Interpretations – how to speak about sparsity in dense graphs

Jakub Gajarský

TU Berlin

November 30, 2018

Prague

Interpretations

$$G \xrightarrow{\psi} H$$

Interpretations:

- We start with a graph G and produce new graph H
- The new vertex set is determined by $\nu(x)$
- The new edge set is determined by $\psi(x, y)$

We will usually be in the situation when we want to a given graph H produce its 'pre-image' G. In this case we will often use colors.

$$G \xrightarrow{\psi} H$$

Research program

$$\mathcal{C} \xrightarrow{\psi} \mathcal{D}$$

For any sparse graph class $\mathcal{C}$:

- Characterize graph classes interpretable in $\mathcal{C}$
- Find an algorithm to ‘reverse’ interpretations
- Find a model checking algorithm

Research program

For any sparse graph class $\mathcal{C}$:

- Characterize graph classes interpretable in $\mathcal{C}$
- Find an algorithm to 'reverse' interpretations
- Find a model checking algorithm

Done for bounded degree and partially for bounded expansion.

The plan

Today:

- Part I: Interpretations of graphs of bounded degree
- Part II: Shrub-depth – interpretations of trees of bounded height
- Part III: Interpretations of bounded expansion classes

Sparse graphs

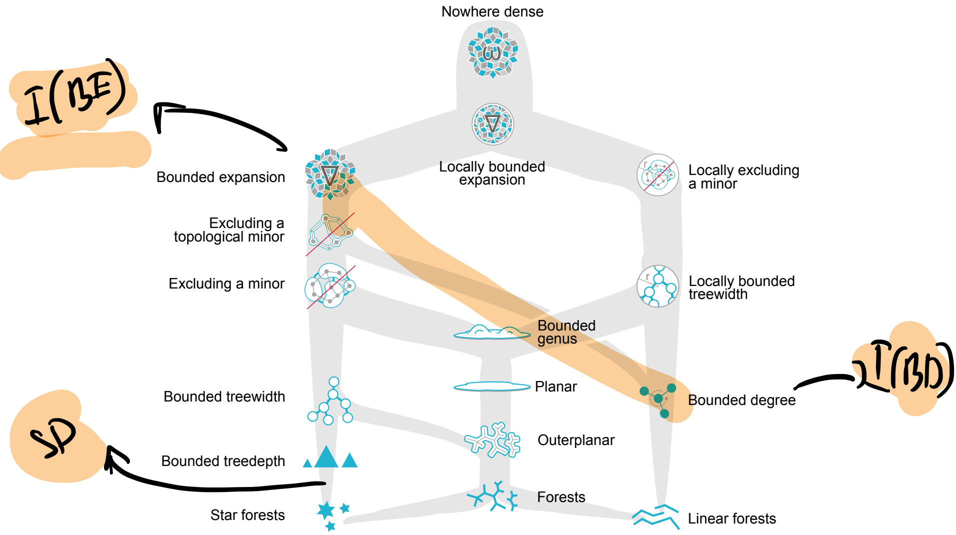


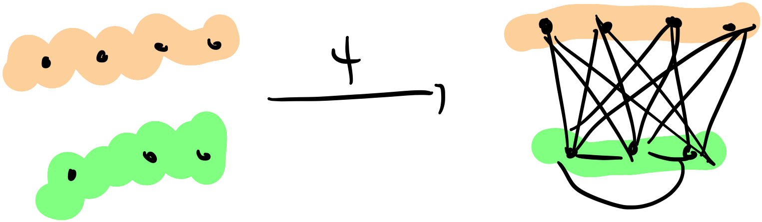
Image by Felix Reidl

Part I: Interpretations of graphs of bounded degree

Interpretations of edgeless graphs with colors

Observation – no matter what the interpretation formula says, it will hold that:

- Each color will end up being a clique or independent set
- Between each two colors there will either be all edges or no edges



Neighborhood diversity

For a graph G, we say that two vertices $u, v \in V(G)$ are twins if $N(u) \backslash v = N(v) \backslash u$

Neighborhood diversity

For a graph G, we say that two vertices $u, v \in V(G)$ are twins if $N(u) \backslash v = N(v) \backslash u$

The twin relation is an equivalence relation on $V(G)$.

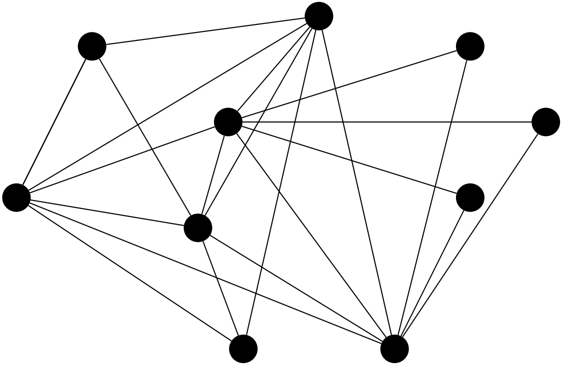
Neighborhood diversity

For a graph G, we say that two vertices $u, v \in V(G)$ are twins if $N(u) \backslash v = N(v) \backslash u$

