

EFFICIENT FIRST-ORDER MODEL CHECKING WITH SHORT LABELS

M. M. KANTÉ B. COURCELLE C. GAVOILLE

Université de Bordeaux, LaBRI, CNRS

Frontiers of Algorithmics Workshop
June 19, 2008

- Given a graph property of vertices, assign labels to each vertex for deciding the property by using only the labels.
- A **label** is a bit sequence and we want it to be “**short**” (say of size $\log^{O(k)}(n)$, $n = \text{number of vertices}$).
- For adjacency we need labels of size at least $\Omega(n)$ in general.
- Two possible questions:
 - ▶ Given a property for what **classes of graphs** can we achieve $O(\log(n))$?
 - ▶ Given a class of graphs for what **properties** can we achieve $O(\log(n))$?



LABELING SCHEME

- Let $P(x_1, \dots, x_p, Y_1, \dots, Y_q)$ be a graph property (Y_i are sets).
- An $f(n)$ -labeling scheme for property P of graphs in a class \mathcal{C} is a pair of algorithms $(\mathcal{A}, \mathcal{B})$ such that:
 - ▶ For all $G \in \mathcal{C}$, \mathcal{A} constructs a labeling $J : V(G) \rightarrow \{0, 1\}^*$ such that $|J(x)| \leq f(n)$ for each $x \in V(G)$, $n = |V(G)|$.
 - ▶ \mathcal{B} checks whether G satisfies $P(a_1, \dots, a_p, U_1, \dots, U_q)$ by using $J(a_1), \dots, J(a_p), J(U_1), \dots, J(U_q)$ where $J(U) = \{J(x) \mid x \in U\}$.

NOTE

$J(x)$ determines x .



SOME RESULTS

- [Adjacency] There exist $O(\log(n))$ -labeling schemes for the classes:
 - bounded arboricity (planar, bounded tree-width, ...),
 - bounded clique-width,
 - interval graphs (unbounded clique-width and unbounded arboricity).
- [Distance (static)] $O(n)$ in general on n -vertex graphs.
 - $O(\log^2(n))$ for trees, bounded-clique-width, bounded tree-width,
 - between $\Omega(n^{1/3})$ and $O(n^{1/2})$ for planar graphs,
 - $O(\log(n))$ for interval graphs.



1 CLIQUE-WIDTH

2 Logic

3 Locally Decomposable Graphs

4 Main Results

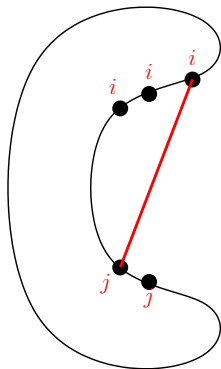
CLIQUE-WIDTH

- **Clique-width** is defined in terms of **few very simple graph operations**.
- Graphs are simple, directed or not.
- **k -graph** = each vertex has one and only one color from $\{1, \dots, k\}$.
- several vertices may have the same color.
- **One binary operation**: \oplus = disjoint union of k -graphs.
- Notice that **$G \oplus G \neq G$** .

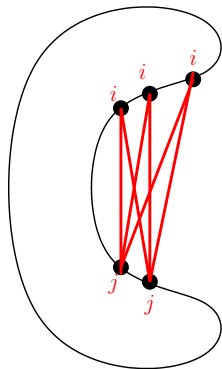


CLIQUE-WIDTH (UNARY OPERATIONS)

- $add_{i,j}(G)$ is the (undirected) k -graph where we add to G edges between vertices colored by i and vertices colored by j .



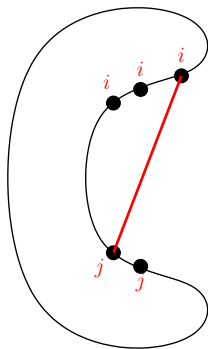
G



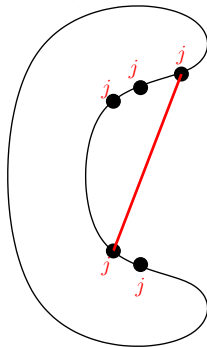
$H = add_{i,j}(G)$

CLIQUE-WIDTH (UNARY OPERATIONS)

- $add_{i,j}(G)$ is the (undirected) k -graph where we add to G edges between vertices colored by i and vertices colored by j .
- $ren_{i \rightarrow j}(G)$ is the graph where we relabel every vertex of G colored by i into j . Then there is **no vertex colored by i** in $ren_{i \rightarrow j}(G)$.



