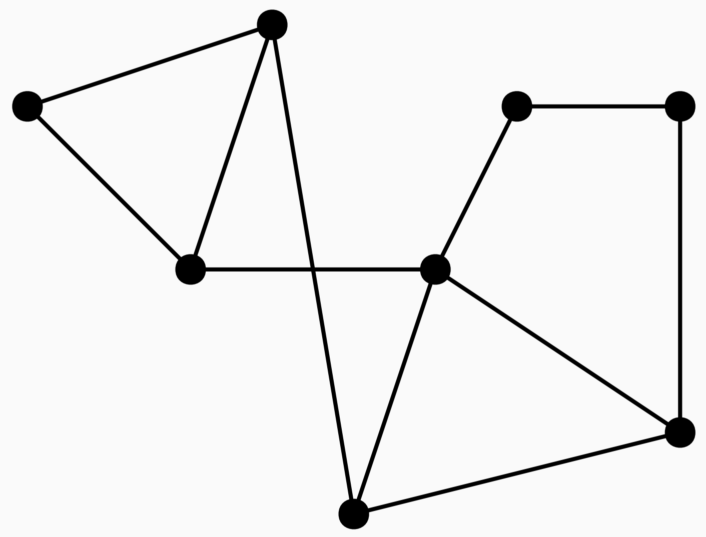
Graph Motif Parameters for Complexity, Expressivity and Representation

Matthias Lanzinger

TU WIEN Informatics

Graph Parameters

A ***graph parameter*** maps graphs to numbers.

 $\to \mathbb{R}$

Graph Parameters – Examples

- Number of Answers to a Graph Query
- Chromatic Number
- Independence Number
- Number of triangles in G
- …

Graph Parameters

Theorem

For every graph parameter Γ there exists a set $\mathcal{F}$ of graphs such that:

$$\Gamma(G) = \sum_{F \in \mathcal{F}} \alpha_F \cdot \text{homs}(F, G)$$

Coefficient $\in \mathbb{Q}$

Number of homomorphisms
From F to G

Refresher: Homomorphisms

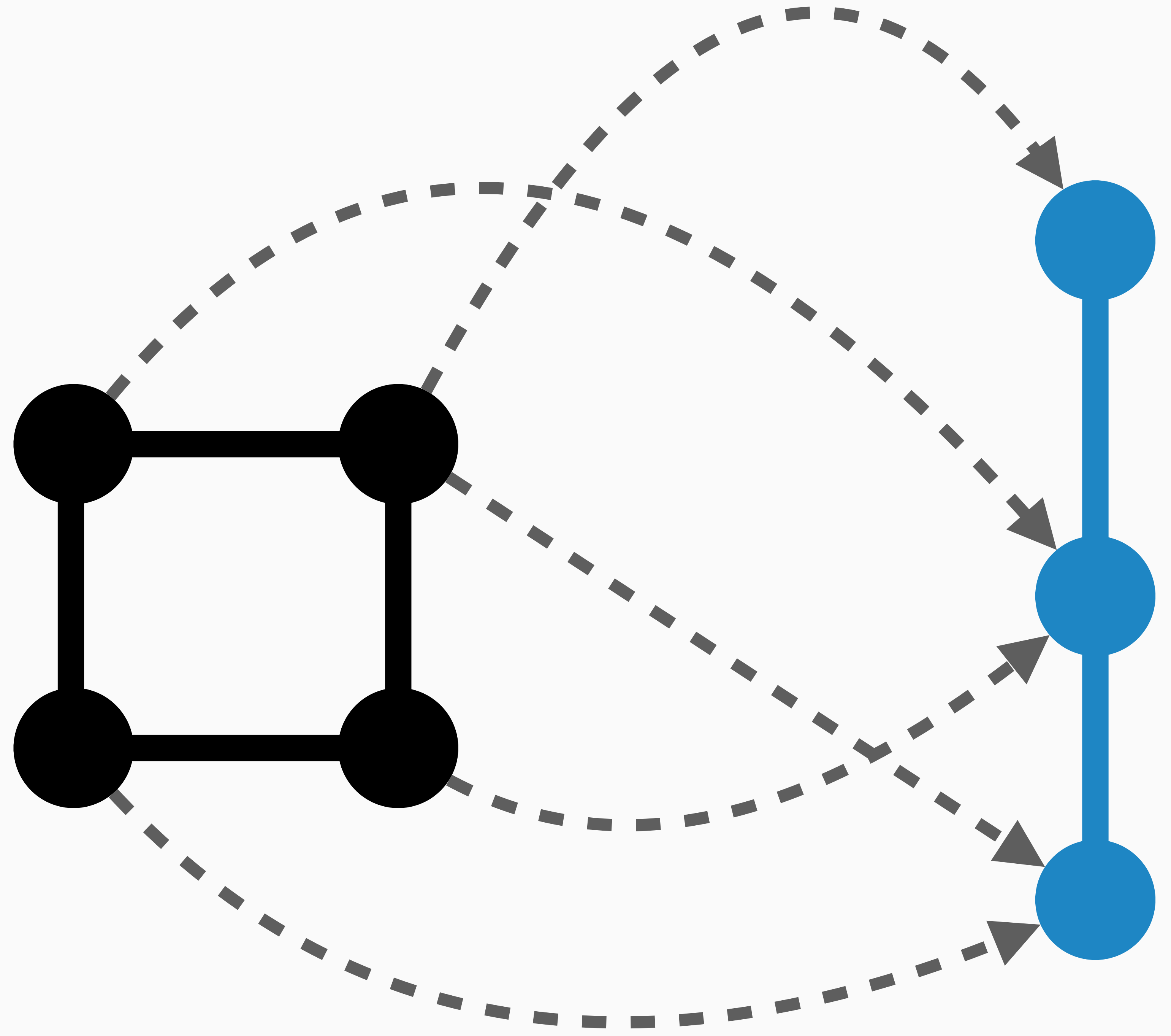
A homomorphism is a mapping

$h : V(G) \to V(H)$ s.t.:

If $v\,u \in E(G)$,

then

$h(v)\,h(u) \in E(H)$



Graph Motif Parameters

Graph parameter Γ is a ***graph motif parameter*** if there is a **finite** set $\mathscr{F}$ of graphs such that:

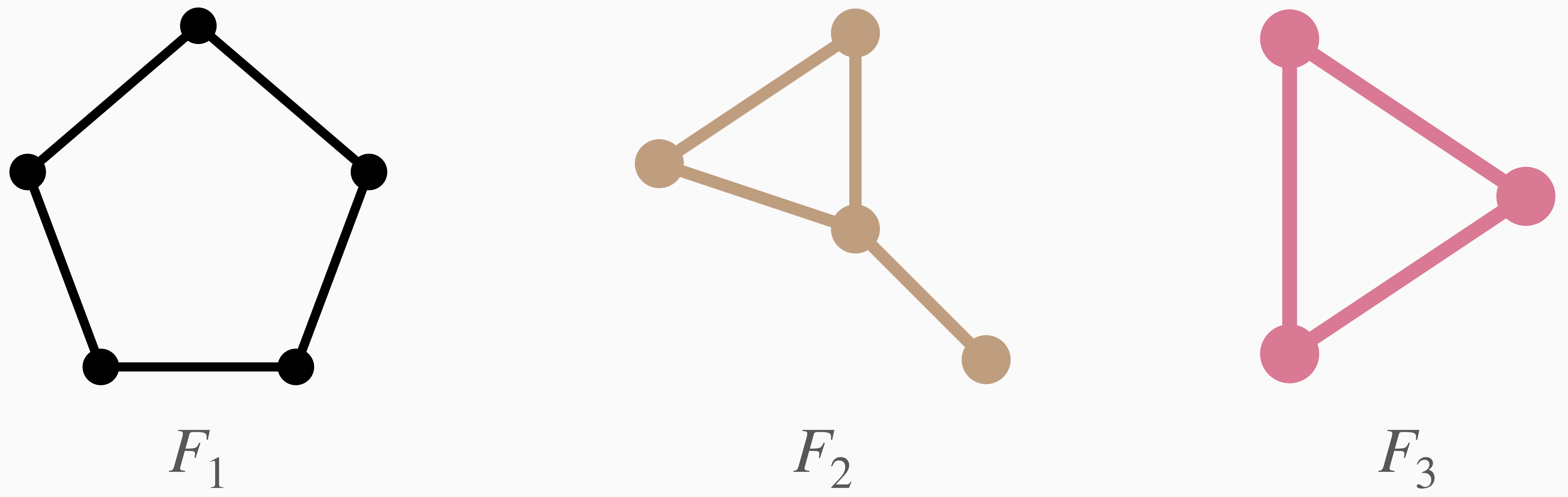
$$\Gamma(G) = \sum_{F \in \mathscr{F}} \alpha_F \cdot \mathsf{homs}(F, G)$$

Always unique!

Example

The number of 5-cycles in G

$$= \frac{1}{10}\text{homs}(F_1, G) - \frac{1}{2}\text{homs}(F_2, G) + \frac{1}{2}\text{homs}(F_3, G)$$



Why?

