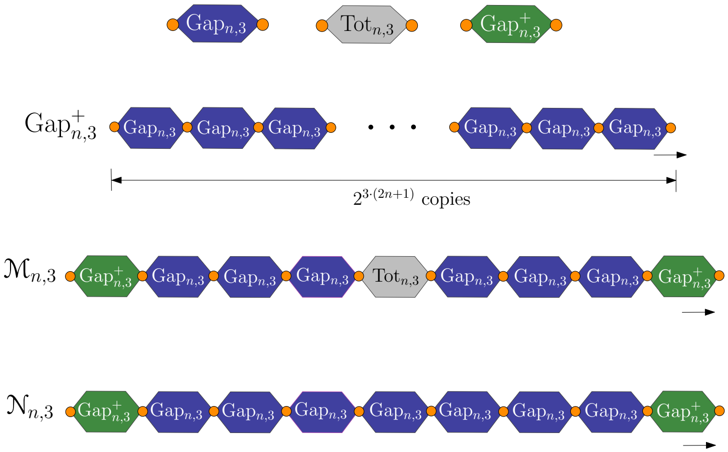
$\mathcal{M}_{1,3}$ and $\mathcal{N}_{1,3}$



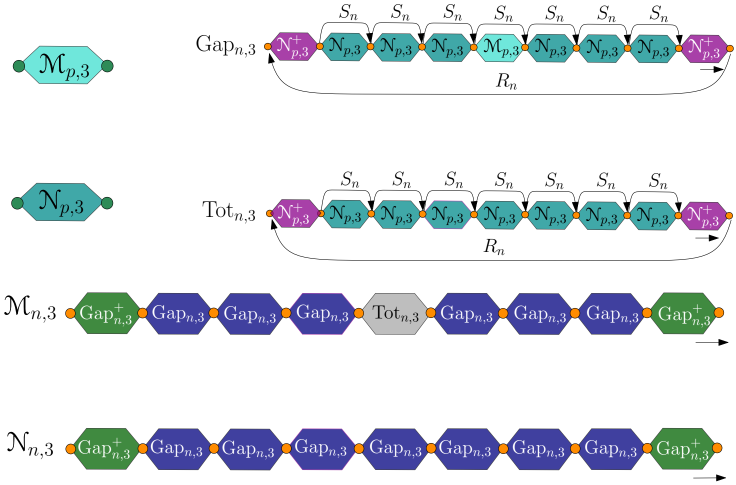
$\mathcal{M}_{n,3}$ and $\mathcal{N}_{n,3}$



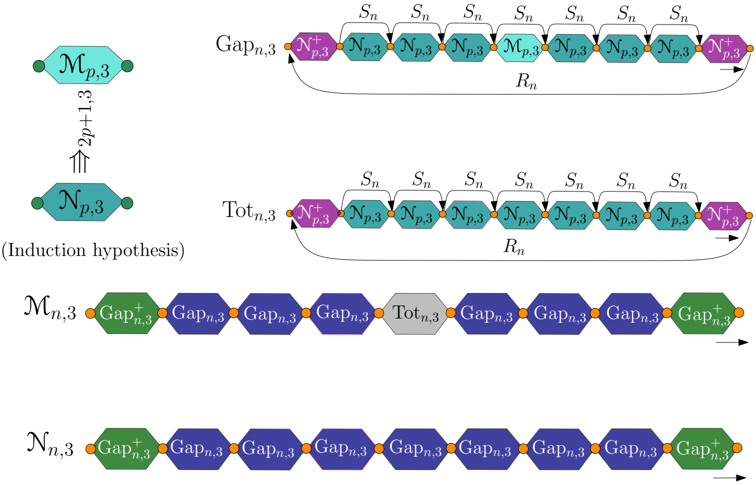
$\mathcal{M}_{n,3}$ and $\mathcal{N}_{n,3}$



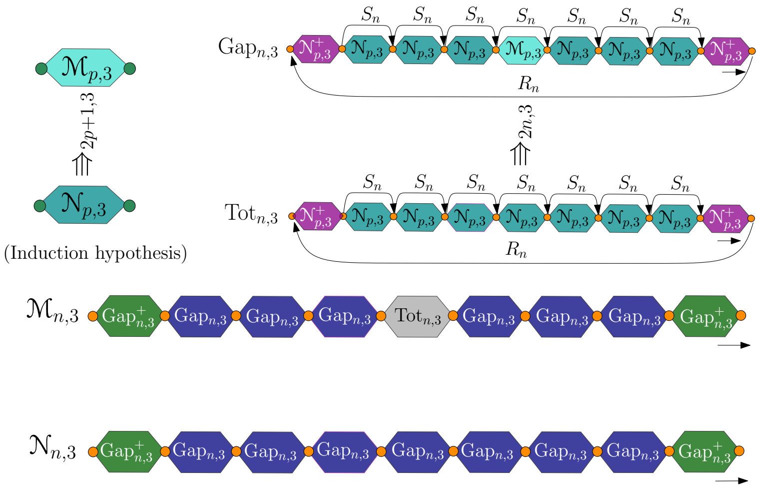
$\mathcal{M}_{n,3}$ and $\mathcal{N}_{n,3}$



