#### **MSO-checking**

Fernando Sánchez Villaamil

Theoretical Computer Science



@NCSU 2014

### As seen in previous lectures

#### Theorem

The languages over trees which are recognizable by DTA are precisely those languages that are definable in  $MSO_1$  logic.

#### Corollary

Every graph property definable in monadic second-order logic can be decided in linear time for trees of bounded degree.

Proof.

- **2** Run  $\mathcal{A}$  on the given tree.

### As seen in previous lectures

#### Theorem

It is possible to convert EMSO-formulas for labeled trees of bounded degree into automata that recognize the same language over trees.

## Today's menu

- How hard MSO-checking is on graphs in general?
- Easy solution for MAXIMUM INDEPENDENT SET on trees of bounded degree.
- An unnecessarily complicated algorithm that solves MAXIMUM INDEPENDENT SET on trees of unbounded degree in linear time.
- Show that what we did for MAXIMUM INDEPENDENT SET can be used to solve any MSO<sub>1</sub> formula on trees of unbounded degree.
- Generalize the proof so that it works on any graph class of bounded treewidth.
- 6 We are gonna talk about the context of these results.

## How hard is MSO<sub>1</sub>-checking?

#### Definition (PSPACE)

The complexity class **PSPACE** contains all problems which are solvable using polynomial space.

Theorem  $NP \subseteq PSPACE.$ 

#### Proof.

For a problem in  $\mathbf{NP}$  it is possible to enumerate all its possible polynomially sized certificates and check them in polynomial time.

(It is assumed, but it has not yet been proven, that  $\mathbf{NP} \subsetneq \mathbf{PSPACE})$ 

### How hard is MSO<sub>1</sub>-checking?

Theorem MSO<sub>2</sub>-checking is **PSPACE**-complete.

#### **Definition** (*k*-INDEPENDENT SET)

Given a graph G, does there exist a set S of vertices of size k such that the graph induced by S is edgeless.

Does there exists a linear time algorithm that solves MAXIMUM INDEPENDENT SET on trees of bounded degree?









For an ordered tree with maximum degree d + 1 we can write the following constant sized formula to check if there is an edge between two nodes.

$$\phi_{\mathsf{edge}}(x,y) := S_1(x,y) \lor \ldots \lor S_d(x,y) \lor S_1(y,x) \lor \ldots \lor S_d(y,x)$$

We can now write a simple constant sized formula for MAXIMUM INDEPENDENT SET on trees of bounded degree.

$$\max S(\forall x \forall y ((x \in S \land y \in S) \to (x = y \lor \neg \phi_{\mathsf{edge}}(x, y))))$$

#### Theorem

A tree of bounded degree T has an independent set of size k iff there exists a set  $S \subseteq G$  of size k such that

$$(G,S) \models \forall x, \forall y ((x \in S \land y \in S) \rightarrow (x = y \lor \neg \phi_{\textit{edge}}(x,y))$$

#### Theorem

A tree of bounded degree T has an independent set of size k iff there exists a set  $S \subseteq G$  of size k such that

$$(G,S) \models \forall x, \forall y ((x \in S \land y \in S) \to (x = y \lor \neg \phi_{\textit{edge}}(x,y))$$

Corollary

It is possible to find an independent set of maximal size in linear time on trees of bounded degree. INDEPENDENT SET on trees of unbounded degree

Question Does there exists a linear time algorithm that solves MAXIMAL INDEPENDENT SET on trees of **unbounded** degree? INDEPENDENT SET on trees of unbounded degree

Idea

- Convert the tree into a nice tree-decomposition of width one.
- Adapt the MAXIMUM INDEPENDENT SET EMSO formula for trees to work on nice tree decompositions of width one.
- Since a nice tree decomposition is a binary tree, we can now convert the new EMSO formula to a deterministic automata and solve the problem in linear time.

# Finding a binary tree decompositions













Can we treat the tree decomposition as a labeled graph?



Not like this: The size of the labeling depends on the number of nodes of the original tree which is problematic.