Canonical representation of many functions in terms of a single (kind of) function: *homs*

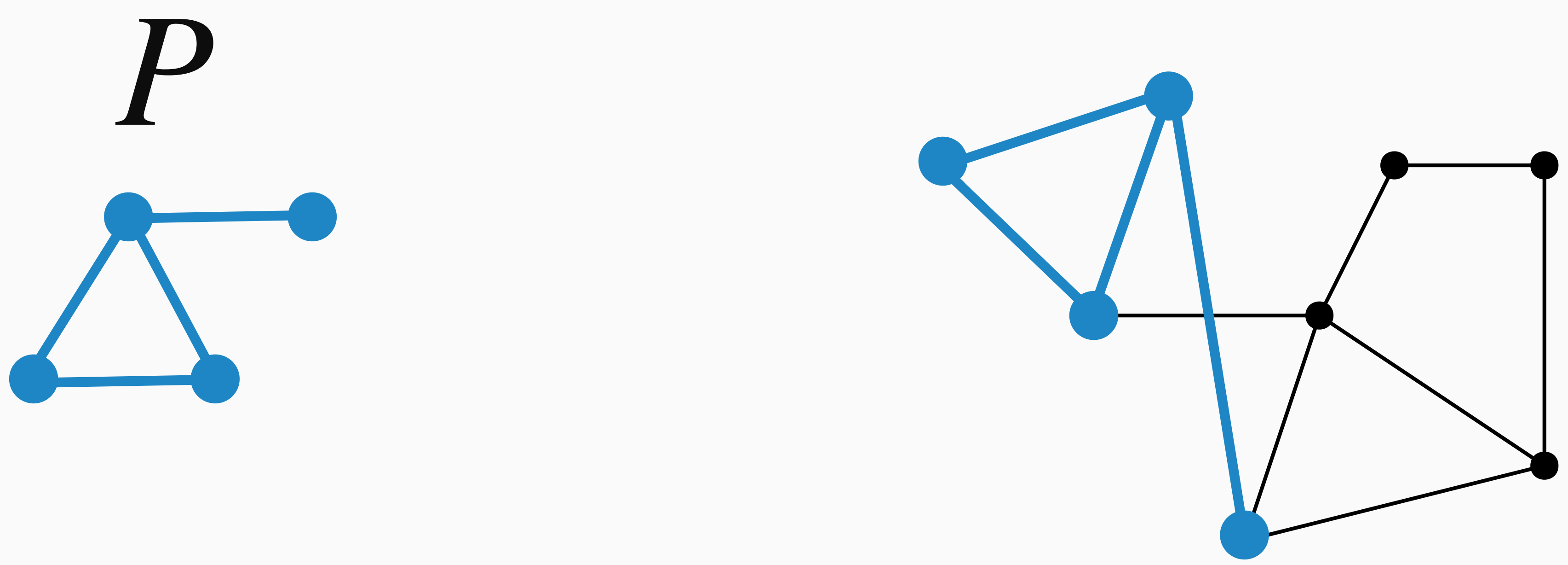
Historically homomorphisms have been easier to understand.

We will be able to transfer our understanding of homomorphisms to graph motif parameters!

Important
Graph Motif Parameters

Counting Subgraphs

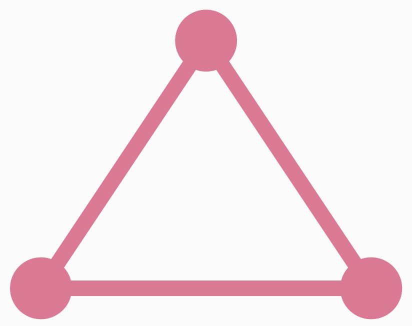
$Sub(P, G)$:= Number of subgraphs of G isomorphic to P.



$Sub(P, \cdot\,)$ is a graph motif parameter (for every P).

Important Problems

Counting the number of triangles in graph G.

$Sub($  $, G)$

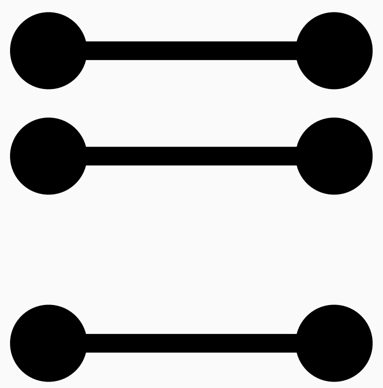
Important Problems

Counting the number of
n-vertex cliques in graph G.

$$Sub(K_n, G)$$

Important Problems

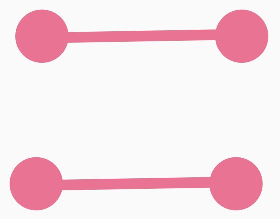
Counting the number of
size n matchings in graph G.

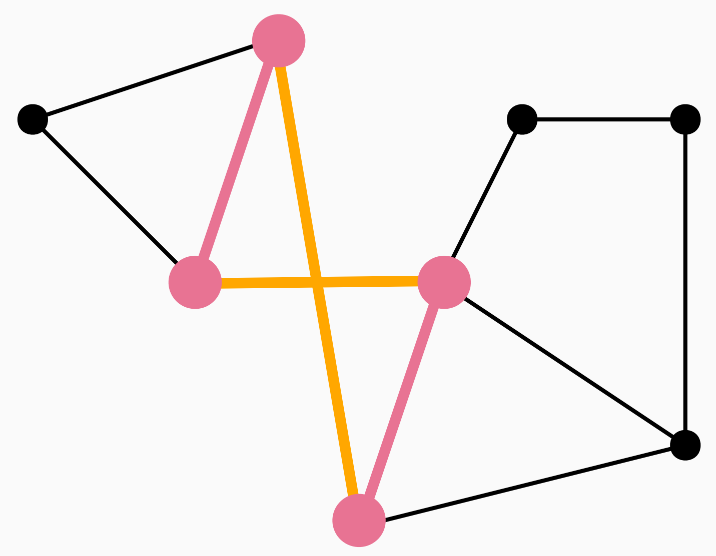
$Sub($  $, G)$

Counting Induced Subgraphs

$IndSub(P, G)$:= Number of induced subgraphs of G isomorphic to P.

P

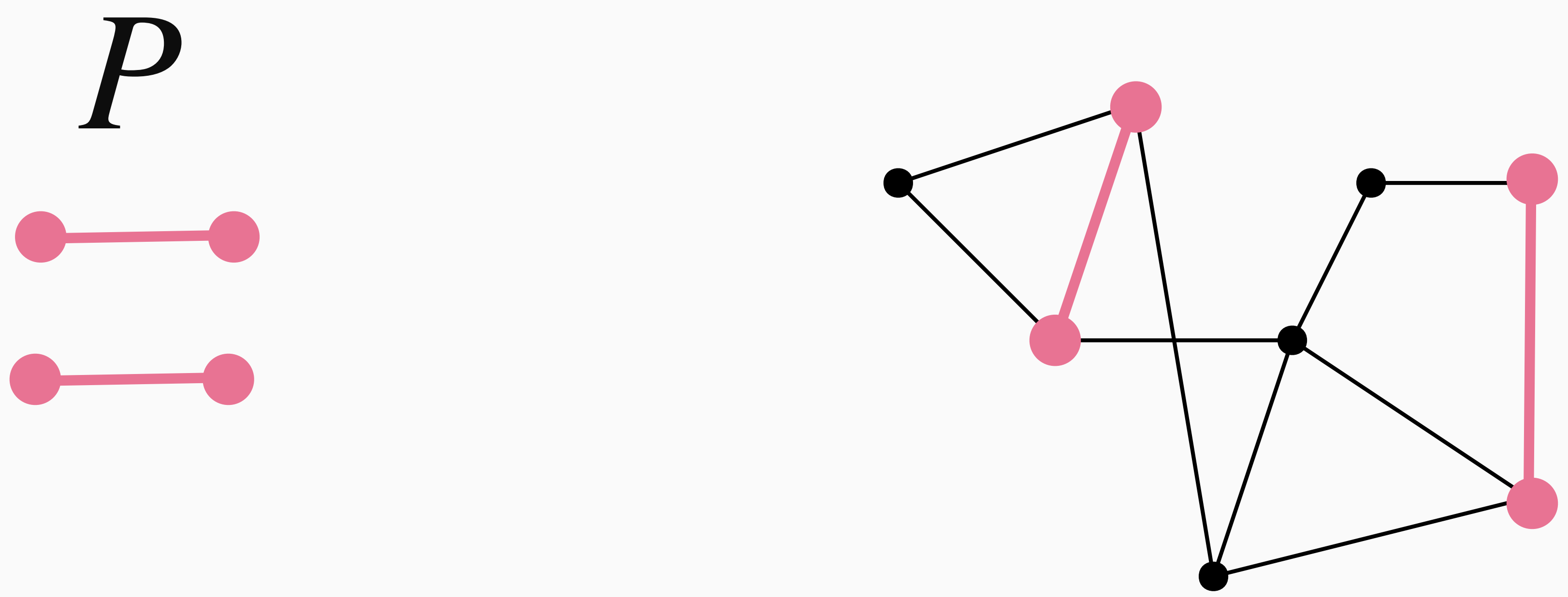




$IndSub(P, \cdot\ )$ is a graph motif parameter (for every P).