G



$H = ren_{i \rightarrow j}(G)$

CLIQUE-WIDTH (UNARY OPERATIONS)

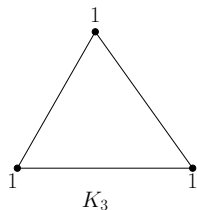
- $add_{i,j}(G)$ is the (undirected) k -graph where we add to G edges between vertices colored by i and vertices colored by j .
- $ren_{i \rightarrow j}(G)$ is the graph where we relabel every vertex of G colored by i into j . Then there is **no vertex colored by i** in $ren_{i \rightarrow j}(G)$.
- **k -Basic graphs** are k -graphs with a single vertex: (i for $i \in [k]$).

DEFINITION

The clique-width of a graph G , denoted by $cwd(G)$, is the minimum k such that G is constructed from k -basic graphs with the operations \oplus , $add_{i,j}$ and $ren_{i \rightarrow j}$ for $i, j \in [k]$.



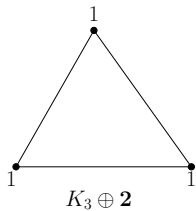
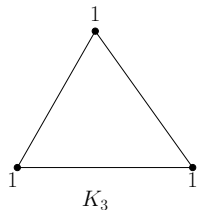
- Cliques have clique-width 2:



EXAMPLES

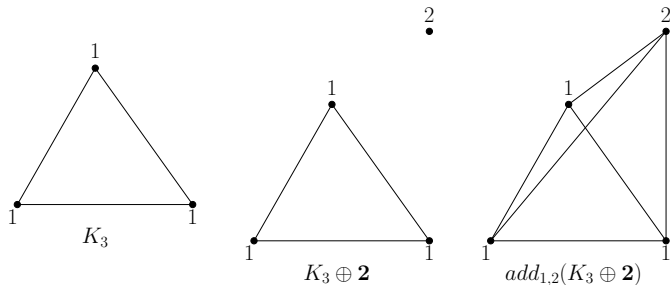
- Cliques have clique-width 2:

2
•



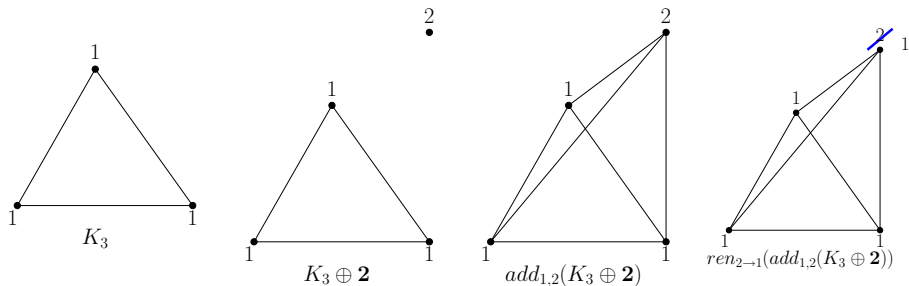
EXAMPLES

- Cliques have clique-width 2:



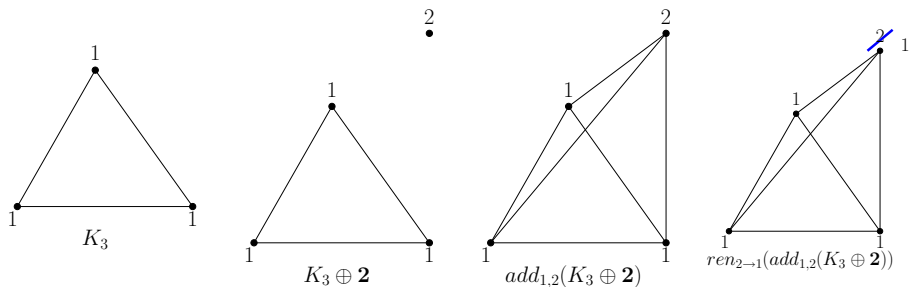
EXAMPLES

- Cliques have clique-width 2:



EXAMPLES

- Cliques have clique-width 2:



- Distance hereditary graphs have clique-width ≤ 3 .
- $cwd(G) \leq 3 \cdot 2^{twd(G)-1}$.
- Planar graphs, interval graphs have unbounded clique-width.