Can we introduce bags that represent nodes?













Are the nodes represented by these leaves connected?



Which bags contain the node represented by this leaf?

## Label interpretation

- The parent bag is exactly the same bag, order included.
- The node on the left position appears in the parent bag at the right position.
- The bag is a leaf representing a node from the original tree. It can have three further meanings:
  - If it is a single child, then the node it represents is in the right position in the parent.
  - If it is not a single child and it is a left child, then it is contained in the left position in the parent.
  - If it is not a single child and it is a right child, then it is contained in the right position in the parent.

Let us write formulas  $\phi_{i,j}(s,t)$  for  $0 \le i, j \le 1$  that are true if s is a parent of t in the tree decomposition and the positions j in s and i in t are the same.

Let us write formulas  $\phi_{i,j}(s,t)$  for  $0 \le i, j \le 1$  that are true if s is a parent of t in the tree decomposition and the positions j in s and i in t are the same.

$$\phi_{0,0}(s,t) := (S_0(s,t) \land (\blacksquare(t) \lor \blacksquare(t))) \\ \lor (S_1(s,t) \land \blacksquare(t))$$

Let us write formulas  $\phi_{i,j}(s,t)$  for  $0 \le i, j \le 1$  that are true if *s* is a parent of *t* in the tree decomposition and the positions *j* in *s* and *i* in *t* are the same.

$$\phi_{0,0}(s,t) := (S_0(s,t) \land (\blacksquare(t) \lor \blacksquare(t))) \\ \lor (S_1(s,t) \land \blacksquare(t))$$

$$\phi_{0,1}(s,t) := (S_0(s,t) \land \blacksquare(t) \land \exists y(S_1(s,y)))$$
$$\lor (S_1(s,t) \land \blacksquare(t))$$
$$\lor ((S_0(s,t) \lor S_1(s,t)) \land \blacksquare(t))$$

Let us write formulas  $\phi_{i,j}(s,t)$  for  $0 \le i, j \le 1$  that are true if *s* is a parent of *t* in the tree decomposition and the positions *j* in *s* and *i* in *t* are the same.

$$\phi_{0,0}(s,t) := (S_0(s,t) \land (\blacksquare(t) \lor \blacksquare(t))) \\ \lor (S_1(s,t) \land \blacksquare(t))$$

$$\phi_{0,1}(s,t) := (S_0(s,t) \land \blacksquare(t) \land \exists y(S_1(s,y)))$$
$$\lor (S_1(s,t) \land \blacksquare(t))$$
$$\lor ((S_0(s,t) \lor S_1(s,t)) \land \blacksquare(t))$$

 $\phi_{1,1}(s,t) := (S_0(s,t) \lor S_1(s,t)) \land [t]$
## All the bags that contain a node

Assume that for any node u the set  $X_0$  contains all bags that contain u at the left position and the set  $X_1$  contains all bags that contain u at the right position. Let us write a formula  $\phi_{\text{conn}}(X_0, X_1)$  which is true on these sets, but false for all sets  $X'_0, X'_1 \subsetneq X_0, X_1$ .

### All the bags that contain a node

Assume that for any node u the set  $X_0$  contains all bags that contain u at the left position and the set  $X_1$  contains all bags that contain u at the right position. Let us write a formula  $\phi_{\text{conn}}(X_0, X_1)$  which is true on these sets, but false for all sets  $X'_0, X'_1 \subsetneq X_0, X_1$ .

$$\begin{split} \phi_{\mathsf{conn}}(X_0, X_1) &:= X_0 \neq \varnothing \land X_1 \neq \varnothing \land \\ \forall s \forall t (((t \in X_0 \to (\phi_{0,0}(s, t) \to s \in X_0) \\ & \lor (\phi_{0,0}(t, s) \to s \in X_0) \\ & \lor (\phi_{0,1}(s, t) \to s \in X_1)) \\ & \land (t \in X_1 \to (\phi_{1,1}(s, t) \to s \in X_1) \\ & \lor (\phi_{1,1}(t, s) \to s \in X_1)))) \end{split}$$

Problem: If  $X_0$  and  $X_1$  contain all bags, then  $\phi_{\text{conn}}(X_0, X_1)$  is also true.