Counting Induced Subgraphs

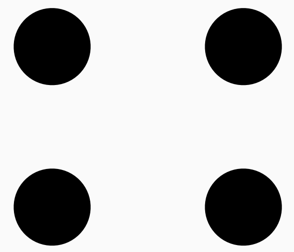
$IndSub(P, G)$:= Number of induced subgraphs of G isomorphic to P.



$IndSub(P, \cdot\ )$ is a graph motif parameter (for every P).

Important Problems

Counting the number of
n vertex independent sets in graph G.

$IndSub($  $\cdots, G)$

Property Counting

$Ind_{\phi}^{k}(G) :=$ number of induced subgraphs on k vertices that have property ϕ

<u>Example:</u>

Number of k-graphlets is the case where ϕ is connectedness.

Ind_{ϕ}^{k} is a graph motif parameter for every computable property ϕ.

Very Robust

- Vertex/Edge-labels
- Partial injectivity
- Weighted counting
- Directed Graphs
- Projection (**#UCQs on graphs**)
- …

Complexity

An Upper Bound

$$\Gamma(G) = \sum_{F \in \mathscr{F}} \alpha_F \cdot homs(F, G)$$

1) Compute $\mathscr{F}$ and all coefficients α

2) Compute $homs(F, G)$ for all $F \in \mathscr{F}$

3) Arithmetic

Computing Γ is no harder than computing the hardest $homs$ term in the sum.

Parameterised Counting

Problem with input I and parameter k.

FPT $= f(k) \cdot poly(|I|)$

#W[1] = The complexity of counting k-cliques in graph I

Standard assumption **FPT** $\neq$ **#W[1]**

Parameterised Counting of *homs*

Theorem

It is possible to compute $homs(F, G)$ in time

$$f(F) \cdot |G|^{tw(F)}$$

Theorem *(Marx, 2010)*

It is not possible to compute $homs(F, G)$ in time

$$f(F) \cdot |G|^{o(k/\log k)}$$

where $k = tw(F)$. Assuming ***ETH***.

You cannot do (much) better than treewidth!

Some History

2002 — **#W[1]**-hardness of counting paths/cycles. (Flum & Grohe)

count P_k or C_k in graph I

.

.

.

2013 — **#W[1]**-hardness of counting matchings. (Curticapean)

count k-matchings in graph I

Complexity Monotonicity

$$\Gamma(G) = \sum_{F \in \mathcal{F}} \alpha_F \cdot homs(F, G)$$

Theorem *(Curticapean, Dell, Marx, 2017)*

Computing Γ is exactly as hard as computing the hardest term $homs(F, G)$ for $F \in \mathcal{F}$.

Exactly as Hard?

$$\Gamma(G) = \sum_{F \in \mathcal{F}} \alpha_F \cdot homs(F, G)$$

There is an efficient (fpt) Turing reduction from computing Γ, to computing any function $homs(F, \cdot\,)$ for $F \in \mathcal{F}$

For a very general argument of why this is the case see Bressan, L., Roth, "The complexity of pattern counting in directed graphs, parameterised by the outdegree", STOC 2023

A Recipe for Dichotomies

1. Figure out the structure of the basis $\mathscr{F}$

2. Understand the complexity of homomorphism counting for the graphs in the bases.
   (often already done 👍)

3. Typically directly gives you a FPT vs. non-FPT dichotomy.
   (And probably fine-grained lower bounds.)

When is Pattern Counting Hard?

Theorem *(Curticapean, Dell, Marx, 2017)*

Computing $Sub(P, \cdot\,)$ for $P \in \mathscr{P}$ is **FPT**

if and only if

All graphs in $\mathscr{P}$ have **vertex cover number** $\leq c$, for some constant c.

If no such c exists the problem is **#W[1]**-hard

Example: Paths

n-hop path

Example: Paths

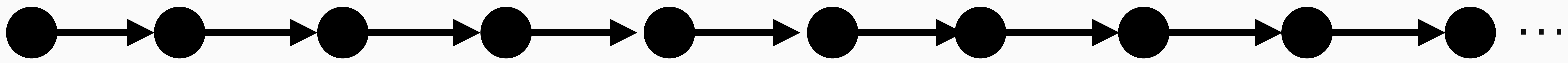
A minimal vertex cover

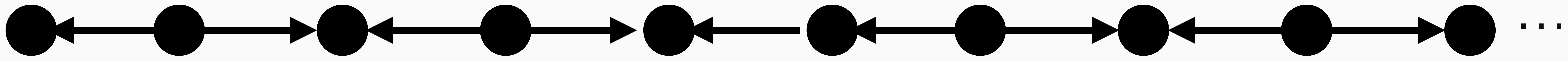
n-hop path

Vertex cover number $\approx \frac{n}{2}$

$\Rightarrow$ **#W[1]**-hardness of counting paths is immediate.

Directed Paths?