OUTLINE

1 Clique-Width

2 **LOGIC**

3 Locally Decomposable Graphs

4 Main Results

NOTATIONS

- Graph $G = \langle V, E \rangle$ where V is the set of vertices and $E(\cdot, \cdot)$ is the adjacency relation.
- $\varphi(x_1, \dots, x_m, Y_1, \dots, Y_q)$: free FO variables in $\{x_1, \dots, x_m\}$ and free set variables in $\{Y_1, \dots, Y_q\}$.
- An *FO sentence* is an FO formula without free variables.



DISTANCE AT MOST t IN $G \setminus Y$: $D_t(x, y, Y)$

$$(x = y) \vee \exists x_1 \cdots x_{t+1} \notin Y \left(\bigwedge_{1 \leq i \leq t} E(x_i, x_{i+1}) \wedge x = x_1 \wedge y = x_{t+1} \right).$$

COMMON NEIGHBOR OF x AND y IN Y : $\varphi(x, y, Y)$

$$\exists z \in Y \left(E(x, z) \wedge E(y, z) \right).$$



MSO LOGIC

- MSO logic = FO logic with quantifications over sets of vertices.

THEOREM 1 (COURCELLE AND VANICAT)

Let $P(x_1, \dots, x_m, Y_1, \dots, Y_q)$ be an MSO-definable property and $k \in \mathbb{N}$.

- There exists an $O(f(k) \log(n))$ -labeling scheme $(\mathcal{A}, \mathcal{B})$ for graphs of clique-width at most k .
- \mathcal{A} computes the labels in $O(n^3)$ -time or in $O(n \log(n))$ -time if the decomposition is given.



1 Clique-Width

2 Logic

3 LOCALLY DECOMPOSABLE GRAPHS

4 Main Results

MOTIVATION

MOTIVATION

- Bounded clique-width is limited but is OK for MSO logic.
- MSO logic cannot be OK for unbounded clique-width unless $P=NP$ (3-colorability is MSO-definable and NP-complete on planar graphs).
- Unbounded clique-width may work for FO logic.

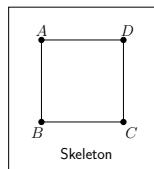
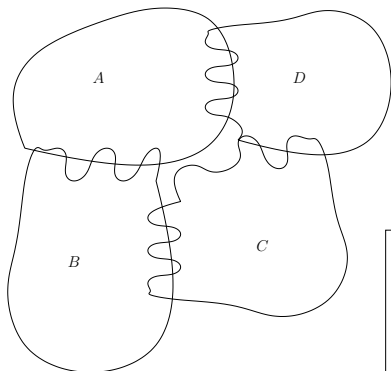
EXAMPLES

Planar Graphs, unit-interval graphs, graphs of bounded degree,



MOTIVATION : NETWORKS BUILT BY COMBINING GRAPHS OF VARIOUS TYPES

- A,B,C,D have clique-width $\leq k$.
- Planar Skeleton
- Skeleton of bounded degree (limited overlaps).
- Concretely ...