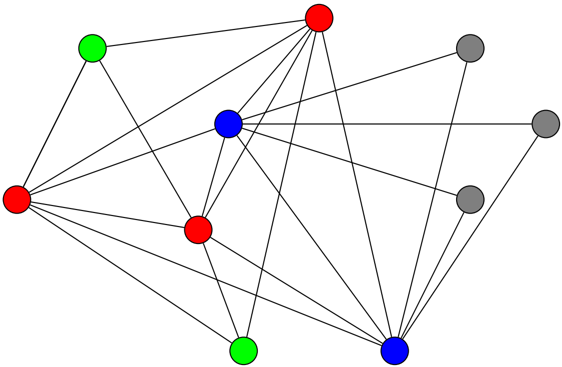
The twin relation is an equivalence relation on $V(G)$.

Neighborhood diversity of a graph G is the number of equivalence classes of the twin relation.

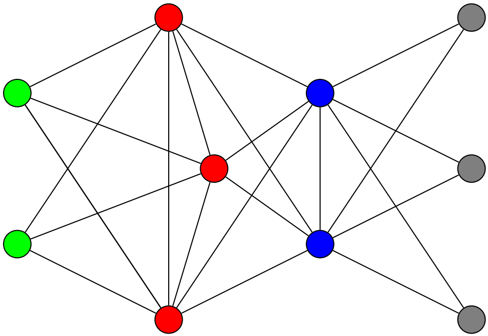
Neighbourhood diversity



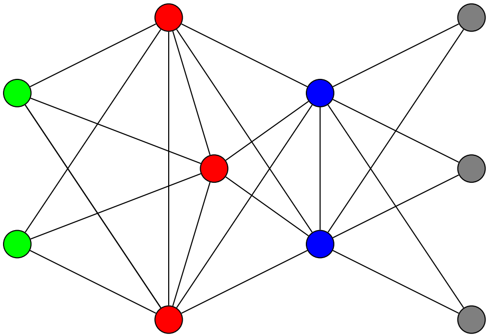
Neighbourhood diversity



Neighbourhood diversity



Neighbourhood diversity



Observe that equivalence class induces either a clique or an independent set.

Interpretations and neighborhood diversity

Theorem

The following are equivalent for a class $\mathcal{C}$ of graphs:

- *Each graph in $\mathcal{C}$ has neighborhood diversity at most m*
- *Each graph in $\mathcal{C}$ is interpretable in an edgeless graph with at most m colors*

Interpretations and neighborhood diversity

Theorem

The following are equivalent for a class $\mathcal{C}$ of graphs:

- *Each graph in $\mathcal{C}$ has neighborhood diversity at most m*
- *Each graph in $\mathcal{C}$ is interpretable in an edgeless graph with at most m colors*

We care about this because graphs interpretable in graphs of bounded degree look approximately like graphs of small neighborhood diversity.

Analysing interpretations

Interpretation $\psi(x, y)$ can be "decomposed" into two parts:

- global part: if two vertices u and v are far apart in G, then whether $\psi(u, v)$ holds depends only on (local) types of u and v.
- local part: if two vertices u and v are close in G, then whether $\psi(u, v)$ holds depends on how they interact.

Analysing interpretations

Interpretation $\psi(x, y)$ can be "decomposed" into two parts:

- global part: if two vertices u and v are far apart in G, then whether $\psi(u, v)$ holds depends only on (local) types of u and v.
- local part: if two vertices u and v are close in G, then whether $\psi(u, v)$ holds depends on how they interact.

Example: $\psi(x, y) \equiv x \neq y \wedge$ (x has degree 1 and y has a neighbor of degree 2 or vice versa) $\vee$ (x and y are at distance at most 2 and there is a vertex of degree exactly 4 on the path from x to y)

Analysing interpretations

Example: $\psi(x, y) \equiv x \neq y \wedge$ (x has degree 1 and y has a neighbor of degree 2 or vice versa) $\vee$ (x and y are at distance at most 2 and there is a vertex of degree exactly 4 on the path from x to y)

(x has degree 1 and y has a neighbor of degree 2 or vice versa) $=$ one of them has degree 1 and the other has a neighbor of degree 2

Global part: there are four types of vertices:

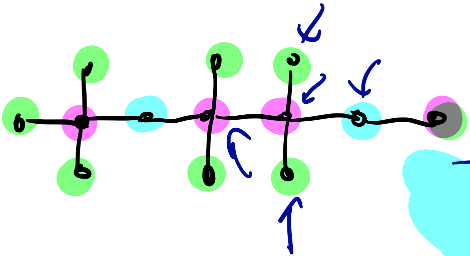
1. Vertices having degree 1
2. Vertices having a neighbor of degree 2
3. Vertices having both properties
4. Vertices having neither property

