Research Project: A Graph Representation for Databases

Arne Meier

13.6.2022

But, first: Ласкаво просимо!

Regarding the Project

Abilities / Knowledge (you should have)

- Programming knowledge (Java, JDK $\geq$1.8)
- A bit of databases knowledge
- Propositional logic
- Graphs

Also: a laptop/computer is needed

Rough Outline of the Project

Write in the team a program that does the following:

1. Take a propositional logic formula φ

Rough Outline of the Project

Write in the team a program that does the following:

1. Take a propositional logic formula φ
2. Compute its incidence graph $\mathrm{IG}(\varphi)$

Rough Outline of the Project

Write in the team a program that does the following:

1. Take a propositional logic formula φ
2. Compute its incidence graph $\mathrm{IG}(\varphi)$
3. Determine the tree-width $k_\varphi \in \mathbb{N}$ of $\mathrm{IG}(\varphi)$

Rough Outline of the Project

Write in the team a program that does the following:

1. Take a propositional logic formula φ
2. Compute its incidence graph $\mathrm{IG}(\varphi)$
3. Determine the tree-width $k_\varphi \in \mathbb{N}$ of $\mathrm{IG}(\varphi)$
4. Translate φ into a database table D_φ

Rough Outline of the Project

Write in the team a program that does the following:

1. Take a propositional logic formula φ
2. Compute its incidence graph $\mathrm{IG}(\varphi)$
3. Determine the tree-width $k_\varphi \in \mathbb{N}$ of $\mathrm{IG}(\varphi)$
4. Translate φ into a database table D_φ
5. Compute its database graph representation $\mathrm{DBG}(D_\varphi)$

Rough Outline of the Project

Write in the team a program that does the following:

1. Take a propositional logic formula φ
2. Compute its incidence graph $\mathrm{IG}(\varphi)$
3. Determine the tree-width $k_\varphi \in \mathbb{N}$ of $\mathrm{IG}(\varphi)$
4. Translate φ into a database table D_φ
5. Compute its database graph representation $\mathrm{DBG}(D_\varphi)$
6. Determine the tree-width $\ell_\varphi \in \mathbb{N}$ of $\mathrm{DBG}(D_\varphi)$

Rough Outline of the Project

Write in the team a program that does the following:

1. Take a propositional logic formula φ
2. Compute its incidence graph $\mathrm{IG}(\varphi)$
3. Determine the tree-width $k_\varphi \in \mathbb{N}$ of $\mathrm{IG}(\varphi)$
4. Translate φ into a database table D_φ
5. Compute its database graph representation $\mathrm{DBG}(D_\varphi)$
6. Determine the tree-width $\ell_\varphi \in \mathbb{N}$ of $\mathrm{DBG}(D_\varphi)$

My motivating question:

How does ℓ_φ relate to k_φ in general?

Step 1: Which propositional formulas shall we use?

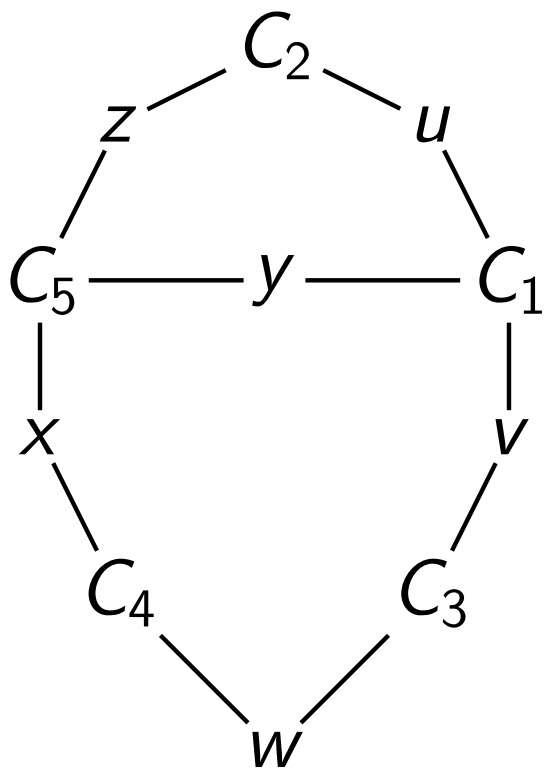
- max. 500 Variables (at first: start with fewer and try how it scales)
- get them from `https://www.cs.ubc.ca/~hoos/SATLIB/benchm.html`
- standard format: $(x_1 \vee x_3 \vee \neg x_4) \wedge (x_4) \wedge (x_2 \vee \neg x_3)$ goes to

```
c Example comment
p cnf 4 3
1 3 -4 0
4 0 2
-3
```

So... Start with writing a parser first.