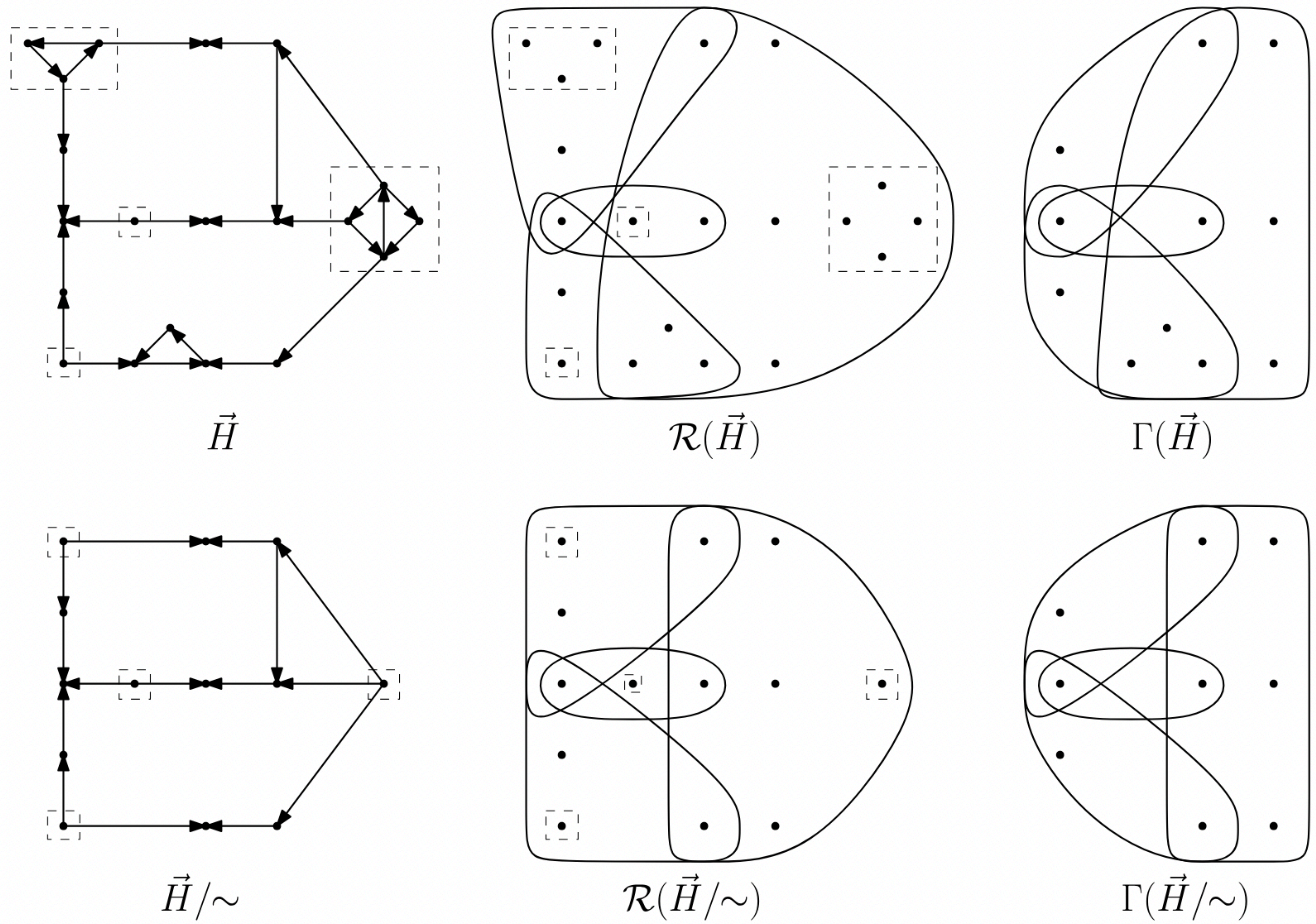




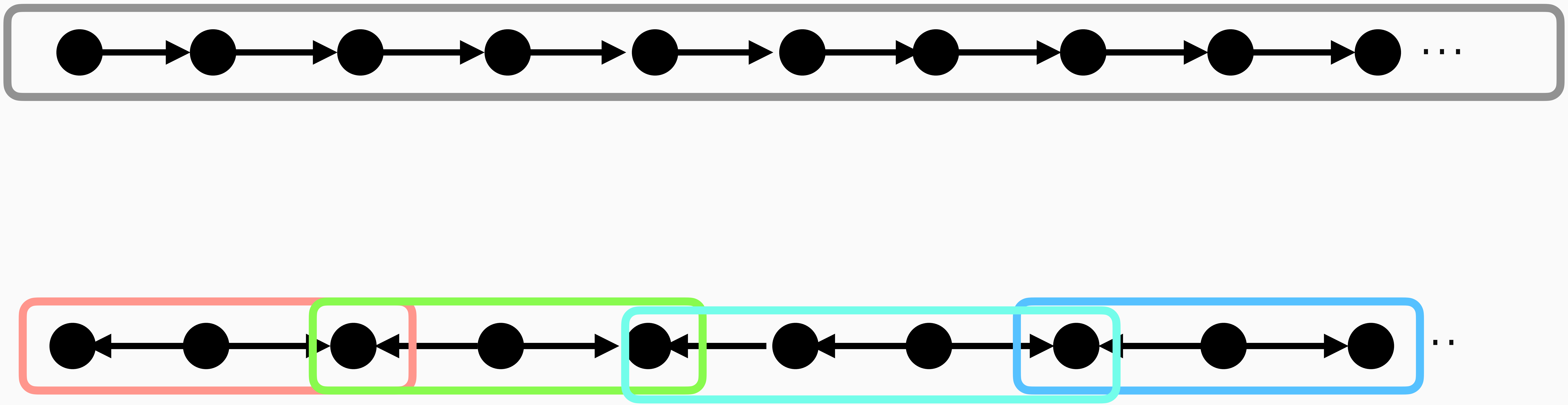
In standard parameterisation by just the pattern:
Exact same boundary as in undirected case.

Folklore / Appendix B in Bressan, L., Roth, “The complexity of pattern counting in directed graphs, parameterised by the outdegree", STOC 2023

We are interested in the case where in/out edges are unbalanced (applies to e.g. low degeneracy). We study this by adding the max *out-degree* to the parameter.



The contour of $\overrightarrow{H}$



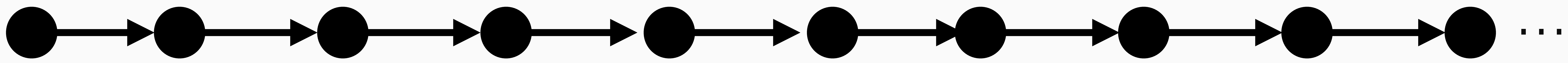
Parameterised by pattern and outdegree

Theorem 4. *If the Exponential Time Hypothesis holds, then:*

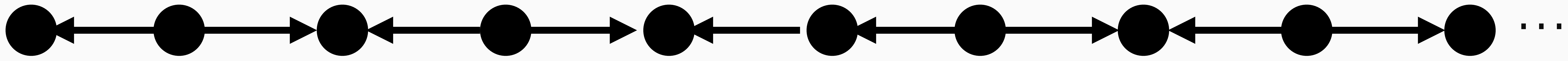
1. $\#\textsc{DirSub}_{\mathrm{d}}(\vec{C}) \in \mathsf{FPT}$ *if and only if* $\rho^*(\vec{C}) < \infty$
2. $\#\textsc{DirIndSub}_{\mathrm{d}}(\vec{C}) \in \mathsf{FPT}$ *if and only if* $\alpha_s(\vec{C}) < \infty$

Bressan, L., Roth, "The complexity of pattern counting in directed graphs, parameterised by the outdegree", STOC 2023

Directed Paths?



FPT

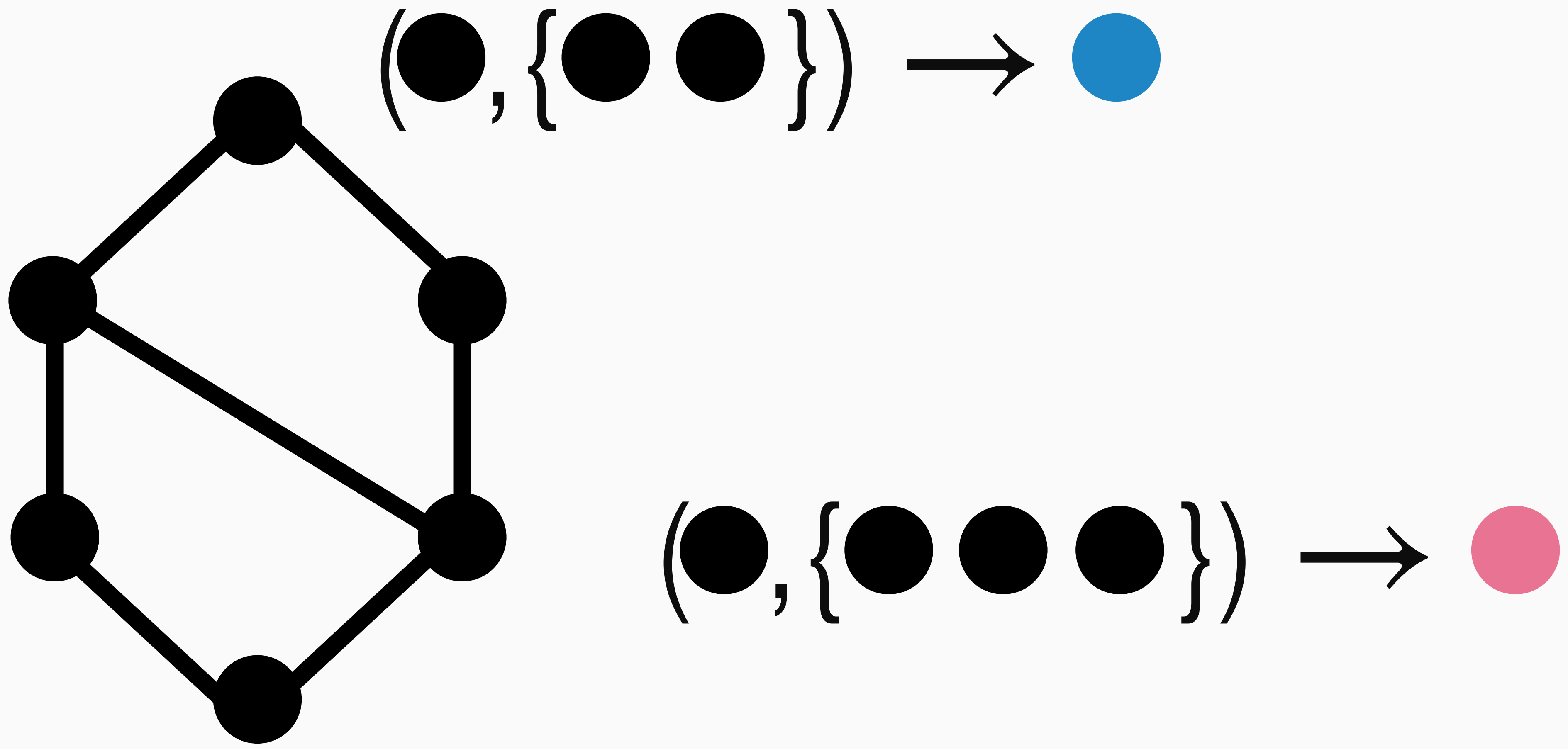


Not FPT

Expressivity

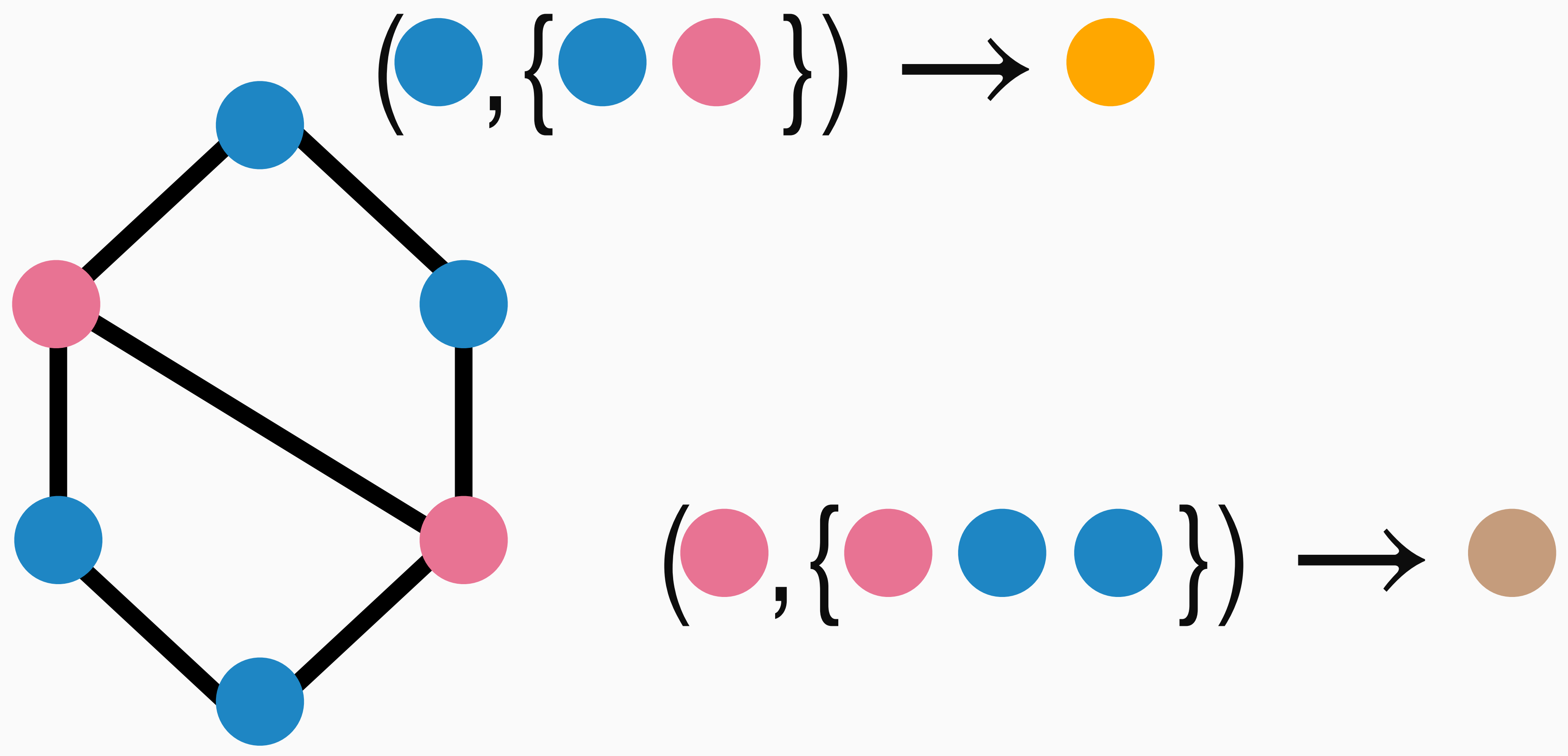
L., Barceló, “On the Power of the Weisfeiler-Leman Test for Graph Motif Parameters.", ICLR 2024