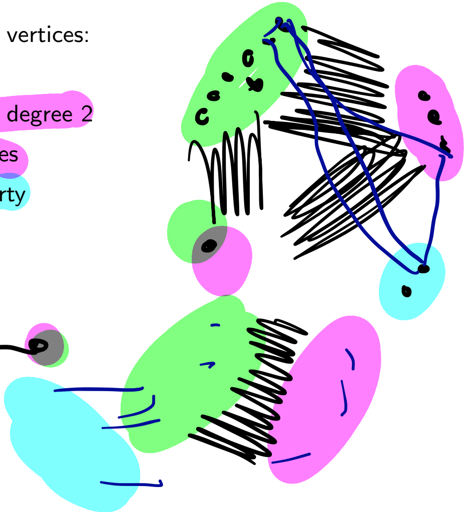
Analysing interpretations

Example: $\psi(x, y) \equiv x \neq y \wedge$ (x has degree 1 and y has a neighbor of degree 2 or vice versa) $\vee$ (x and y are at distance at most 2 and there is a vertex of degree exactly 4 on the path from x to y)

Global part: there are four types of vertices:

1. Vertices having degree 1
2. Vertices having a neighbor of degree 2
3. Vertices having both properties
4. Vertices having neither property





Analysing interpretations

Analysing interpretations

Typically there will be many more types of vertices, but always finitely many (the number is determined by $\psi(x, y)$ and independent of G)

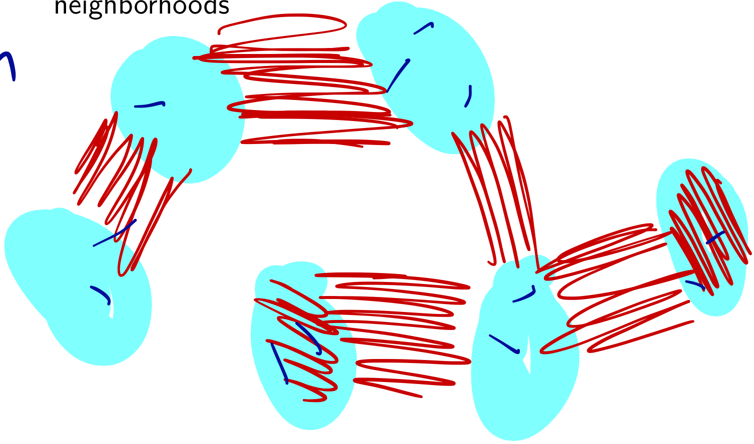
Consequence - **the structure of a graph H interpreted in bounded degree graph G:**

- We can partition $V(H)$ into m parts and we will have:
- Global edges: between any two parts there will be basically all or no edges
- Local edges: every vertex can have some exceptions caused by the local part of the interpretations, but only a bounded number of them

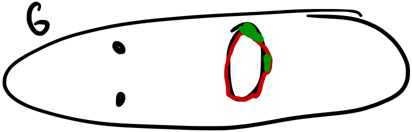
Structure of a graph intepreted in a graph of bounded degree

- graph of small neighborhood diversity (global structure)
- small number of exceptions – from locality and small neighborhoods

m



Near-k-twin relation

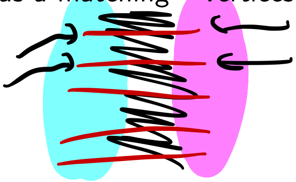


Definition

Two vertices $u, v \in V(G)$ are *near-k-twins* if $|N(u) \Delta N(v)| \leq k$, i.e. their neighborhoods differ in at most k vertices.

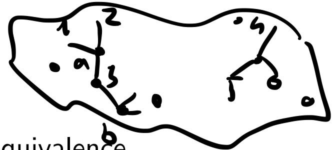
Notation: u, v are near-k-twins – $\{u, v\} \in \rho_k$

Example: complete bipartite graph minus a matching – vertices from the same parts are near-2-twins



Near-k-twin relation

$k = 2d$



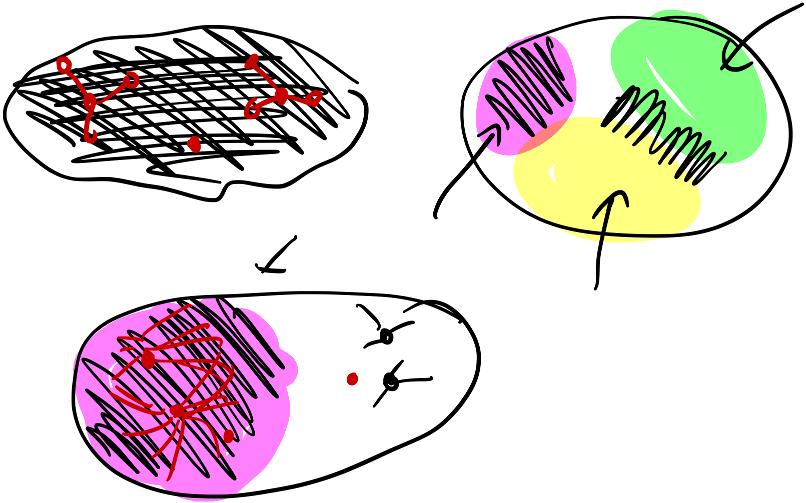
Basic properties:

- ρ_k is always reflexive and symmetric
- ρ_k is **not** necessarily an equivalence
- ρ_k equivalence does **not** imply ρ_{k+1} equivalence

We are mainly interested in situations when ρ_k on $V(G)$ is an equivalence with small number of classes.