Step 2: Incidence Graph of Formulas

$(u \vee \neg v \vee \neg y) \wedge (\neg u \vee z) \wedge (v \vee \neg w) \wedge (w \vee \neg x) \wedge (x \vee y \vee \neg z)$
becomes



- each variable becomes a vertex
- each clause number becomes a vertex
- draw edges between clauses and their vertices

Step 2.5: How to Represent Graphs?

short answer: `https://pacechallenge.org/2017/treewidth/`

```
c path with five vertices and four edges.
p tw 5 4
1 2
2 3
3 4
4 5
```

Step 3: Treewidth, Definition

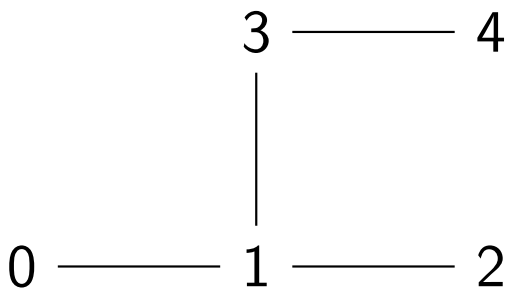
Definition

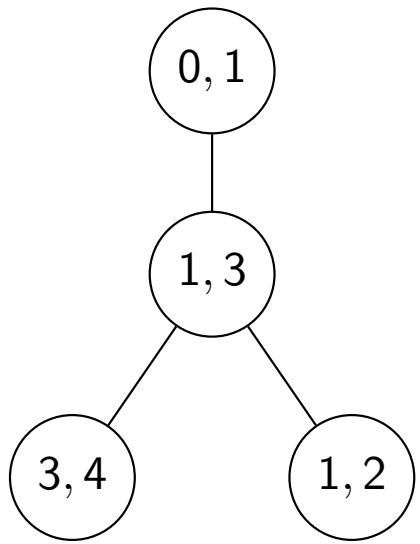
The tree-decomposition of a given graph $G = (V, E)$ is a tree $T = (B, E_T)$ such that:

- $\bigcup_{b \in B} = V$,
- for every $\{u, v\} \in E$ there is a bag $b \in B$ with $u, v \in b$ and
- for all $v \in V$ the restriction of T to v is connected.

Width of a given tree-decomposition $T = (B, E_T)$ is $\max_{b \in B} |b| - 1$.
The treewidth of a given graph G is the minimum over all widths of tree-decompositions of G.

Step 3: Treewidth, Example





Step 3: Treewidth, Bad News

Definition

Problem	TW (Treewidth Problem)
Instances	Graph $G = (V, E)$, natural number $k \in \mathbb{N}$
Question	Does G have a treewidth of k?

This problem is NP-complete.

Step 3: Treewidth, Silver Lining

- Use a blackbox for it
- `https://github.com/twalgor/tw`

Step 4: Translate a Formula into a Database

$(x_1 \vee x_3 \vee \neg x_4) \wedge (x_4) \wedge (x_2 \vee \neg x_3)$ becomes two tables

c	v	s
C_1	1	$+$
C_1	3	$+$
C_1	4	$-$
C_2	4	$+$
C_3	2	$+$
C_3	3	$-$

v	s
1	$+$
1	$-$
2	$+$
2	$-$
3	$+$
3	$-$
4	$+$
4	$-$

Step 5: Graph Representation DBG

Vertices

- one per row in the table

Edges

- between rows of same c-value
- between rows of same (v, s)-value

Step 6: Like Step 3

Last Step

- compare the two treewidth values
- how do they relate to each other?
- if they are different: how much do they differ?

Formalities

Where Do You Store the Code?

`https://github.com/ArneMeier/db-repr-students`

Which e-mail addresses shall I use for invitation?

The ones on the learning agreements?

How often do we meet?

Bi-weekly at 14:00–15:30 (27.6., 11.7., . . .) in this room 1611.

Stud.IP

Projekt: Graph Representations for Databases