Color Refinement

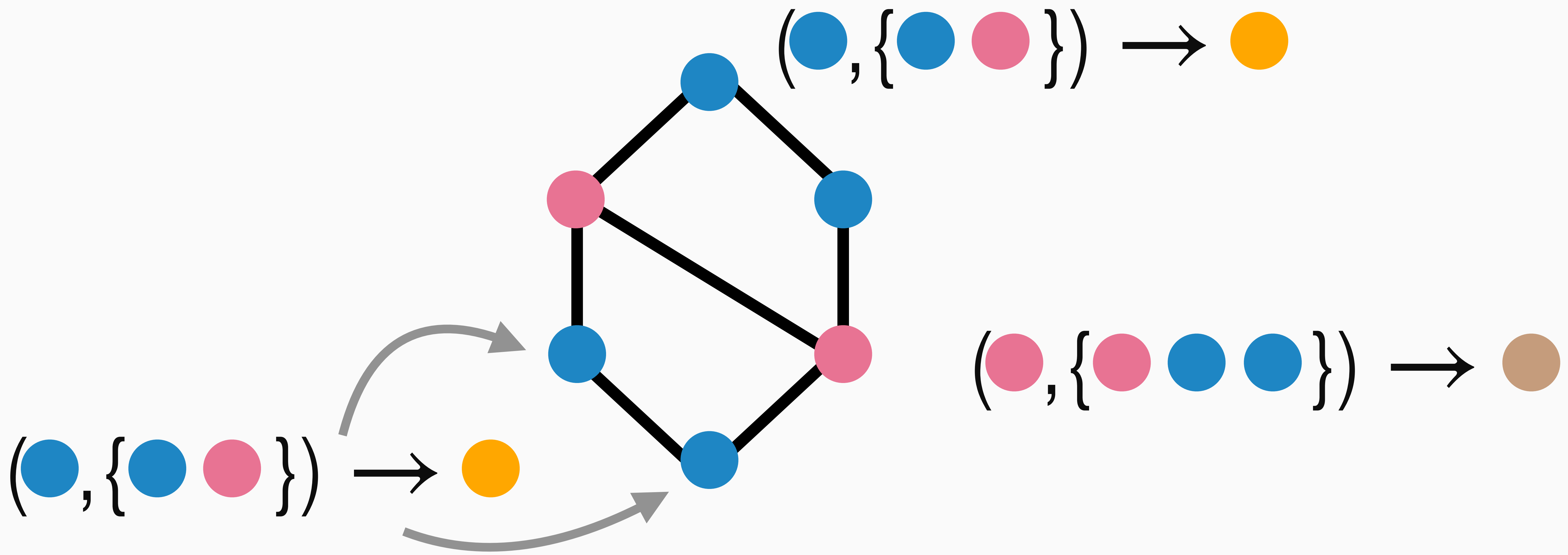


New node color = old color + multi-set of neighbours

Color Refinement

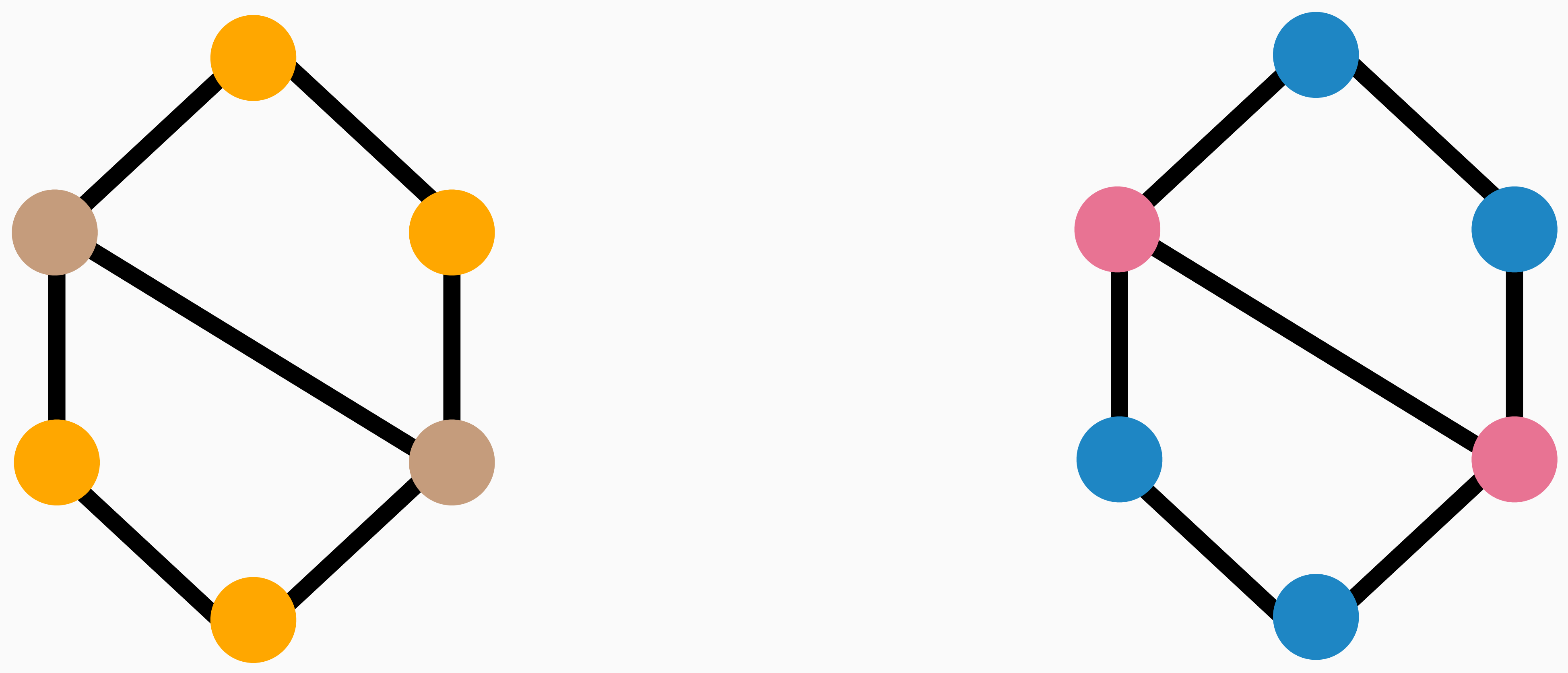


Color Refinement



Color Refinement

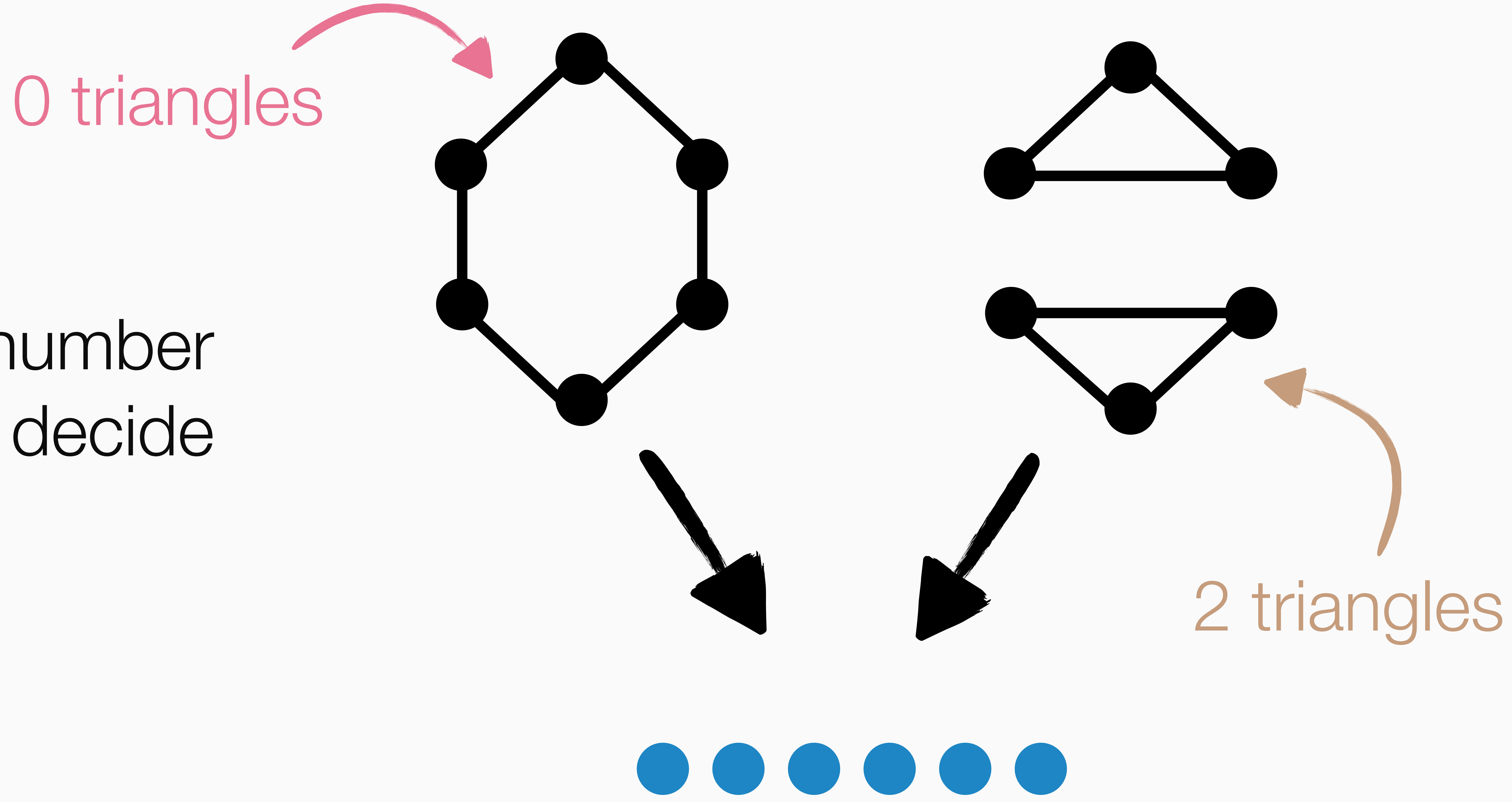
Historical use: different stable colouring $\Rightarrow$ not isomorphic



Pattern will repeat: **stable** colouring!

Expressivity of 1-WL

GNNs cannot count the number of triangles in a graph (or decide connectedness)!



We enjoy hierarchies

1-**WL**: Standard Color Refinement

2-**WL**: “Color refinement on all pairs of vertices”

⋮

k-**WL**: “Color refinement on all k-tuples of vertices”

What can k-WL (not) express?

Expressing a Function — Formally

k-WL can ***express*** function Γ if

$$G \equiv_{kWL} H \Rightarrow \Gamma(G) = \Gamma(H)$$

The ***WL-dimension*** of Γ is the least k such that
k-WL can express Γ

Intuitively, the minimal level of GNN that we need to reason about Γ

Early Steps

Fürer (2017):

- WL-dimension of counting k-Cycles for $3 \leq k \leq 6$ is 2.
- WL-dimension of counting k-Cycles for $8 \leq k \leq 16$ is *at least* 2.