Examples: graphs of degree at most d, complements of graphs of degree at most d, several subset complementations of graphs of degree at most d.

Near-k-twin equivalence with small number of classes



If ρ_k is an equivalence ...

We want ρ_k on $V(G)$ to be an **equivalence with small number of classes**.

This leads us to the following definition:

Definition

Graph class $\mathcal{C}$ is *near uniform* if there exists k_0 and m such that for every for every $G \in \mathcal{C}$:

- the relation ρ_k is an equivalence for at least one $k \in \{1, \ldots, k_0\}$
- For this k the equivalence has at most m classes

Near uniform graph classes and interpretations

Theorem

Graph class $\mathcal{D}$ is interpretable in class of graphs of degree at most d if and only if and only if $\mathcal{D}$ is near-uniform.

We split the proof into two directions:

- Near uniformity implies interpretability in small degree
- Interpretability in small degree implies near uniformity

Near uniformity implies interpretability

When ρ_k is an equivalence:

Lemma (Informally)

Informally: If ρ_k is an equivalence, then for every two large enough equivalence classes (in terms of k) there are almost all or almost no edges between these classes, and every vertex has bounded number (in terms of k) exceptions to this rule.

Near uniformity implies interpretability

When ρ_k is an equivalence:

Lemma (Informally)

Informally: If ρ_k is an equivalence, then for every two large enough equivalence classes (in terms of k) there are almost all or almost no edges between these classes, and every vertex has bounded number (in terms of k) exceptions to this rule.

Corollary

Graph for which ρ_k is an equivalence consists of:

- *Large and almost homogeneous groups of vertices*
- *Small groups of vertices connected arbitrarily to the rest of the graph*

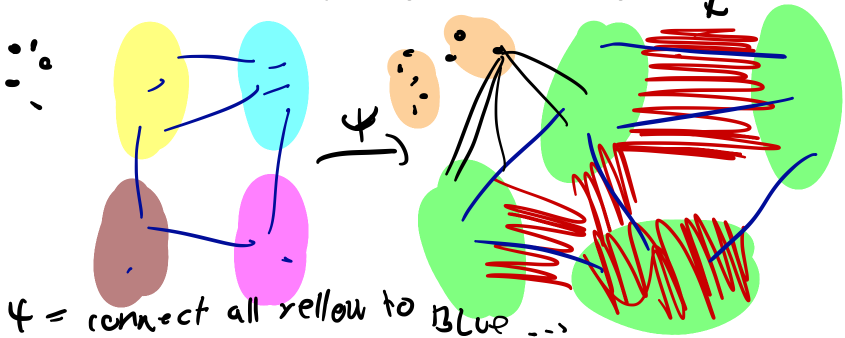
ρ_k an equivalence implies interpretability

Corollary

Graph for which ρ_k is an equivalence consists of:

- *Large and almost homogeneous groups of vertices*
- *Small groups of vertices connected arbitrarily to the rest of the graph*

This implies interpretability in a graph of small degree.



φ = connect all yellow to Blue ...

Interpretability implies near uniformity

Formally, the "decomposition" of the interpretation formula into global and local types follows from from Gaifman's theorem.

The corollary we will use:

Lemma

Let $\psi(x, y)$ be an interpretation formula and let u and v be two vertices of the same logical type. There exists $r \in \mathbb{N}$ such that for every vertex w which has distance at least r from both u and v it holds that $G \models \psi(u, w)$ if and only if $G \models \psi(v, w)$.

Interpretability implies near uniformity

Formally, the "decomposition" of the interpretation formula into global and local types follows from from Gaifman's theorem.

The corollary we will use:

Lemma

Let $\psi(x, y)$ be an interpretation formula and let u and v be two vertices of the same logical type. There exists $r \in \mathbb{N}$ such that for every vertex w which has distance at least r from both u and v it holds that $G \models \psi(u, w)$ if and only if $G \models \psi(v, w)$.

Consequence:
Let $H = I_\psi(G)$. If u and v had the same type in G and if w was far away from both u and v in G, then w is adjacent to both u and v in H or none of them.

Interpretability implies near uniformity

$G \xrightarrow{\psi} H$

Near uniform

Let G be of degree at most d and let $H = I_\psi(G)$.

ψ

We know that: If u and v had the same type in G and if w was at distance more than r from both u and v in G, then w is adjacent to both u and v in H or none of them.

Crucial observation: Two vertices of H which were of the **same type in** G will have almost identical neighborhoods in H – they may disagree on at most constantly many (depending on d and r) vertices in H (these are the vertices which were close to them in G). This means that they will be **near-twins in** H.