Problem: If  $X_0$  and  $X_1$  contain all bags, then  $\phi_{\text{conn}}(X_0, X_1)$  is also true.

Solution: Force exclusion minimality.

$$\begin{split} \phi_{\mathsf{exc}}(X_0, X_1) &:= \phi_{\mathsf{conn}}(X_0, X_1) \land \\ \forall X'_0 \forall X'_1 ((X'_0 \subsetneq X_0 \lor X'_1 \subsetneq X_1) \\ & \land (X'_0 \subseteq X_0 \land X'_1 \subseteq X_1) \\ & \to \neg \phi_{\mathsf{conn}}(X'_0, X'_1)) \end{split}$$

Problem: If  $X_0$  and  $X_1$  contain all bags, then  $\phi_{\text{conn}}(X_0, X_1)$  is also true.

Solution: Force exclusion minimality.

$$\begin{split} \phi_{\mathsf{exc}}(X_0, X_1) &:= \phi_{\mathsf{conn}}(X_0, X_1) \land \\ \forall X'_0 \forall X'_1 ((X'_0 \subsetneq X_0 \lor X'_1 \subsetneq X_1) \\ & \land (X'_0 \subseteq X_0 \land X'_1 \subseteq X_1) \\ & \to \neg \phi_{\mathsf{conn}}(X'_0, X'_1)) \end{split}$$

Finally  $\phi_{\text{exc}}(X_0, X_1)$  is true iff there exists a node u such that the set of bags that contain u is precisely  $X_0 \cup X_1$ .

## Interpretation

MSO-formula for trees of bounded degree.

$$\begin{aligned} \phi_{\mathsf{edge}}(x,y) &:= S_1(x,y) \lor \ldots \lor S_d(x,y) \lor S_1(y,x) \lor \ldots \lor S_d(y,x) \\ \max S(\forall x \forall y ((x \in S \land y \in S) \to (x = y \lor \neg \phi_{\mathsf{edge}}(x,y)))) \end{aligned}$$

## Interpretation

We can now rewrite the formula for edges with the help of  $\phi_{\text{exc}}$ :

$$\begin{split} \phi_{\mathsf{edge}}(x,y) &:= x \neq y \land \exists X_0 \exists X_1 \exists Y_0 \exists Y_1 (x \in X_0 \land y \in Y_0 \\ \land \phi_{\mathsf{exc}}(X_0, X_1) \land \phi_{\mathsf{exc}}(Y_0, Y_1) \\ \land \exists z ((z \in X_0 \lor z \in X_1) \land (z \in Y_0 \lor z \in Y_1))) \end{split}$$

And finally rewrite the formula for INDEPENDENT SET on trees.

$$\max S(\forall x \forall y( (x) \land (x) \land (y) \land (x \in S \land y \in S))) \rightarrow (x = y \lor \neg \phi_{\mathsf{edge}}(x, y))))$$

# Linear algorithm for INDEPENDENT SET on trees

- 1 Compute a tree decomposition T of the input tree T with the following properties:
  - T has width one.
  - T is a binary tree.
  - Every node of T appears as a leaf bag in  $\mathcal{T}$ .

This is doable in linear time.

- ② Compute a label tree T' that represent T.
- Since the formula we created for INDEPENDENT SET on tree decompositions of width one has constant size, we can compute a DTA for it.
- **4** Run the DTA on T' and repeat the output.

Since there exists a rather simple dynamic programming algorithm to compute independent set on trees in linear time our proof via tree-automata has been more of an exercise in mental resilience than anything else...unless we can use these techniques to prove something more general!

# MSO<sub>1</sub>-checking on trees of unbounded degree

We can use a similar algorithm to solve all problems expressible with a MSO<sub>1</sub> formula  $\psi$  of constant size on trees in linear time.