Arvind et al. (2020):

- WL-dimension of counting 7-Cycles is 2
- Only subgraph counting problems with WL-dimension 1 are stars and the 2-matching graph.

Answers!

Theorem (Neuen 2024, L. & Barceló 2024):

For graphs with vertex and edge labels,
k-WL can distinguish graph motif parameter Γ

$$\iff$$

maximum treewidth in basis of $\Gamma \leq k$

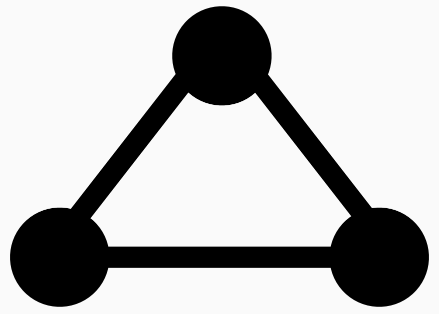
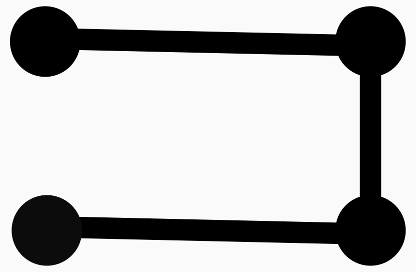
Reuse Complexity Results

The complexity of computing graph motif parameters is inherently about understanding the treewidth in the basis.

The same analysis from complexity theory can be reused for GNN expressivity!

Examples

Counting patterns:

			P_{12}	C_8
WL-Dimension	2	2	4	3

Subsumes results of Fürer and Arvind et al.
and provides a reason!

Fancy Example

k-graphlets & k-vertex disconnected graphs:

WL-Dimension $(k-1)$

(needs some basic algebraic topology)

Special cases of counting Ind_{Ψ}^{k} where Ψ is disconnectedness.

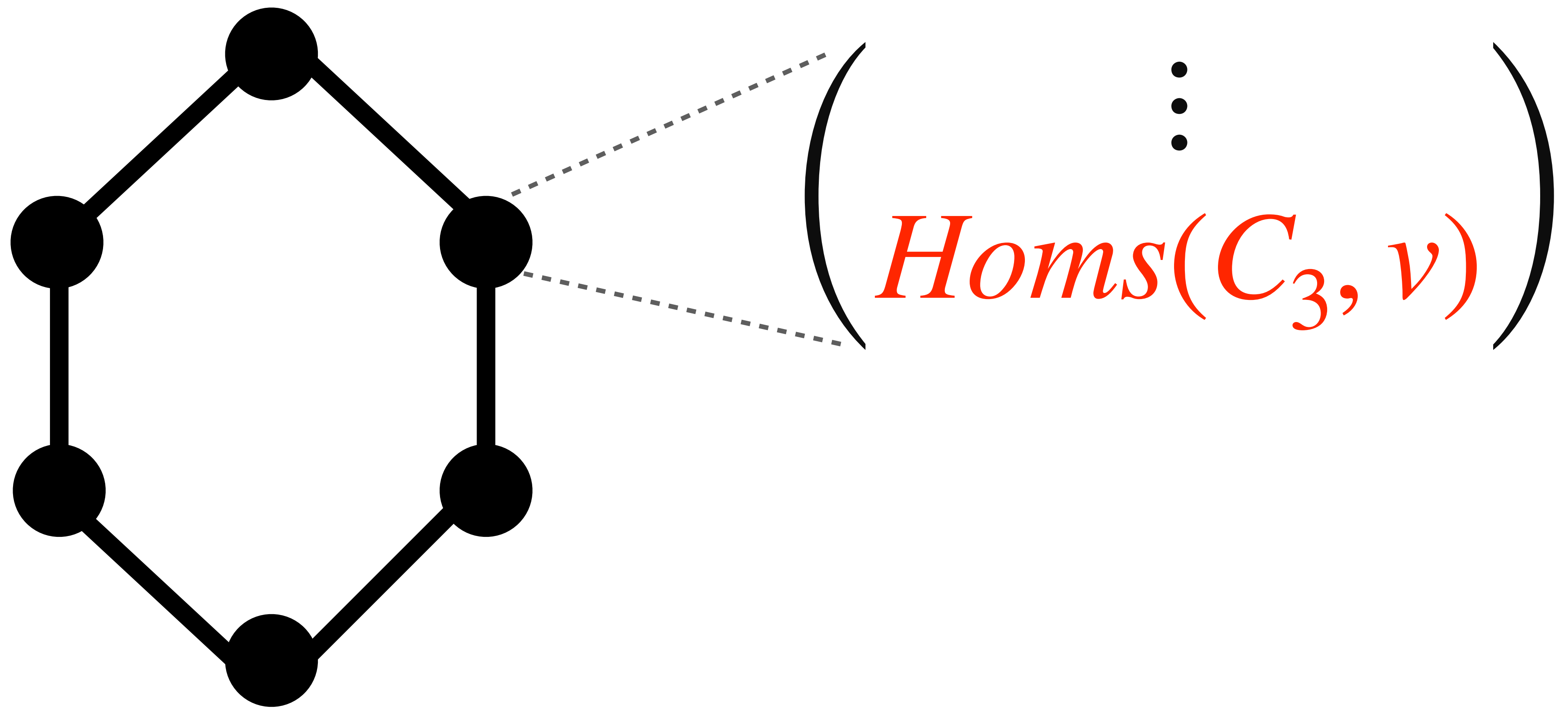
Roth & Schmitt: k-Clique is in the basis if the Euler characteristic for the simplical complex of hereditary ψ is non-zero!