Finishing the proof:

- Partition the vertices of G based on their type
- This partition corresponds to the near-k-twin equivalence

Wrapping it up

Algorithmic remark: near-k-twin relation can be computed quickly, an so to a given graph H we can compute its bounded degree pre-image efficiently.

Takeaway:

- Interpretations have a global and local part
- After identifying and eliminating the global part the resulting graph is (usually) going to be locally simple
- Complete characterization of graph classes interpretable in bounded degree graphs

Part II: Interpretations of trees of small height = shrub-depth

Shrub-depth

Fix d (depth) and set C of colors of size m

Fix a relation $R \subseteq C \times C \times \{2, 4, \ldots, 2d\}$

Consider a tree T of depth d such that every leaf has one of m colors

Such a tree is called a *tree-model* and it defines graph G as follows:

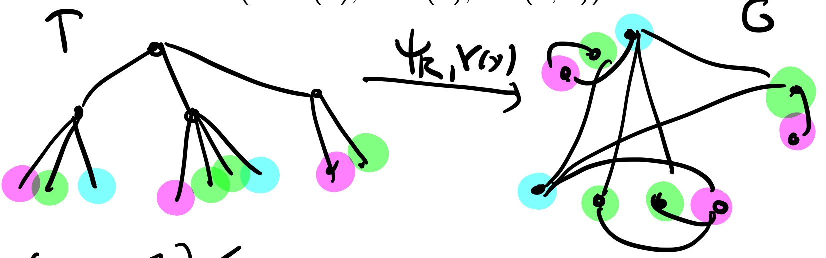
- Vertices of G are leaves of T
- There is an edge between u and v if

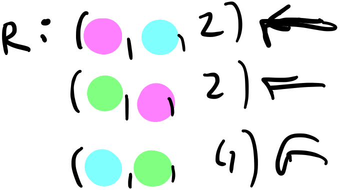
$$(color(u), color(v), dist(u, v)) \in R$$

Shrub-depth

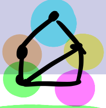
- Vertices of G are leaves of T
- There is an edge between u and v if

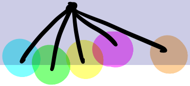
$$(color(u), color(v), dist(u, v)) \in R$$





Shrub depth





Notice that every graph G has a tree model of depth 1 and at most $|V(G)|$ colors

Consequence – it does not make sense to speak about depth of a single graph; depth has to be defined for classes of graphs

Definition

Shrub depth of a class $\mathcal{C}$ of graphs is the least d such that there exists m such that every graph in $\mathcal{C}$ has a tree model of depth d with at most m colors.

Theorem

Class $\mathcal{C}$ of graphs has bounded shrubdepth if and only if it is interpretable in the class of trees of bounded height

Shrub depth

Theorem

Class $\mathcal{C}$ of graphs has bounded shrubdepth if and only if it is interpretable in the class of trees of bounded height

Shrub depth

Theorem

Class $\mathcal{C}$ of graphs has bounded shrubdepth if and only if it is interpretable in the class of trees of bounded height

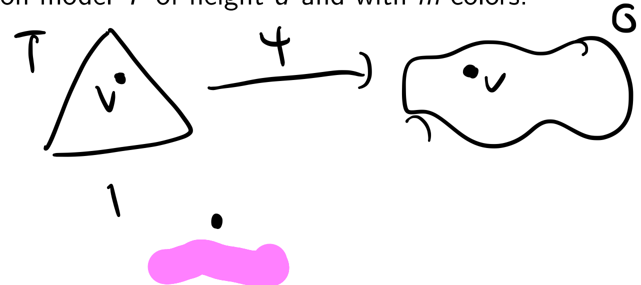
- $\Rightarrow$: simple, $\psi(x, y)$ expresses the (finite) instructions encoded by relation R and $\nu(x)$ just says that x is a leaf
- $\Leftarrow$: technical, see next slides

Shrub depth

Lemma (The hard direction of our theorem)

Every graph class interpretable in trees of bounded height has bounded shrub-depth

What we need to show:
Let $\mathcal{C}$ be a class of graphs interpretable in a class of trees of height h. There exist d and m such that for every $G \in \mathcal{C}$ there is a connection model T of height d and with m colors.



Shrub depth

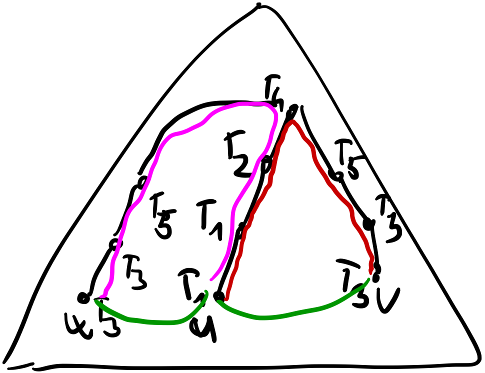
Lemma (The hard direction of our theorem)

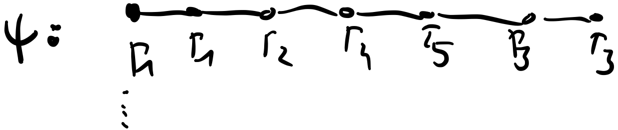
Every graph class interpretable in trees of bounded height has bounded shrub-depth

What we need to show:
Let $\mathcal{C}$ be a class of graphs interpretable in a class of trees of height h. There exist d and m such that for every $G \in \mathcal{C}$ there is a connection model T of height d and with m colors.