- 1 Compute a tree decomposition  $\mathcal{T}$  of the input tree T with the following properties:
  - T has width one.
  - T is a binary tree.
  - Every node of T appears as a leaf bag in  $\mathcal{T}$ .

This is doable in linear time.

- ② Compute a label tree T' that represent T.
- Translate  $\psi$  into a formula  $\psi'$  on tree decompositions, by forcing all variables to be leaves in the tree decomposition and replacing the edge predicate by our MSO<sub>1</sub>-formula for the edge relation.
- 4 We can compute a DTA for  $\psi'$  in constant time.
- Sun the DTA on T' and repeat the output.

Question Can we do something similar for any constant treewidth *w*?

Question

Can we do something similar for any constant treewidth w?

Theorem (Courcelle's Theorem)

Every graph property definable in monadic second-order logic can be decided in linear time on graphs of bounded treewidth.

Question

Can we do something similar for any constant treewidth w?

Theorem (Courcelle's Theorem)

Every graph property definable in monadic second-order logic can be decided in linear time on graphs of bounded treewidth.

[This] result was first proved by Bruno Courcelle in 1990 and is considered the archetype of algorithmic meta-theorems.

—Wikipedia

# **Bruno Courcelle**



# Labeled trees for general tree decompositions

Assume you are given a tree decomposition of width w, how can we convert it into a tree with a finite labeling (where the size can only depend on w) that is usable?





# Labeled trees for general tree decompositions

- Write every bag set as a tuple of size w + 1 (repeat vertices if necessary).
- 2 For every bag t generate a labeling  $\lambda(t) := [eq(t)], [overlap(t)], [edge(t)]$  with the following properties:
  - eq(t) Is a list of pairs of positions in the bag which represent the same vertex of graph.
- overlap(t) If there is a parent  $p = (p_0, \dots, p_w)$  of  $t = (t_0, \dots, t_w)$  it contains a pair (i, j) if  $t_i = p_j$ .
  - edge(t) Contains a pair (i, j) if for  $t = (t_0, \dots, t_w)$  there is an edge in the original graph between  $t_i$  and  $t_j$ .

The number of possible words representing bags of the tree decomposition is bounded by  $2^{(w+1)^2} \cdot 2^{(w+1)^2} \cdot 2^{(w+1)^2}$ , thus the alphabet of the tree has constant size.

## Interpretation for width w

We can repeat the same principles we used for labeled trees representing tree decompositions of width one for labeled trees representing trees of width w.

Instead of using formulas like  $\phi_{0,0}$ ,  $\phi_{0,1}$  and  $\phi_{1,1}$  to write a formula which forces the inclusion of bags when they overlap with bags already contained we use the relation  $overlap_{i,j}$  to do the same.

$$\begin{split} \phi_{\text{overlap}}(X_0, \dots, X_w) &:= \forall s \forall t (\\ & \bigwedge_{i,j} t \in X_i \land ((S_0(s,t) \lor S_1(s,t) \land overlap_{i,j}(t)) \\ & \lor (S_0(t,s) \lor S_1(t,s) \land overlap_{j,i}(s)))) \to s \in X_j \end{split}$$

### Interpretation for width w

We can now force the inclusion of bags that overlap with bags already contained for a set for some position. We have to be careful to also deal with values repeating in a bag.

$$\phi_{\mathsf{eq}}(X_0, \dots, X_w) := \bigwedge_i \forall t (t \in X_i \to (\bigwedge_{i \neq j} eq_{i,j}(t) \to t \in X_j))$$

The above formula forces a bag that contains repeated nodes to be contained in all the appropriate position sets.

