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 $O(\log(n))$?
FORMALLY

**Labeling Scheme**

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

**Note**

$J(x)$ determines $x$. 

M. M. Kanté (LaBRI)
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.
OUTLINE

1. CLIQUE-WIDTH
2. Logic
3. Locally Decomposable Graphs
4. Main Results
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 $\{1, \ldots, k\}$. 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)**

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

\[
G \\
H = \text{ren}_{i\rightarrow j}(G)
\]

\( H = \text{ren}_{i\rightarrow j}(G) \)
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: \(\textbf{i}\) 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 \(i, j \in [k]\).
**Examples**

- Cliques have clique-width 2:

```
  1
 /|
/  |
1  1
K3
```

Distance hereditary graphs have clique-width ≤ 3.

\[ \text{cwd}(G) \leq 3 \cdot \text{twd}(G) - 1. \]

Planar graphs, interval graphs have unbounded clique-width.
**Examples**

- Cliques have clique-width 2:

\[ K_3 \]

\[ K_3 \oplus 2 \]

Distance hereditary graphs have clique-width \( \leq 3 \).

\[ \text{cwd}(G) \leq 3 \cdot 2^{\text{twd}(G) - 1} \]

Planar graphs, interval graphs have unbounded clique-width.
Examples

- Cliques have clique-width 2:
  - $K_3$
  - $K_3 \oplus 2$
  - $\text{add}_{1,2}(K_3 \oplus 2)$
**Examples**

- Cliques have clique-width 2:

  \[
  K_3 \\
  1 \quad 1
  \]

  \[
  K_3 \oplus 2 \\
  1 \quad 1
  \]

  \[
  add_{1,2}(K_3 \oplus 2) \\
  1 \quad 1
  \]

  \[
  ren_{2\rightarrow 1}(add_{1,2}(K_3 \oplus 2)) \\
  1 \quad 1
  \]

Distance hereditary graphs have clique-width \( \leq 3 \).

\( \text{cwd}(G) \leq 3 \cdot 2^{\text{twd}(G)} - 1 \).

Planar graphs, interval graphs have unbounded clique-width.
**Examples**

- Cliques have clique-width 2:

![Diagram showing cliques and operations]

- Distance hereditary graphs have clique-width $\leq 3$.

- $cwd(G) \leq 3 \cdot 2^{twd(G) - 1}$.

- Planar graphs, interval graphs have unbounded clique-width.
OUTLINE

1. Clique-Width
2. LOGIC
3. Locally Decomposable Graphs
4. Main Results
FO Logic: Examples

Notations

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

**Distance at most** $t$ **in** $G \setminus Y$: $D_t(x, y, Y)$

$$(x = y) \lor \exists x_1 \cdots x_{t+1} \not\in Y \left( \bigwedge_{1 \leq i \leq t} E(x_i, x_{i+1}) \land x = x_1 \land y = x_{t+1} \right).$$

**Common neighbor of** $x$ **and** $y$ **in** $Y$: $\varphi(x, y, Y)$

$$\exists z \in Y \left( E(x, z) \land E(y, z) \right).$$
**Theorem 1 (Courcelle and Vanicat)**

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

- There exists an $O(f(k) \log(n))$-labeling scheme $(A, B)$ for graphs of clique-width at most $k$.

- $A$ computes the labels in $O(n^3)$-time or in $O(n \log(n))$-time if the decomposition is given.
OUTLINE

1. Clique-Width

2. Logic

3. Locally Decomposable Graphs

4. Main Results
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 $(r, \ell, g)$-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 $\ell$ many $V \in \mathcal{T}$ such that $U \cap V \neq \emptyset$.

3. For each $U$ we have $\text{cwd}(G[U]) \leq g(1)$.
Let \( r, \ell \geq 1 \) and \( g : \mathbb{N} \to \mathbb{N} \). An \((r, \ell, g)\)-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 \( \ell \) many \( V \in \mathcal{T} \) such that \( U \cap V \neq \emptyset \).
Let $r, \ell \geq 1$ and $g : \mathbb{N} \to \mathbb{N}$. An $(r, \ell, g)$-cwd cover of $G$ is a family $T$ of subsets of $V_G$ such that:

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

CWD-COVER

Let $r, \ell \geq 1$ and $g : \mathbb{N} \to \mathbb{N}$. An $(r, \ell, g)$-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 $N^r_G(a) \subseteq U$.
2. For each $U \in \mathcal{T}$ there exist less than $\ell$ many $V \in \mathcal{T}$ such that $U \cap V \neq \emptyset$.
3. For each $U$ we have $cwd(G[U]) \leq g(1)$.