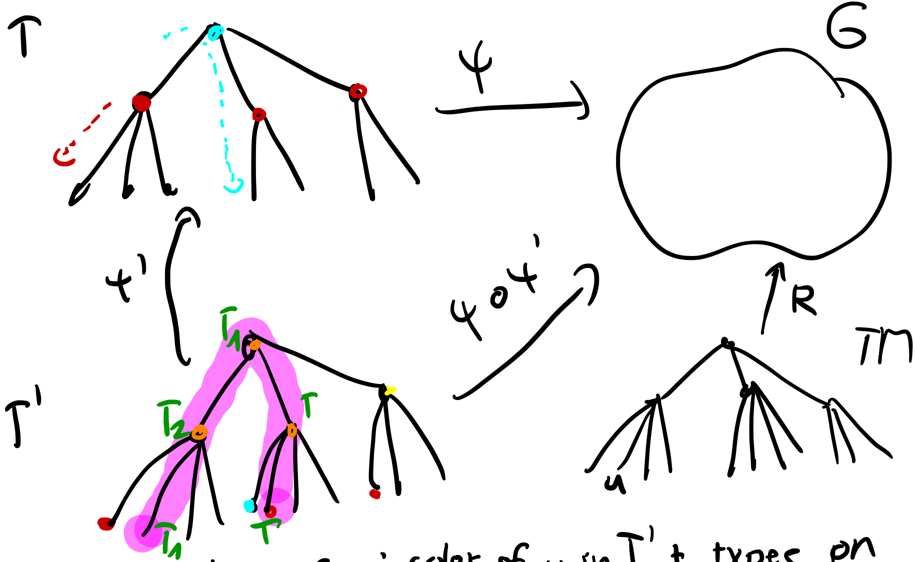
What we will use:
Let T be a tree of height h and let $\psi(x, y)$ be an interpretation formula. Whether $\psi(u, v)$ holds in T depends only on the types of u and v and on the types of vertices on the (short and unique) path between u and v.

Shrub depth





Shrub depth



color of u: color of u in T' + types on the path to the root

Wrapping it up

What you should know/remember:

- Bounded shrubdepth ⇒ bounded clique-width (shrubdepth is 'clique-depth')
- Class $\mathcal{C}$ of graphs has bounded shrubdepth if and only if it is interpretable in a class of colored trees.
- Bounded treedepth implies bounded shrubdepth.
- Class $\mathcal{C}$ of graphs of bounded shrubdepth has bounded treedepth if it does not contain arbitrary large complete bipartite graphs as subgraphs.
- Shrub-depth is closed under inderpretations.

Conjecture: Class $\mathcal{C}$ of graphs has bounded shrub depth if and only if it is not possible interpret arbitrary long induced paths in $\mathcal{C}$.

Part III: Interpretations of bounded expansion graph classes

Interpretations of bounded expansion graph classes

Definition

A class $\mathcal{D}$ of graphs has *structurally bounded expansion* if there exists a class $\mathcal{C}$ of graphs of bounded expansion and an interpretation I such that $\mathcal{D} \subseteq I(\mathcal{C})$

Interpretations of bounded expansion graph classes

Definition

A class $\mathcal{D}$ of graphs has *structurally bounded expansion* if there exists a class $\mathcal{C}$ of graphs of bounded expansion and an interpretation I such that $\mathcal{D} \subseteq I(\mathcal{C})$

Our goal: find a structural/combinatorial characterization of graph classes with structurally bounded expansion

Low treedepth covers

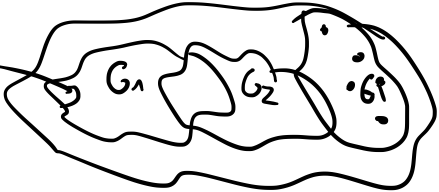
Definition

Class $\mathcal{C}$ of graphs has *low treedepth covers* if for every k there exist N and t such that for each $G \in \mathcal{C}$ we there is a system $G_1, \ldots, G_N$ of induced subgraphs of G such that:

- Each G_i has treedepth at most t
- Each k-tuple of vertices is in at least one G_i

TD $\leq t$

N



Low treedepth covers

Definition

Class $\mathcal{C}$ of graphs has *low treedepth covers* if for every k there exist N and t such that for each $G \in \mathcal{C}$ we there is a system $G_1, \ldots, G_N$ of induced subgraphs of G such that:

- Each G_i has treedepth at most t
- Each k-tuple of vertices is in at least one G_i

Lemma

Class $\mathcal{C}$ of graphs has bounded expansion if and only if it has low treedepth covers.