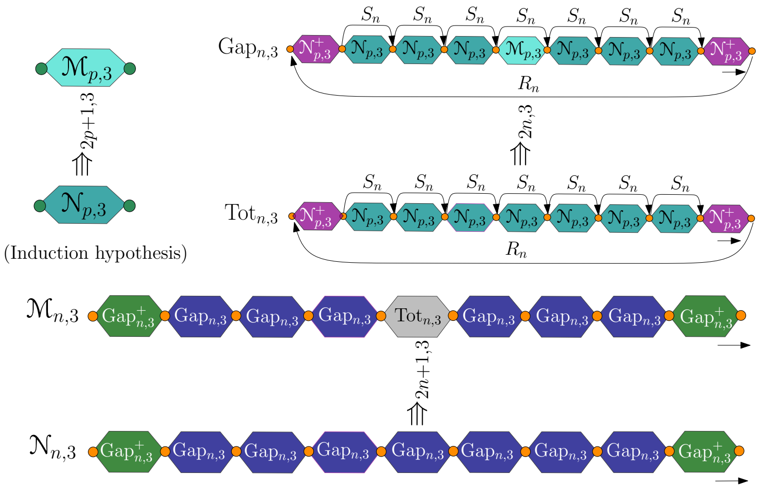
$\mathcal{M}_{n,3}$ and $\mathcal{N}_{n,3}$



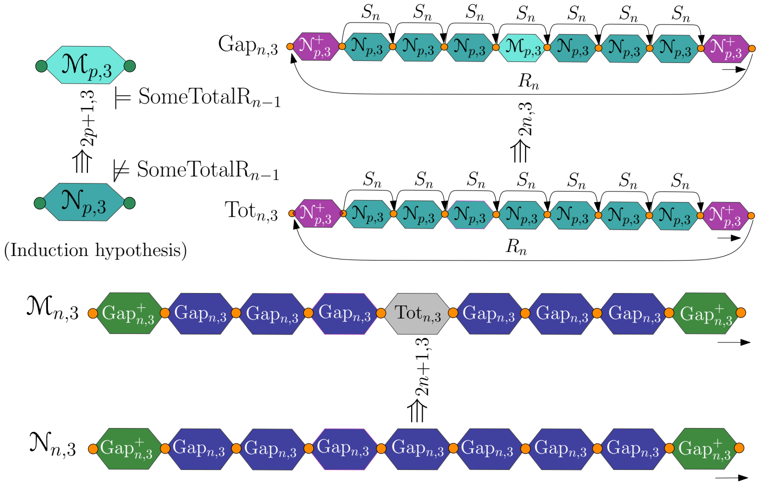
$\mathcal{M}_{n,3}$ and $\mathcal{N}_{n,3}$



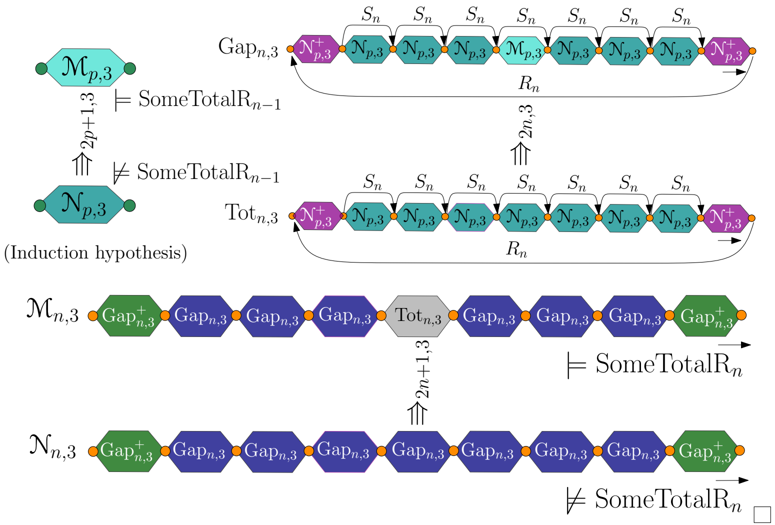
$\mathcal{M}_{n,3}$ and $\mathcal{N}_{n,3}$



$\mathcal{M}_{n,3}$ and $\mathcal{N}_{n,3}$



$\mathcal{M}_{n,3}$ and $\mathcal{N}_{n,3}$



Conclusion

Main results

Theorem

Tait's sentence is a $\Sigma_3 \cap \texttt{Datalog}(\neg)$ sentence that is extension preserved over all finite structures, but is not equivalent over this class to any Π_3 sentence.

Theorem

No prefix class of FO is expressive enough to capture:

- Extension closed FO properties in the finite, and even
- FO $\cap$ $\texttt{Datalog}(\neg)$ queries in the finite[a]

[a] Resolves an open problem posed by Rosen and Weinstein in '94

Future directions

- The sentence $\mathsf{SomeTotalR}_n$ is over a vocabulary σ_n that grows with n.
- Further, σ_n can be seen as the vocabulary of ordered vertex colored and edge colored graphs.

Question 1.

Is there a fixed (finite) vocabulary σ^* such that prefix classes fail to capture extension preserved FO properties of finite σ^*-structures?

Question 2.

Do prefix classes fail to capture extension preserved FO properties of undirected graphs (possibly vertex colored)?

Future directions

- The sentence $\mathsf{SomeTotalR}_n$ is over a vocabulary σ_n that grows with n.
- Further, σ_n can be seen as the vocabulary of ordered vertex colored and edge colored graphs.

Question 1. (Resolved: Yes! $|\sigma^*| \leq 4$)

Is there a fixed (finite) vocabulary σ^* such that prefix classes fail to capture extension preserved FO properties of finite σ^*-structures?

Question 2. (Not resolved yet)

Do prefix classes fail to capture extension preserved FO properties of undirected graphs (possibly vertex colored)?

Thank you!