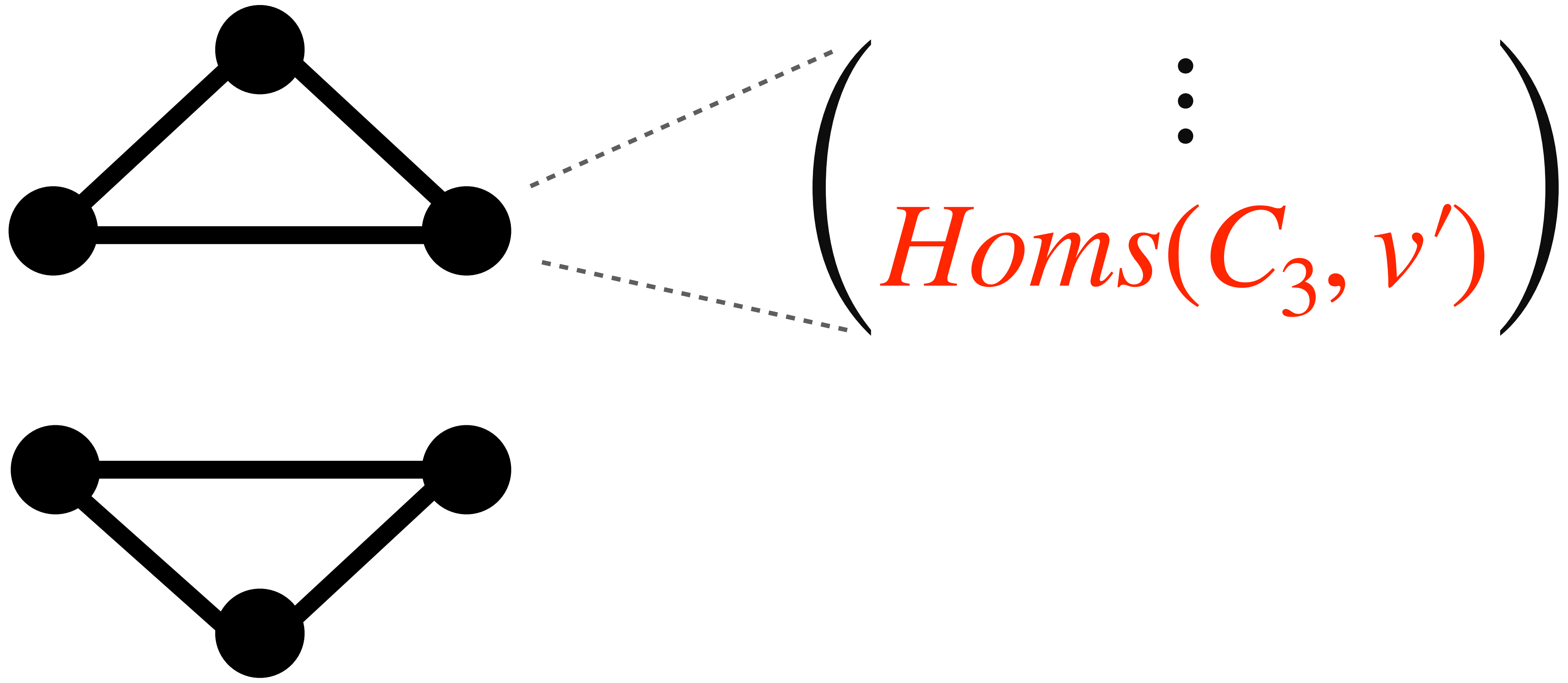
For this Ψ, the implical complex is the $k-3$ dimension wedge sum of spheres. Well-studied — has Euler characteristic $\pm(k-1)!$

Impact in Practice

Idea: we can make GNNs more powerful by adding local information at each node.

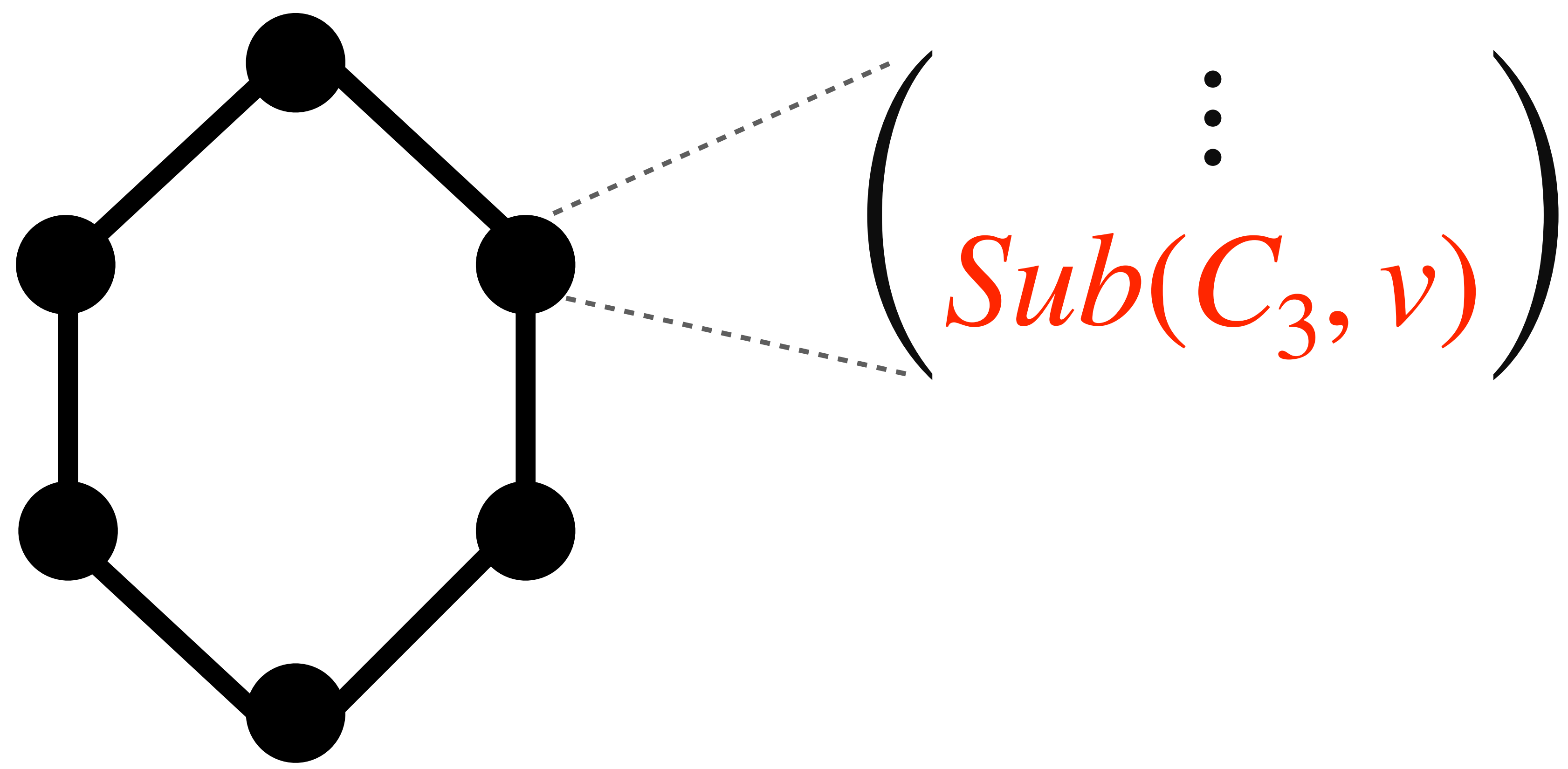
Barceló, P., Geerts, F., Reutter, J., & Ryschkov, M. “Graph neural networks with local graph parameters.” NeurIPS 2021





Impact in Practice

Idea: we can make GNNs more powerful by adding local information at each node.



Bouritsas, Giorgos, Fabrizio Frasca, Stefanos Zafeiriou, and Michael M. Bronstein. "Improving graph neural network expressivity via subgraph isomorphism counting." IEEE Transactions on Pattern Analysis and Machine Intelligence 45,.