Low shrub-depth covers

Dense counterpart of low treedepth covers:

Definition

Class $\mathcal{C}$ of graphs has *low shrubdepth covers* if for every k there exist N and t such that for each $G \in \mathcal{C}$ we there is a system $G_1, \ldots, G_N$ of induced subgraphs of G such that:

- Each G_i has shrubdepth at most t
- Each k-tuple of vertices is in at least one G_i

Low shrub-depth covers

Dense counterpart of low treedepth covers:

Definition

Class $\mathcal{C}$ of graphs has *low shrubdepth covers* if for every k there exist N and t such that for each $G \in \mathcal{C}$ we there is a system $G_1, \ldots, G_N$ of induced subgraphs of G such that:

- Each G_i has shrubdepth at most t
- Each k-tuple of vertices is in at least one G_i

Low shrub-depth covers

Dense counterpart of low treedepth covers:

Definition

Class $\mathcal{C}$ of graphs has *low shrubdepth covers* if for every k there exist N and class $\mathcal{S}$ of bounded shrubdepth such that for each $G \in \mathcal{C}$ we there is a system $G_1, \ldots, G_N$ of induced subgraphs of G such that:

- Each G_i is from $\mathcal{S}$
- Each k-tuple of vertices is in at least one G_i

Main result

Theorem

Class $\mathcal{C}$ of graphs has structurally bounded expansion if and only if it has low shrubdepth covers.

$$
\begin{tikzcd}
\textit{bounded treedepth} \arrow[r, "\text{interpretation}"] \arrow[d, "\text{cover}"] & \begin{array}{c}\textit{interpretation of bounded treedepth} \\ =_{(1)} \textit{bounded shrubdepth}\end{array} \arrow[d, "\text{cover"] \\
\begin{array}{c}\textit{bounded treedepth covers} \\ =_{(2)} \textit{bounded expansion}\end{array} \arrow[r, "\text{interpretation}"] & \begin{array}{c}\textit{low shrubdepth covers} =_{(\star)} \\ \textit{structurally bounded expansion}\end{array}
\end{tikzcd}
$$

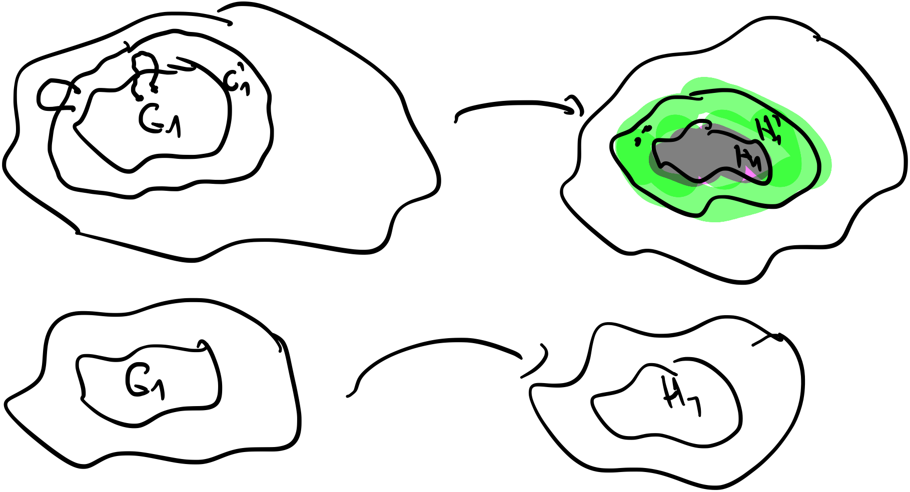
Proof overview

We split the proof in two parts:

1. Structurally bounded expansion implies low shrubdepth covers
2. Low shrubdepth covers imply structurally bounded expansion

Structurally bounded expansion ⇒ low shrubdepth covers

What we need to prove: If we start with a graph G from bounded expansion graph class and apply an interpretation then the resulting graph has a low shrubdepth cover.



Structurally bounded expansion $\Rightarrow$ low shrubdepth covers

What we need to prove: If we start with a graph G from bounded expansion graph class and apply an interpretation then the resulting graph has a low shrubdepth cover.

More precisely: For every bounded expansion graph class $\mathcal{C}$ and every interpretation I it holds: for every k there is N and class of bounded shrubdepth $\mathcal{S}$ such that every $H \in I(\mathcal{C})$ can be k-covered by graphs $H_1, \ldots, H_N$ from $\mathcal{S}$.

Structurally bounded expansion $\Rightarrow$ low shrubdepth covers

What we need to prove: If we start with a graph G from bounded expansion graph class and apply an interpretation then the resulting graph has a low shrubdepth cover.