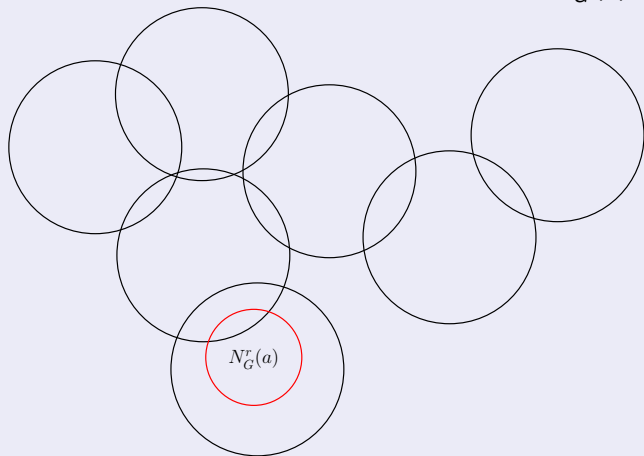


LOCALLY CWD-DECOMPOSABLE

CWD-COVER

Let $r, \ell \geq 1$ and $g : \mathbb{N} \rightarrow \mathbb{N}$. An (r, ℓ, g) -*cwd cover* of G is a family \mathcal{T} of subsets of V_G such that:

- 1 For every $a \in V_G$ there exists $U \in \mathcal{T}$ such that $N_G^r(a) \subseteq U$.

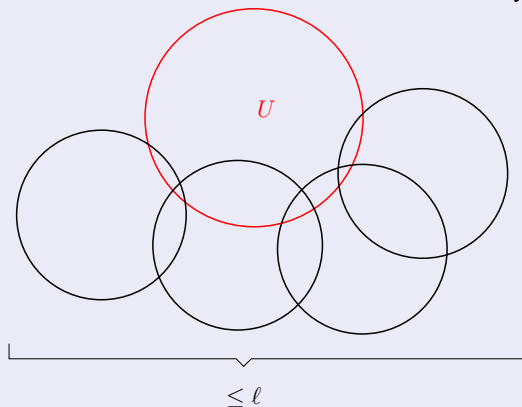


LOCALLY CWD-DECOMPOSABLE

CWD-COVER

Let $r, \ell \geq 1$ and $g : \mathbb{N} \rightarrow \mathbb{N}$. An (r, ℓ, g) -*cwd cover* of G is a family \mathcal{T} of subsets of V_G such that:

- 1 For every $a \in V_G$ there exists $U \in \mathcal{T}$ such that $N_G^r(a) \subseteq U$.
- 2 For each $U \in \mathcal{T}$ there exist less than ℓ many $V \in \mathcal{T}$ such that $U \cap V \neq \emptyset$.

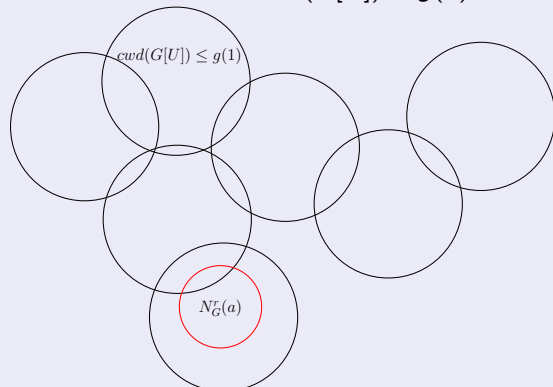


LOCALLY CWD-DECOMPOSABLE

CWD-COVER

Let $r, \ell \geq 1$ and $g : \mathbb{N} \rightarrow \mathbb{N}$. An (r, ℓ, g) -*cwd cover* of G is a family \mathcal{T} of subsets of V_G such that:

- 1 For every $a \in V_G$ there exists $U \in \mathcal{T}$ such that $N_G^r(a) \subseteq U$.
- 2 For each $U \in \mathcal{T}$ there exist less than ℓ many $V \in \mathcal{T}$ such that $U \cap V \neq \emptyset$.
- 3 For each U we have $\text{cwd}(G[U]) \leq g(1)$.



CWD-COVER

Let $r, \ell \geq 1$ and $g : \mathbb{N} \rightarrow \mathbb{N}$. An (r, ℓ, g) -*cwd cover* of G is a family \mathcal{T} of subsets of V_G such that:

- 1 For every $a \in V_G$ there exists $U \in \mathcal{T}$ such that $N_G^r(a) \subseteq U$.
- 2 For each $U \in \mathcal{T}$ there exist less than ℓ many $V \in \mathcal{T}$ such that $U \cap V \neq \emptyset$.
- 3 For each U we have $\text{cwd}(G[U]) \leq g(1)$.

