University of Antwerp
Faculty of Science

Expressive power of graph neural networks

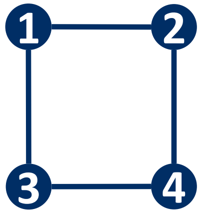
Floris Geerts

Relational databases & Structured Query Language (SQL)

S	T
v1	v2
v2	v3
v3	v4
v4	v1
v2	v1
v3	v2
v4	v3
v1	v4

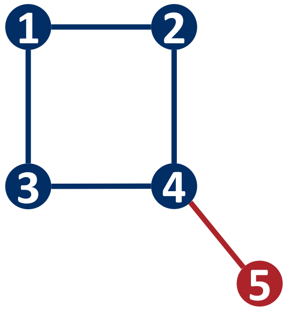
``Find all degree two vertices in a graph''

```
SELECT E1.S AS S
FROM E AS E1
WHERE  2= (SELECT COUNT(E2.T)
              FROM E AS E2 WHERE E2.S=E1.S);
```



S
v1
v2
v3
v4

S
v1
v2
v3



S	T
v1	v2
v2	v3
v3	v4
v4	v1
v2	v1
v3	v2
v4	v3
v1	v4
v4	v5
v5	v4

Course: introduction to databases

SQL & logic

- **SQL: standardized query language for relational databases**
- **First-order predicate logic with aggregation: formal mathematical abstraction of SQL**
- **In this lecture: First-order predicate logic with counting quantifiers**

```
SELECT E1.S AS S
FROM E AS E1
WHERE  2= (SELECT COUNT(E2.T)
                FROM E AS E2 WHERE E2.S=E1.S);
```

→

$$\varphi(x) = \exists^{=2} y\, E(x, y)$$

Course: logic

Logic example

$$\psi = \exists x(\exists^{\geq 3} y\ (E(x,y) \wedge \exists z(E(x,z) \wedge \exists^{=1} v\ E(z,v)))$$

$$\exists^{\geq 3} y\ (E(x,y) \wedge \exists z(E(x,z) \wedge \exists^{=1} v\ E(z,v)))$$

$G \vDash \psi$ **or G satisfies** ψ

Optimization: only two variables are needed

$$\exists x(\exists^{\geq 3} y(E(x,y)) \wedge \exists y(E(x,y) \wedge \exists^{=1} x\ (E(y,x))))$$

FOC_k = k-variable fragment of first order logic with counting quantifiers

Expressive power

- Interested in which properties can or cannot be expressed in SQL
- From SQL to Logic: Expressive power of logics

- Given two graphs, do they satisfy the same logic sentences?
- Given a graph property, e.g., is a graph connected, can this be expressed by a logic formula?

- Insights in these questions give insights in capabilities of practical query languages and drives innovation.

Expressive power of logics

- Many many many different logics around (not only in query languages)
- Determining whether two objects are *equivalent* for a logic, i.e., whether one cannot detect a difference between the two objects using formulas in the logic, is one of the basic problems.

- Computationally: complexity of deciding equivalence
- Conceptually: characterization of equivalence
- Mathematically: tools (games, finite model theory, …) to analyze logics