We can finally redefine  $\phi_{\text{conn}}(X_0, \ldots, X_w)$  to be a formula which is true when  $\bigcup_i X_i$  contains all bags that contain some node u.

$$\phi_{\mathsf{conn}}(X_0,\ldots,X_w) = \phi_{\mathsf{overlap}}(X_0,\ldots,X_w) \land \phi_{\mathsf{eq}}(X_0,\ldots,X_w)$$

We can finally redefine  $\phi_{\text{conn}}(X_0, \ldots, X_w)$  to be a formula which is true when  $\bigcup_i X_i$  contains all bags that contain some node u.

$$\phi_{\mathsf{conn}}(X_0,\ldots,X_w) = \phi_{\mathsf{overlap}}(X_0,\ldots,X_w) \land \phi_{\mathsf{eq}}(X_0,\ldots,X_w)$$

We can use as before exclusion minimality to make sure that we get exactly the bags that contain some node u.

$$\phi_{\mathsf{exc}}(X_0, \dots, X_w) := \phi_{\mathsf{conn}}(X_0, \dots, X_w) \land (\bigvee_i X_i \neq \varnothing) \land$$
$$\forall X'_0, \dots, X'_w((\bigvee_i X'_i \subsetneq X_i) \land (\bigwedge_i X'_i \subseteq X_i)) \to \neg \phi_{\mathsf{conn}}(X'_0, \dots, X'_w))$$

We can finally redefine  $\phi_{\text{conn}}(X_0, \ldots, X_w)$  to be a formula which is true when  $\bigcup_i X_i$  contains all bags that contain some node u.

$$\phi_{\mathsf{conn}}(X_0,\ldots,X_w) = \phi_{\mathsf{overlap}}(X_0,\ldots,X_w) \land \phi_{\mathsf{eq}}(X_0,\ldots,X_w)$$

We can use as before exclusion minimality to make sure that we get exactly the bags that contain some node u.

$$\phi_{\mathsf{exc}}(X_0, \dots, X_w) := \phi_{\mathsf{conn}}(X_0, \dots, X_w) \land (\bigvee_i X_i \neq \varnothing) \land$$
$$\forall X'_0, \dots, X'_w((\bigvee_i X'_i \subsetneq X_i) \land (\bigwedge_i X'_i \subseteq X_i)) \to \neg \phi_{\mathsf{conn}}(X'_0, \dots, X'_w))$$

And finally we can get the sought after formula for the edge relation.

$$\begin{split} \phi_{edge}(x,y) &:= x \neq y \land \exists X_0 \dots X_w \exists Y_0 \dots Y_w (x \in X_0 \land y \in Y_0 \\ \land \phi_{\mathsf{exc}}(X_0, \dots, X_w) \land \phi_{\mathsf{exc}}(Y_0, \dots, Y_w) \\ \land \exists z ((\bigvee_{i,j} z \in X_i \land z \in Y_j \land edge_{i,j}(z)))) \end{split}$$

## Size of the formula for edges

Question How big is our new  $\phi_{edge}(x, y)$ ?

## Size of the formula for edges

Question How big is our new  $\phi_{edge}(x, y)$ ?

Answer Constant, since the size only depends on w.

# Almost Courcelle's Theorem

Theorem (Almost Courcelle's Theorem)

Every graph property definable in  $MSO_1$  can be decided in linear time on graphs of bounded treewidth.

Proof.

The following algorithm achieves linear running time:

- Compute a tree decomposition of minimal width w for the graph in linear time.
- Onvert the tree decomposition into a leaf decomposition.
- Convert the leaf decomposition into a labeled tree T' whose labeling is bounded in w.
- Interpret the given MSO<sub>1</sub> formula to work on tree decompositions of width w.
- G Convert this formula into a DTA.
- **6** Run the DTA on T'.

#### Theorem

 $MSO_1$  and  $MSO_2$  have the same expressive power on graph classes of bounded treewidth.

#### Proof.

A graph class with bounded treewidth w is w-degenerate (every subgraph contains a node of maximal degree w). MSO<sub>1</sub> and MSO<sub>2</sub> have the same expressive power on graph classes with bounded degeneracy.