NICE CWD-COVER

An (r, ℓ, g) -*cwd cover* is *nice* if condition 3 is replaced by:

For $q \geq 1$ and all U_1, \dots, U_q in \mathcal{T} we have

$$\text{cwd}(G[U_1 \cup \dots \cup U_q]) \leq g(q).$$

LOCALLY CWD-DECOMPOSABLE

A class \mathcal{C} of graphs is *locally cwd-decomposable* if there is a polynomial time algorithm that given $G \in \mathcal{C}$ and $r \geq 1$,

computes ℓ, g and an (r, ℓ, g) -cwd cover of G where ℓ, g depend only on r .



LOCALLY CWD-DECOMPOSABLE

LOCALLY CWD-DECOMPOSABLE

A class \mathcal{C} of graphs is *locally cwd-decomposable* if there is a polynomial time algorithm that given $G \in \mathcal{C}$ and $r \geq 1$,

computes ℓ, g and an (r, ℓ, g) -cwd cover of G where ℓ, g depend only on r .

NICELY LOCALLY CWD-DECOMPOSABLE

A class \mathcal{C} of graphs is *nicely locally cwd-decomposable* if there is a polynomial time algorithm that given $G \in \mathcal{C}$ and $r \geq 1$,

computes ℓ, g and a nice (r, ℓ, g) -cwd cover of G where ℓ, g depend only on r .



OUTLINE

1 Clique-Width

2 Logic

3 Locally Decomposable Graphs

4 **MAIN RESULTS**

MAIN THEOREM

There exist $O(\log(n))$ -labeling schemes for the following queries and graph classes:

- 1 FO queries without set arguments on locally cwd-decomposable classes.
- 2 FO queries with set arguments on nicely locally cwd-decomposable.

MAIN RESULTS

MAIN THEOREM

There exist $O(\log(n))$ -labeling schemes for the following queries and graph classes:

- 1 FO queries without set arguments on locally cwd-decomposable classes.
- 2 FO queries with set arguments on nicely locally cwd-decomposable.

USING SET ARGUMENTS

In a network, we may specify the failed or forbidden nodes in the formula by using **set variables**.



t -LOCAL FORMULAS

An FO formula $\varphi(x_1, \dots, x_m, Y_1, \dots, Y_q)$ is *t -local around (x_1, \dots, x_m)* if for every G and, every $a_1, \dots, a_m \in V_G$, $W_1, \dots, W_q \subseteq V_G$ we have

$$G \models \varphi(a_1, \dots, a_m, W_1, \dots, W_q)$$

iff

$$G[N] \models \varphi(a_1, \dots, a_m, W_1 \cap N, \dots, W_q \cap N)$$

where $N = N_G^t(a_1, \dots, a_m) = \{y \in V_G \mid d(y, a_i) \leq t \text{ for some } i = 1, \dots, m\}$.

t -LOCAL FORMULAS

An FO formula $\varphi(x_1, \dots, x_m, Y_1, \dots, Y_q)$ is *t -local around (x_1, \dots, x_m)* if for every G and, every $a_1, \dots, a_m \in V_G$, $W_1, \dots, W_q \subseteq V_G$ we have

$$G \models \varphi(a_1, \dots, a_m, W_1, \dots, W_q)$$

iff

$$G[N] \models \varphi(a_1, \dots, a_m, W_1 \cap N, \dots, W_q \cap N)$$

where $N = N_G^t(a_1, \dots, a_m) = \{y \in V_G \mid d(y, a_i) \leq t \text{ for some } i = 1, \dots, m\}$.

REMARK

The query $d(x, y) \leq 2t$ is t -local.

(t, s) -LOCAL SENTENCES

An FO sentence is *basic (t, s) -local* if it is equivalent to a sentence of the form

$$\exists x_1 \cdots \exists x_s \left(\bigwedge_{1 \leq i < j \leq s} d(x_i, x_j) > 2t \wedge \bigwedge_{1 \leq i \leq s} \psi[x_i/x] \right)$$

where $\psi(x)$ is t -local around its unique free variable x .