We show that adding the basis of Γ gives more expressiveness, <u>for free</u>!

ingredients for cake

$>$

“more expressive than”

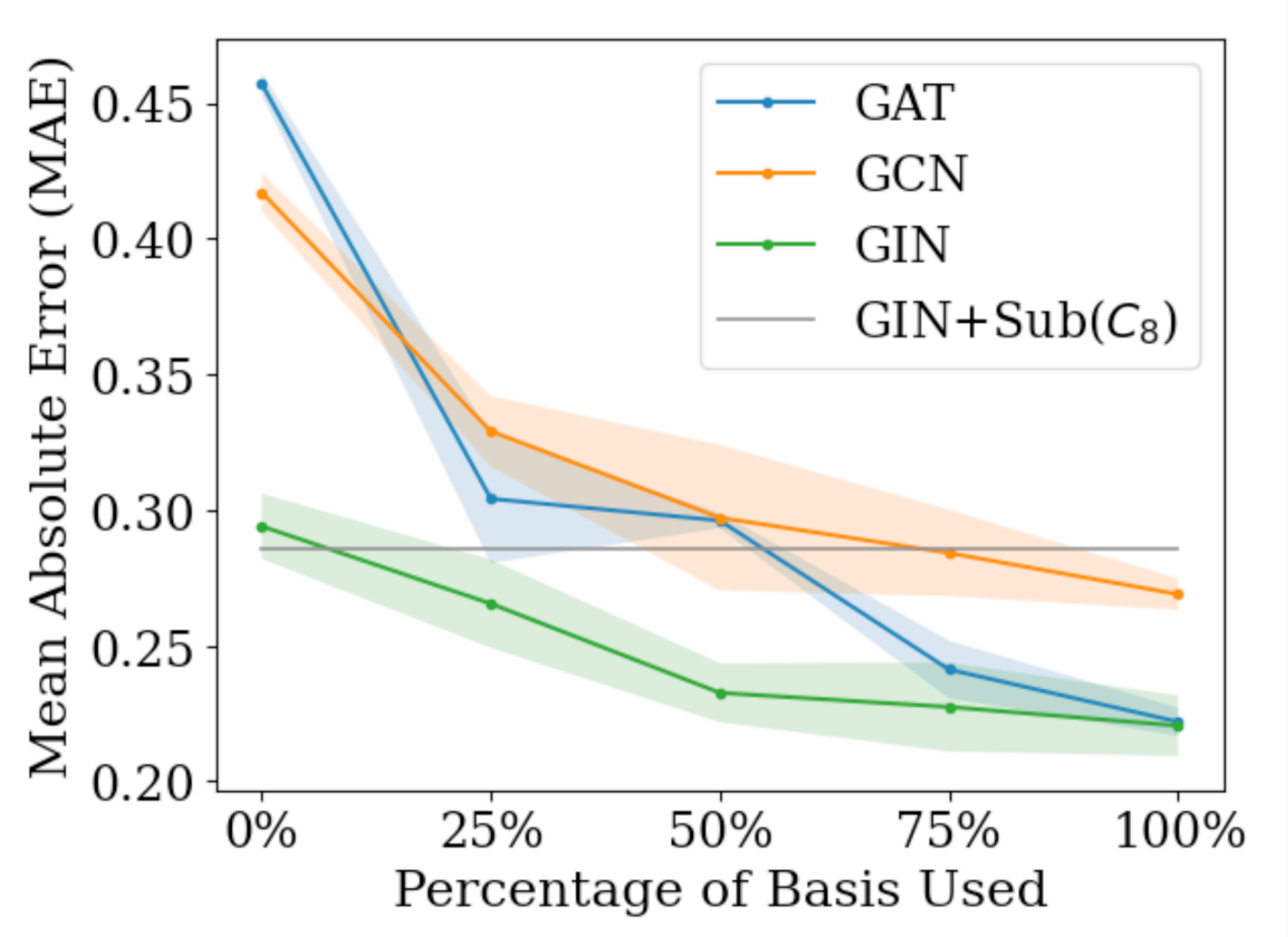
cake

Jin, Bronstein, Ceylan, L., “Homomorphism Counts for Graph Neural Networks: All About That Basis”, ICML 2024

ZINC

We know that cycles are important in molecules.

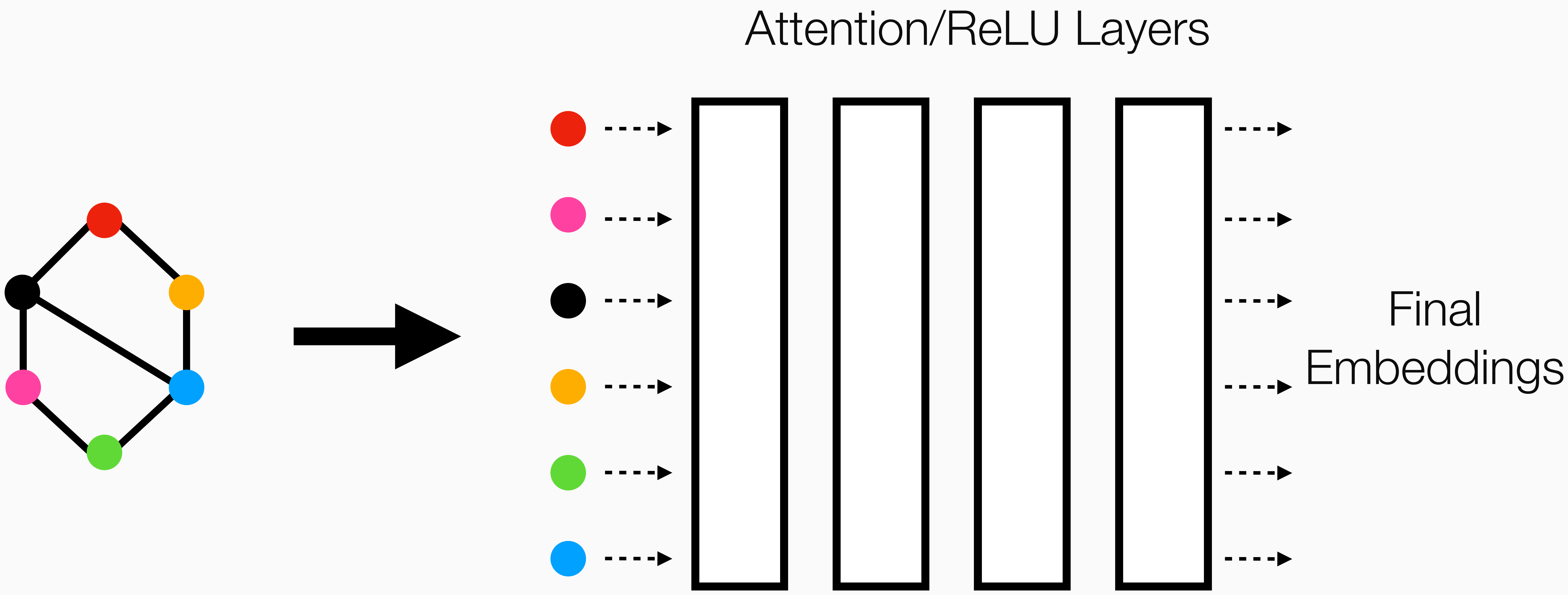
We compare adding $Subs(C_8, \cdot\,)$ to adding the $homs$ values for the basis of the function



Representation

Bao, Jin, Bronstein, Ceylan, L.. "Homomorphism Counts as Structural Encodings for Graph Learning.” ICLR 2025

Graph Transformers



The Problem

We often hear attention described as a "sequence to sequence" operation.

But, on a mechanical level the inputs are not ordered!

Sequence structure is obtained by adding a positional encoding, representing the position in the sequence, to inputs.

Graphs are not sequences.

Basic graph transformer architectures lose all information about adjacency structure.

Common Encodings

Random Walk Encodings

$f_i(v)$ … likelihood of an i-hop random walk from v to end at v

$$RWSE_\ell(v) = [f_i(v)]_{i \in [\ell]}$$

Laplacian Encodings

Derived from the eigenvectors of the (normalised) graph Laplacian for v.

Issue: invariant under sign changes.

Motif Encodings

$$\text{MoSE}_{\mathcal{G},\omega}(v, H) = \left[\omega\text{-Hom}_{\star\to v}(G_i, H)\right]_{i=1}^{d}$$

ω maps vertices of H to weights in $\mathbb{R}$

$\omega\text{-hom}_{\star\to v}(G_i, H)$ is the number of ω weighted homomorphisms that map a fixed node $\star$ of G_i to v in H

Benefits

- Homomorphism counts form the basis of all local functions on graphs.
- MoSE aligns with the existing theory of GNN expressivity → theoretical insight into the expressivity of our structural encoding.
- Generalises RWSE: $RWSE_{\ell}(v) = MoSE_{\mathcal{G},\omega}(v)$

where $\mathcal{G} = \{C_k\}_{k \in [\ell]}$ and $\omega(v) = \dfrac{1}{deg(v)}$

RWSE is weaker than 2-WL

Encoding	PE/SE	ZINC-12K	PCQM4Mv2-subset	CIFAR10
		MAE ↓	MAE ↓	Acc. ↑
none	-	0.113±0.007	0.1355±0.0035	71.49±0.19
PEG^{LapEig}	PE	0.161±0.006	0.1209±0.0003	72.10±0.46
LapPE	PE	0.116±0.009	0.1201±0.0003	72.31±0.34
SignNet^{MLP}	PE	0.090±0.007	0.1158±0.0008	71.74±0.60
$\text{SignNet}^{DeepSets}$	PE	0.079±0.006	0.1144±0.0002	72.37±0.34
RWSE	SE	0.070±0.002	0.1159±0.0004	71.96±0.40
MoSE (*ours*)	SE	**0.062**±0.002	**0.1133**±0.0014	**73.50**±0.44

Conclusion

Summary

Graph motif parameters are a powerful and general framework for studying functions on graphs.

They provide a tight link between machine learning, model theory, and complexity theory.

Probably, there is much low-hanging fruit still left.

Future Work

Generalise expressivity results from graphs to relational structures and to common query languages:

"Which databases can I (not) distinguish by counting answers to some fixed CQ/UCQ/ Datalog query?"