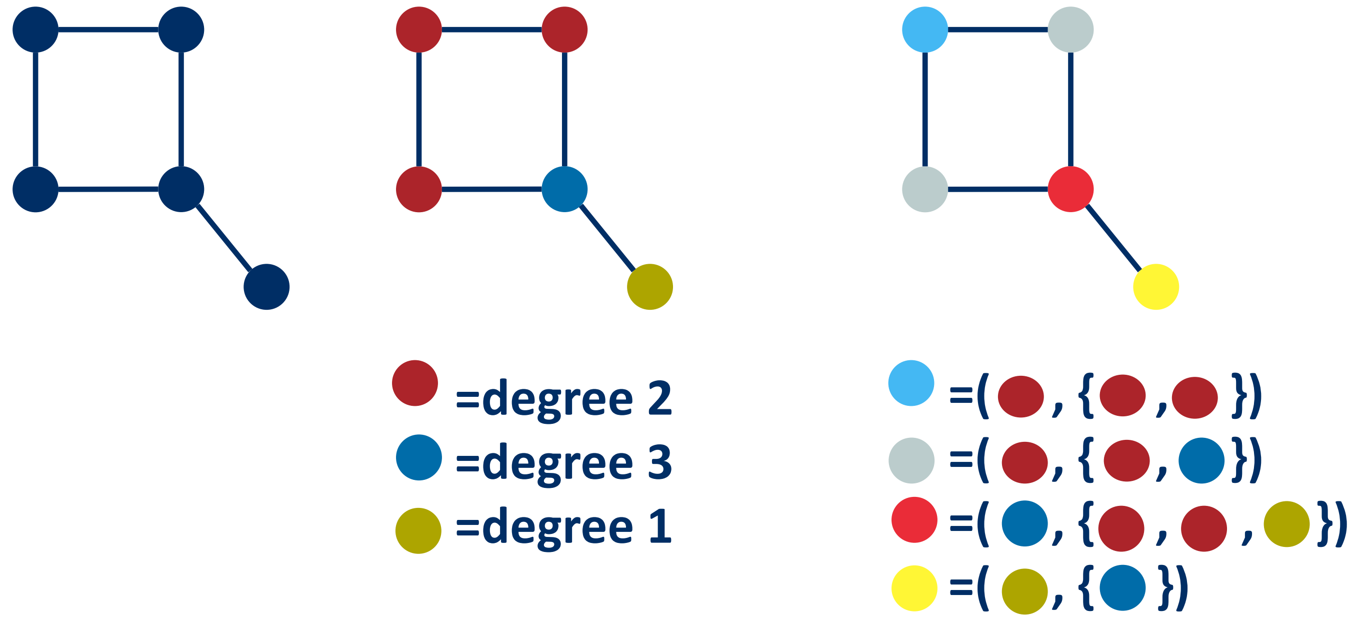
Let us focus on FOC_2

FOC$_2$ & color refinement

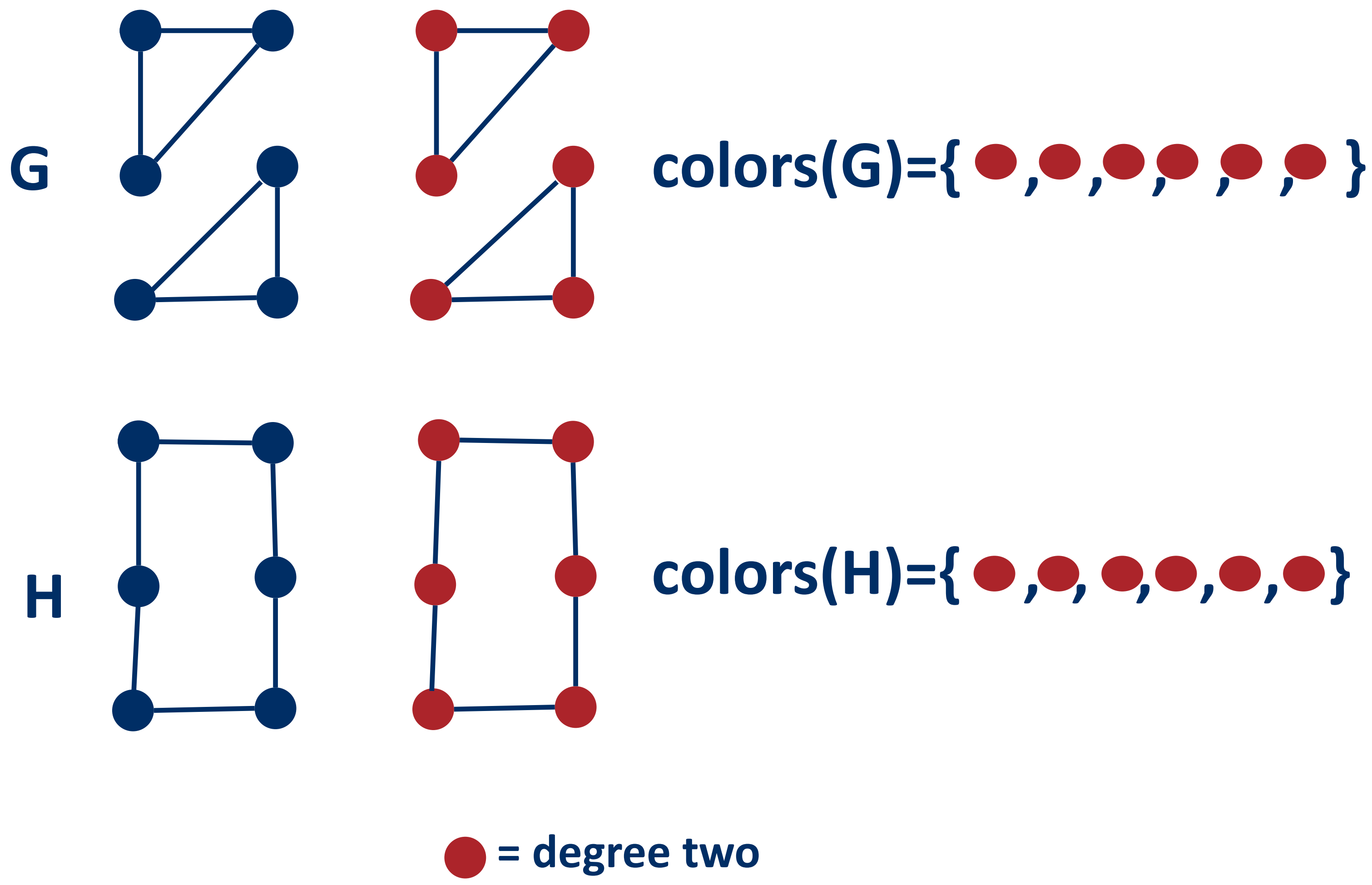
Theorem

*Two graphs cannot be told apart using sentences **in FOC$_2$** if and only if they are equivalent with regards to **color refinement** (if and only if Duplicator has a winning strategy in **the bijective two-pebble game**).*

Color refinement



Color refinement example



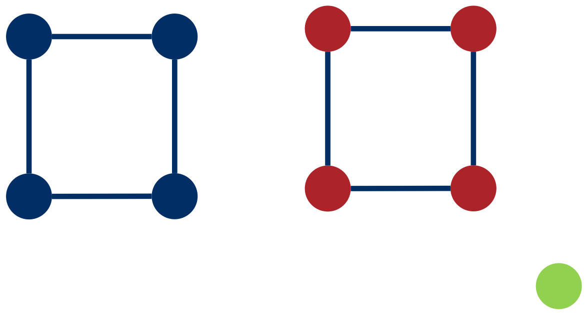
G and H are equivalent for color refinement