THEOREM 2

Let $\varphi(\bar{x})$ be a FO formula where $\bar{x} = (x_1, \dots, x_m)$. Then φ is logically equivalent to a Boolean combination $B(\varphi_1(\bar{u}_1), \dots, \varphi_p(\bar{u}_p), \psi_1, \dots, \psi_h)$ where:

- each φ_i is a **t -local formula around $\bar{u}_i \subseteq \bar{x}$** (sub-tuple).
- each ψ_i is a basic **(t', s) -local sentence**.

B can be **computed** and, t, t' and s can be **bounded** in terms of m and the **quantifier-rank of φ** .

VERIFICATION OF BASIC (t, s) -LOCAL SENTENCES

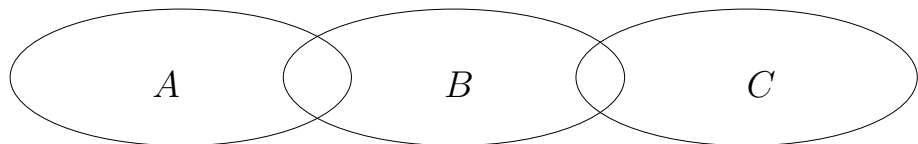
LEMMA 1

Let G be in a locally cwd-decomposable class. Every basic (t, s) -local sentence can be decided in polynomial time on G .

PROOF.

Adaptated from Frick and Grohe. □

HOW TO CHECK DISTANCE AT MOST t



- For each x , $N_G^t(x) \subseteq A$ or B or C .
- Each A, B, C of clique-width $\leq k$.
- By Courcelle-Vanicat, labelings J_A, J_B, J_C for $d(x, y) \leq t$ in $G[A], G[B], G[C]$ respectively.
- $L(x) = (J_A(x), J_B(x), J_C(x))$ ($J_A(x) =$ “undefined” if $x \notin A$).
- $d(x, y) \leq t$ iff some pair $(J_U(x), J_U(y))$ says $d_{G[U]}(x, y) \leq t$ for $U \in \{A, B, C\}$.

A t -FORMULA AROUND ONE FREE FO VARIABLE



$\varphi(x, Y, Z)$ WITH SET VARIABLES Y AND Z

- For $\varphi(x, Y, Z)$ apply Courcelle-Vanicat to each U and get a labeling J_U .
- x may be in many $U \in \{A, B, C\}$ and the same for $N_G^t(x)$.
- For simplicity, let each x in A, B and C .
- $L(x) = (J_A(x), J_B(x), J_C(x))$.
- Given $L(x), L(Y)$ and $L(Z)$ determine the U such that $N_G^t(x) \subseteq U$ and recover $Y \cap U$ and $Z \cap U$.
- $G \models \varphi(x, Y, Z)$ iff $J_U(x)$ says that $G[U] \models \varphi(x, Y \cap U, Z \cap U)$.

t -LOCAL FORMULAS (STRONGER STATEMENT)

LEMMA 2

There exists an $O(\log(n))$ -labeling scheme for t -local formulas with set arguments on locally cwd-decomposable classes.

PROOF.

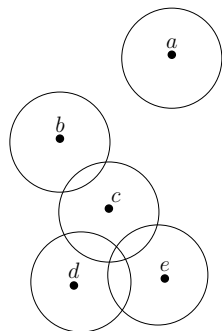
- We will use a decomposition of t -local formulas by Frick.
- Recall that Gaifman Theorem extends to FO formulas with set arguments (handled as colors of vertices; distances do not depend on colors).



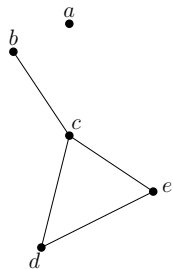
t -DISTANCE TYPE

DEFINITION 1

Let $m, t \geq 1$. The t -distance type of an m -tuple \bar{a} is the undirected graph $\epsilon = ([m], \text{edg}_\epsilon)$ where $\text{edg}_\epsilon(i, j)$ iff $d(a_i, a_j) \leq 2t + 1$.



Balls of radius t



t -distance type

DEFINITION 1

Let $m, t \geq 1$. The t -distance type of an m -tuple \bar{a} is the undirected graph $\epsilon = ([m], \text{edg}_\epsilon)$ where $\text{edg}_\epsilon(i, j)$ iff $d(a_i, a_j) \leq 2t + 1$.