NICE CWD-COVER

An $(r, \ell, g)$-cwd cover is nice if condition 3 is replaced by:

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

$$cwd(G[U_1 \cup \cdots \cup U_q]) \leq g(q).$$
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 $\ell, g$ and an $(r, \ell, g)$-cwd cover of $G$ where $\ell, g$ depend only on $r$. 
A class $C$ of graphs is **locally cwd-decomposable** if there is a polynomial time algorithm that given $G \in C$ and $r \geq 1$,

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

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

computes $\ell, g$ and a nice $(r, \ell, g)$-cwd cover of $G$ where $\ell, g$ depend only on $r$. 
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 $\varphi(x_1, \ldots, x_m, Y_1, \ldots, Y_q)$ is *t-local around* $(x_1, \ldots, x_m)$ if for every $G$ and, every $a_1, \ldots, a_m \in V_G$, $W_1, \ldots, W_q \subseteq V_G$ we have

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

iff

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

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

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

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

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

iff

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

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

**Remark**

The query $d(x, y) \leq 2t$ is $t$-local.
An FO sentence is \textit{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 $\bar{x} = (x_1, \ldots, x_m)$. Then $\varphi$ is logically equivalent to a Boolean combination

$$B(\varphi_1(\bar{u}_1), \ldots, \varphi_p(\bar{u}_p), \psi_1, \ldots, \psi_h)$$

where:

- each $\varphi_i$ is a $t$-local formula around $\bar{u}_i \subseteq \bar{x}$ (sub-tuple).
- each $\psi_i$ 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 $\varphi$. 
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.

Adapted from Frick and Grohe.
How to Check Distance at Most $t$

- For each $x$, $N^t_G(x) \subseteq A$ 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 \textit{t-Formula around one Free FO Variable}

\[ \varphi(x, Y, Z) \text{ with set variables } Y \text{ and } Z \]

- 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 \( G[U] \models \varphi(x, Y \cap U, Z \cap U) \).
**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).
**Definition 1**

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

$$\rho_{t,\epsilon}(x_1, \ldots, x_m) := \bigwedge_{(i,j) \in \text{edg}_\epsilon} d(x_i, x_j) \leq 2t + 1 \land \bigwedge_{(i,j) \notin \text{edg}_\epsilon} d(x_i, x_j) > 2t + 1.$$
**Lemma 3**

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

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

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

iff

$$G \models \rho_{t, \epsilon}(\bar{a}) \land F^{t, \epsilon}(\ldots, \varphi_{i,j}(\bar{a} | \epsilon_i, W_1, \ldots, W_q), \ldots).$$
Proof of Lemma 2

- Let $T$ be an $(r, \ell, g)$-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 $\epsilon$ be a fixed $t$-distance type and call the algorithm of Frick (Lemma 3).
- For each $\varphi_{i,j}$, 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 $J_\epsilon$.
- There exists at most $k' = 2^{k(k-1)/2}$ $t$-distance types, we let
  \[ J(x) = \{\lceil x \rceil, K(x), J_\epsilon^1, \ldots, J_\epsilon^{k'} \}. \]
- It has length $O(\log(n))$ (Huge Constants).
Proof of Lemma 2 (Validity)

- Let $J(a_1), \ldots, J(a_m)$ and $J(W_1), \ldots, J(W_q)$.
- By using $K(a_i)$ we can construct the $t$-distance type $\epsilon$ satisfied by $a_1, \ldots, a_m$. We can then recover $J_\epsilon(a_i)$.
- We let $\epsilon_1, \ldots, \epsilon_p$ be the connected components of $\epsilon$.
- For each $\bar{a} \mid \epsilon_i$ there exists at least one $U \in \mathcal{T}$ such that $N_G^t(\bar{a} \mid \epsilon_i) \subseteq U$. (There are less than $\ell$.)
- We can now decide whether $G$ satisfies $\varphi$ by Lemma 3.
Sizes of labels: \( a_\varphi, T \cdot \log(n) \). Decrease \( a_\varphi, T \) 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!