#### Theorem

 $MSO_1$  and  $MSO_2$  have the same expressive power on graph classes of bounded treewidth.

#### Proof.

A graph class with bounded treewidth w is w-degenerate (every subgraph contains a node of maximal degree w). MSO<sub>1</sub> and MSO<sub>2</sub> have the same expressive power on graph classes with bounded degeneracy.

We have now proven Courcelle's Theorem.

#### Theorem (Courcelle's Theorem)

Every graph property definable in monadic second-order logic can be decided in linear time on graphs of bounded treewidth.

## Improvement

#### Definition (L)

The complexity class  $\mathbf{L}$  contains all problems which can be solved in logarithmic space.

#### Theorem (Elberfeld, Jakoby and Tantau)

Every graph property definable in monadic second-order logic can be decided in **logarithmic space** on graphs of bounded treewidth.

# Running time

The running time of the previous algorithm for a formula  $\psi$  on treewidth w is  $f(||\psi||,w)\cdot n$  where  $f(||\psi||,w)$  is

$$f(||\psi||, w) := 2^{2^{w^{-2^{w}}}} \Big\} \Theta(||\psi||)$$

# Running time

The running time of the previous algorithm for a formula  $\psi$  on treewidth w is  $f(||\psi||,w)\cdot n$  where  $f(||\psi||,w)$  is

$$f(||\psi||, w) := 2^{2^{-1}} \Theta(||\psi||)$$

This is also basically the best one can do.

#### Theorem (Frick and Grohe)

Unless  $\mathbf{P} = \mathbf{NP}$  the the model-checking problem for monadic second-order logic on finite words is not solvable in time  $f(||\psi||) \cdot p(n)$ , for any elementary function f and any polynomial p.



Question Is there something in between which is polynomial?



Question Is there something in between which is polynomial?

Answer Not really...

Question

Is there something in between which is polynomial?

Answer Not really...

#### Theorem (Kreutzer and Tazari)

Let C be a graph class that is closed under subgraphs, and has polylogarithmically unbounded treewidth  $(\log^c(|G_n|) \le tw(G_n))$ . Then given  $G \in C, \psi \in MSO_1$  with  $|\psi|$  as parameter, deciding whether  $G \models \psi$  is not in **XP**, unless the Exponential Time Hypothesis fails.

Question

Is there something in between which is polynomial?

Answer Not really...

#### Theorem (Kreutzer and Tazari)

Let C be a graph class that is closed under subgraphs, and has polylogarithmically unbounded treewidth  $(\log^c(|G_n|) \le tw(G_n))$ . Then given  $G \in C, \psi \in MSO_1$  with  $|\psi|$  being a constant, deciding whether  $G \models \psi$  is not solvable in time  $O(n^{f(|\psi|)})$ , unless something terrible happens.
## Extension of Courcelle's Theorem

The following logics are also solvable in polynomial time on graph classes of bounded treewidth:

- EMSO [Arnborg et al.]
- EMSO-counting [Arnborg et al.]
- MSO-evaluation [Courcelle and Mosbah]

## Takeaway

- MSO<sub>2</sub> and extensions can be model-checked in linear time for graphs of bounded treewidth.
- The running time dependence on the given formula is terrible, and the worst case can not be good.
- This result can not be extended to graph classes where the treewidth is not constant but grows very slowly.
- This gives a certain kind of characterization of MSO<sub>2</sub> model-checking: It is very easy (linear time/logarithmic space) for graphs of bounded treewidth and hard otherwise.

## Takeaway

- MSO<sub>2</sub> and extensions can be model-checked in linear time for graphs of bounded treewidth.
- The running time dependence on the given formula is terrible, and the worst case can not be good.
- This result can not be extended to graph classes where the treewidth is not constant but grows very slowly.
- This gives a certain kind of characterization of MSO<sub>2</sub> model-checking: It is very easy (linear time/logarithmic space) for graphs of bounded treewidth and hard otherwise.
- (It has cool implications in the realm of **FPT**, but more about that on Monday).

## EOF