SATISFACTION

The satisfaction of a t -distance type by an m -tuple can be expressed by a t -local formula:

$$\rho_{t,\epsilon}(x_1, \dots, x_m) := \bigwedge_{(i,j) \in \text{edg}_\epsilon} d(x_i, x_j) \leq 2t + 1 \wedge \bigwedge_{(i,j) \notin \text{edg}_\epsilon} d(x_i, x_j) > 2t + 1.$$

DECOMPOSITION OF t -LOCAL FORMULAS

LEMMA 3

Let $\varphi(\bar{x}, Y_1, \dots, Y_q)$ be a t -local formula around $\bar{x} = (x_1, \dots, x_m)$. For each t -distance type ϵ with $\epsilon_1, \dots, \epsilon_p$ as connected components, one can compute a Boolean combination $F^{t,\epsilon}(\varphi_{1,1}, \dots, \varphi_{1,j_1}, \dots, \varphi_{p,1}, \dots, \varphi_{p,j_p})$ of formulas $\varphi_{i,j}$ such that:

- The FO free variables of each $\varphi_{i,j}$ are among $\bar{x} \upharpoonright \epsilon_i$ ($\bar{x} \upharpoonright \epsilon_i$ is the restriction of \bar{x} to ϵ_i) and the set arguments remains in $\{Y_1, \dots, Y_q\}$.
- $\varphi_{i,j}$ is t -local around $\bar{x} \upharpoonright \epsilon_i$.
- For each m -tuple \bar{a} , each q -tuple of sets W_1, \dots, W_q :

$$G \models \rho_{t,\epsilon}(\bar{a}) \wedge \varphi(\bar{a}, W_1, \dots, W_q)$$

iff

$$G \models \rho_{t,\epsilon}(\bar{a}) \wedge F^{t,\epsilon}(\dots, \varphi_{i,j}(\bar{a} \upharpoonright \epsilon_i, W_1, \dots, W_q), \dots).$$

PROOF OF LEMMA 2

- Let \mathcal{T} be an (r, ℓ, g) -cwd cover of G where $r = m(2t + 1)$.
- Label each vertex with a label $K(x)$ of length $O(\log(n))$ for deciding if $d(x, y) \leq 2t + 1$ (Distance at most t - Slide 21).
- Let ϵ be a fixed t -distance type and call the algorithm of Frick (Lemma 3).
- For each $\varphi_{i,j}^\epsilon$, there exists a labeling $J_{i,j,U}^\epsilon$ in $G[U]$ (Courcelle-Vanicat).
- For each x we **append** all these labels $J_{i,j,U}^\epsilon$ in order to get a label J_ϵ .
- There exists at most $k' = 2^{k(k-1)/2}$ t -distance types, we let

$$J(x) = \{\ulcorner x \urcorner, K(x), J_{\epsilon^1}, \dots, J_{\epsilon^{k'}}\}.$$

- It has length $O(\log(n))$ (**Huge Constants**).



PROOF OF LEMMA 2 (VALIDITY)

- Let $J(a_1), \dots, J(a_m)$ and $J(W_1), \dots, J(W_q)$.
- By using $K(a_i)$ we can construct the t -distance type ϵ satisfied by a_1, \dots, a_m . We can then recover $J_\epsilon(a_i)$.
- We let $\epsilon_1, \dots, \epsilon_p$ be the connected components of ϵ .
- For each $\bar{a} \mid \epsilon_j$ there exists at least one $U \in \mathcal{T}$ such that $N_G^t(\bar{a} \mid \epsilon_j) \subseteq U$. (There are less than ℓ .)
- We can now decide whether G satisfies φ by Lemma 3.



- Sizes of labels: $a_{\varphi, \mathcal{T}} \cdot \log(n)$. Decrease $a_{\varphi, \mathcal{T}}$ by concrete constructions avoiding logic.
- Better understanding of locally cwd-decomposable classes.
- Extension to larger classes of graphs.
- Can we extend the logic, or give at least specific constructions for interesting properties like connectivity ?
- **Thank you !**