G and H are equivalent for FOC_2

FOC_2 not powerful enough to express connectivity or checking for cycles…

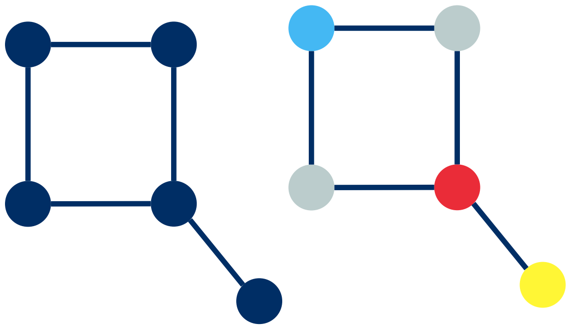
Color refinement example

G



colors(G)={}

H



colors(H)={}

G and H are not equivalent for color refinement

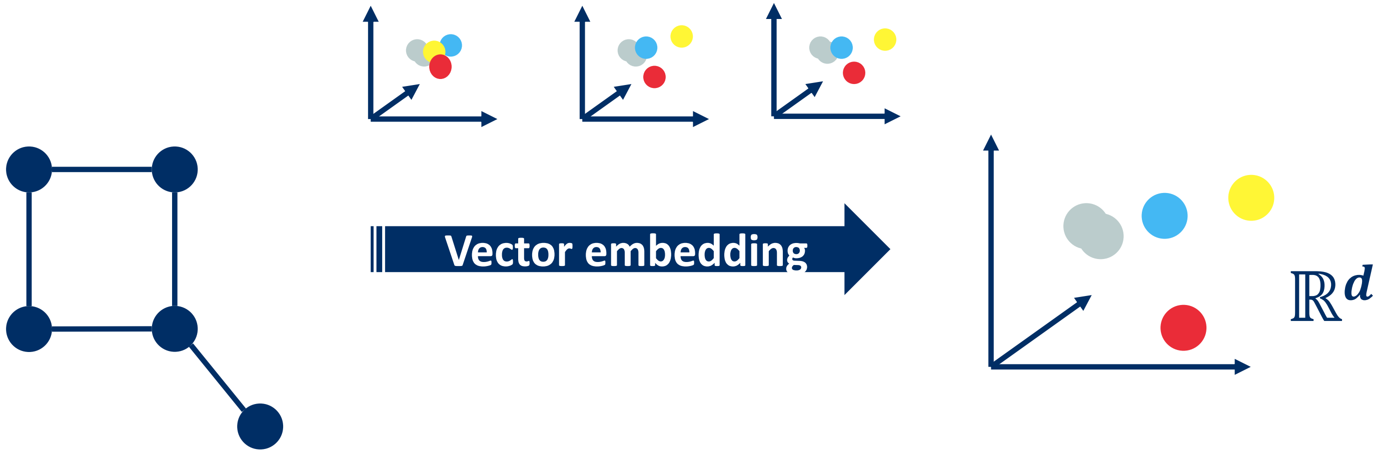
G and H are not equivalent for FOC_2

E.g., $\exists^{=1} x \, \neg \exists^{>0} y \, E(x,y)$

Color refinement

- Provides a complete (and easy to check) characterization of FOC_2 equivalence
- Similar characterizations are in place for FOC_k in terms of k-dimensional Weisfeiler-Leman algorithm
- Color refinement (and k-WL) play a crucial role in graph isomorphism testing

Graph neural networks (GNN)



- **Iterative, layer based embedding computation**

$$emb^{t+1}(x) = UPD_{\Theta}(emb^{t}(x), AGG_{\Gamma}(emb^{t}(y) | E(x,y) = 1))$$

- **UPD is learnable update function**
- **AGG is learnable aggregation function**

Expressive power of GNNs

- **When can two vertices be embedded differently by a GNN?**

GNN

$emb^{t+1}(x)$

a in $\mathbb{R}$

→

Infinitary FOC_2

$$\varphi^a(x) = \bigvee_{\{i \in I\}} \bigwedge_{\{j \in J\}} \varphi^t_{ij}(x)$$

$$emb^{t+1}(G, v) = a$$

→

$$(G, v) \vDash \varphi^a$$

FOC_2 & color refinement

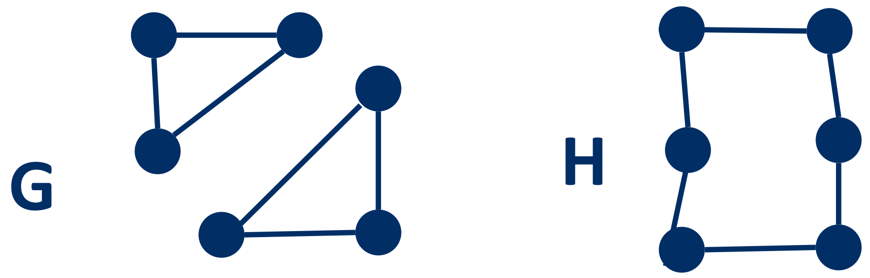
Theorem

$(G, v) \vDash \varphi^a$ *and* $(G, w) \vDash \varphi^a \Rightarrow emb(G, v) = emb(G, w)$

Corollary

If color refinement cannot distinguish G from H then neither can any GNN!

Suppose emb(G,v) not equal to emb(G,w) for some GNN, then (G,v) satisfies some φ^a in infinitary FOC_2 while (G,w) does not. This implies that (G,v) satisfies some FOC_2 ψ^a while (G,w) does not.
But if color refinement does not distinguish G from H then G and H must satisfy the same formulae, including ψ^a.



Conclusion

- Expressive power of logics not only of interest for query languages also for GNNs
- Expressive power of k-GNNs related to that of FOC_k for k>1
- FOC_k is less expressive than FOC_{k+1}
- Connection with logic has resulted in many new GNNs based with higher expressive power