More precisely: For every bounded expansion graph class $\mathcal{C}$ and every interpretation I it holds: for every k there is N and class of bounded shrubdepth $\mathcal{S}$ such that every $H \in I(\mathcal{C})$ can be k-covered by graphs $H_1, \ldots, H_N$ from $\mathcal{S}$.

Outline:

- Since $\mathcal{C}$ has bounded expansion it has low treedepth covers
- For any k there exists N and class $\mathcal{T}$ of bounded treedepth such that for each $G \in \mathcal{D}$ is covered by graphs $G_1, \ldots, G_N$ from $\mathcal{T}$
- Idea – apply I to each G_i to obtain graphs H_i of bounded shrubdepth (coming from $I(\mathcal{T})$) which cover H – does not quite work

Structurally bounded expansion $\Rightarrow$ low shrubdepth covers

Idea – apply I to each $G_i \in \mathcal{T}$ to obtain graphs H_i of bounded shrubdepth (coming from $I(\mathcal{T})$) which cover H – does not quite work

Let G be a graph, U subset of $V(G)$ and I an interpretation. In general it does *not* hold that

$$I(G)[U] = I(G[U])$$

because the interpretation can use information outside $G[U]$ to decide whether two vertices are adjacent. (Take $U = V(G_i)$ to see the problem.)

Structurally bounded expansion $\Rightarrow$ low shrubdepth covers

Result of Dvořák, Král', Thomas (quantifier elimination):
On a class of bounded expansion every formula $\psi(x_1, \ldots, x_k)$ is equivalent to a quantifier-free formula $\psi'(x_1, \ldots, x_k)$

Consequence:
Whether $\psi(u, v)$ is true in a graph depends only on a small number (fixed for ψ) of vertices reachable from u and v by function applications

Let S_v denote the set of vertices of G reachable by function applications from v. It holds that $|S_v| \leq c$, where c depends on ψ.

Structurally bounded expansion ⇒ low shrubdepth covers

We want a p-shrubdepth cover of H:

1. Let G be such that $H = I_{\psi(x,y)}(G)$. We take a $(p \cdot c)$-treedepth cover $G_1, \ldots, G_N$ of G.
2. In each G_i we consider set W of vertices v such that all c relevant vertices for v are also in G_i, i.e. $S_v \subseteq W$.
3. We consider $I_\psi(G_i)$ and look at what happens with the set W – the graph induced by it has small shrubdepth (since $I_\psi(G_i)$ has small shrubdepth) any two vertices in W have all the information necessary to determine whether $\psi(x, y)$ holds, and so the adjacency in $I_\psi(G_i)[W]$ is the same as in $I_\psi(G_i)$

Low shrubdepth covers $\Rightarrow$ structurally bounded expansion

We need to show: If we can cover graph H by subgraphs of small shrubdepth then it is interpretable in a graph G from a class of graphs of bounded expansion.

Idea: Take a 2-shrubdepth cover of H consisting of graphs $H_1, \ldots, H_N$. These are inerpretable in trees $T_1, \ldots, T_N$ of bounded height. Define *sparsify*(H) to be the union of these trees.

Surprisingly, this simple idea works (i.e. it produces a graph class of bounded expansion)

Low shrubdepth covers $\Rightarrow$ structurally bounded expansion

Let $\mathcal{D}$ be a graph class with low shrubdepth covers.
How do we show that the class

$$\mathit{Sparsify}(\mathcal{D}) = \{\mathit{Sparsify}(H) \mid H \in \mathcal{D}\}$$

is of bounded expansion?

1. We show that if we start with a class with low shrub-depth covers and apply an interpretation, then the resulting graph class has low shrub-depth covers (quantifier elimination again)
2. We show that $\mathit{Sparsify}(\mathcal{D})$ is actually an interpretation of $\mathcal{D}$ and so by (1) it also has low shrub-depth covers
3. We observe that $\mathit{Sparsify}(\mathcal{D})$ does not have large cliques or bi-cliques
4. Combining (2) and (3) implies that $\mathit{Sparsify}(\mathcal{D})$ has low treedepth covers

Transductions

Show that sparsification is a FO interpretation (crucial step, involves showing that one can define in G its decompositionl using FO logic)

The issue: Tree-model has more vertices than the graph it defines. How do we interpret it in the graph?

Transductions

Show that sparsification is a FO interpretation (crucial step, involves showing that one can define in G its decompositionl using FO logic)

The issue: Tree-model has more vertices than the graph it defines. How do we interpret it in the graph?

We consider *transductions*. A transduction is a sequence of the following atomic operations:

- Creating new relations using formulas
- Removing relations
- Restricting the universe (deleting vertices)
- Adding arbitrary colors
- Copying
- Adding definable unary functions

Conclusions

Consequence of our result:
If we are given together with a graph H its 2-shrubdepth cover then we can perform FO model checking quickly on classes of graphs with structurally bounded expansion.

Main open questions:

- Computing low shrubdepth covers (it is sufficient to a compute 2-shrubdepth cover)
- Different characterizations of structurally bounded expansion
- Characterizing interpretations of nowhere dense graph classes