ScaDS.AI
DRESDEN LEIPZIG

CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

38th TBI Winterseminar in Bled
February 2023

Talk: **Comparison of Atom Maps**

Speaker: **PhD student - Marcos E. Gonzalez Laffitte**

Advisor: **Prof. Dr. Peter F. Stadler**

A work in collaboration with Nora Beier and Nico Domschke

Bioinformatik Uni Leipzig
ScaDS.AI Leipzig

GEFÖRDERT VOM

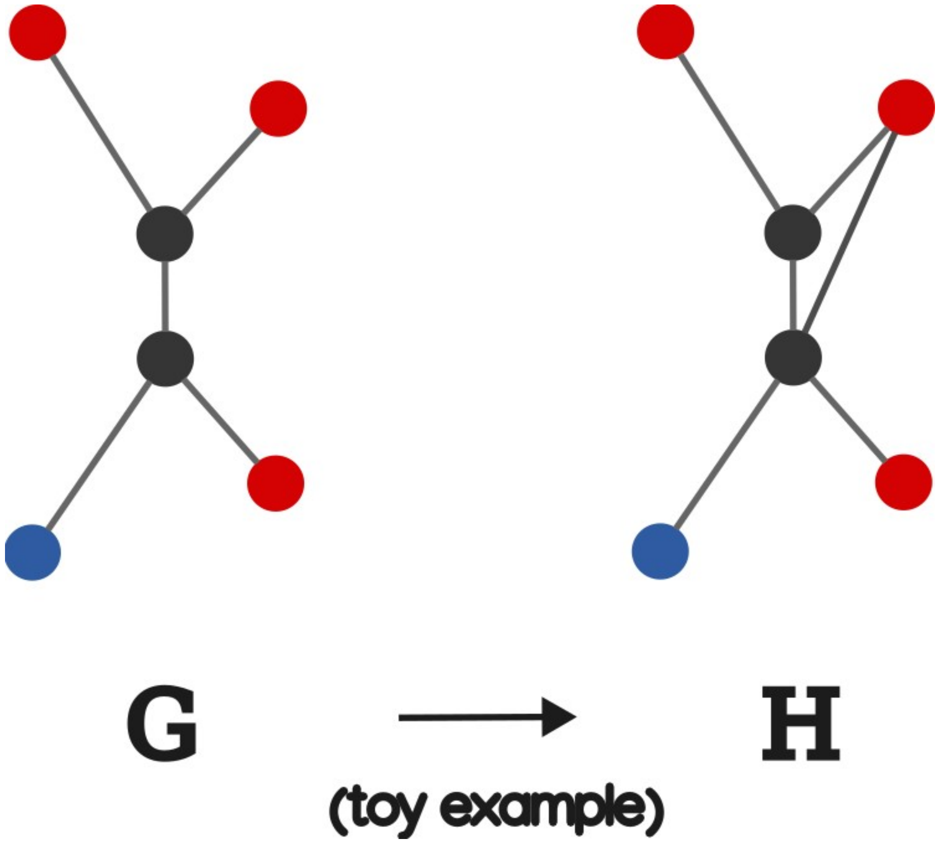
Bundesministerium für Bildung und Forschung

SACHSEN Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

TECHNISCHE UNIVERSITÄT DRESDEN

UNIVERSITÄT LEIPZIG

Introduction: Atom-to-Atom Maps

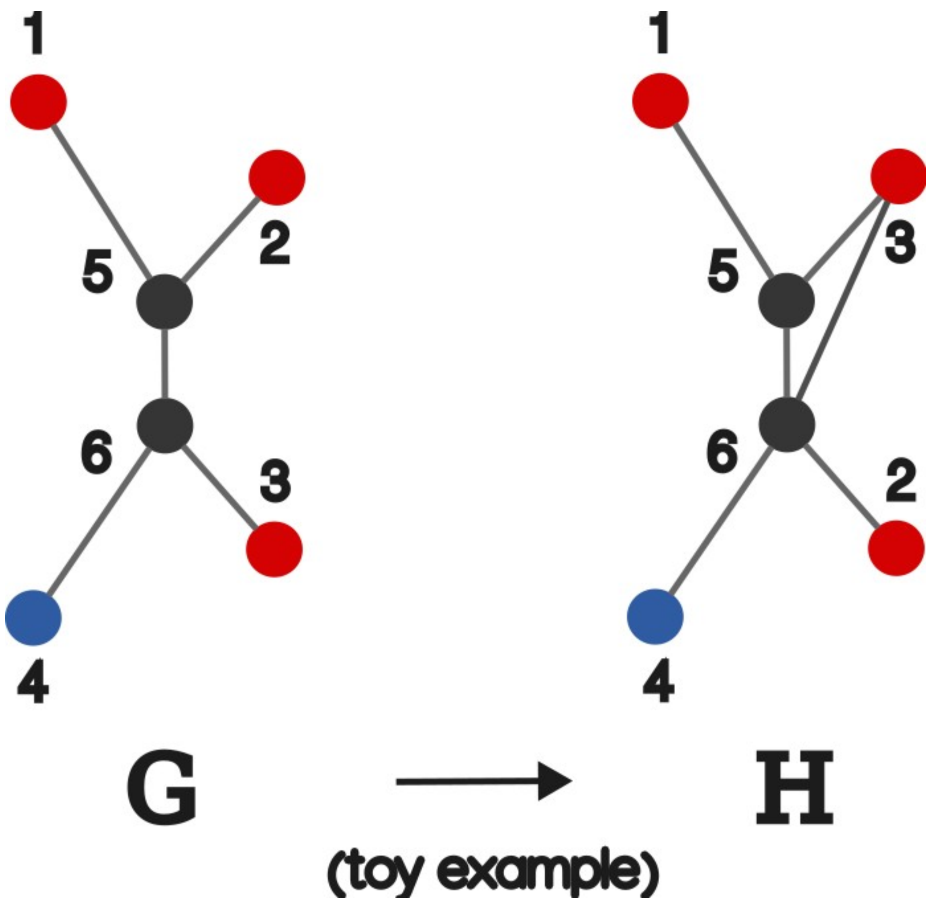


Intuitively speaking an **atom map** is a one-to-one correspondence describing how the atoms in the reactants have to be rearrange in order to obtain the products of a chemical reaction, thus *encoding* the mechanism of the reaction.

Definition

Let G and H be two simple graphs with associated vertex labelings $a_G : V(G) \to L_V$ and $a_H : V(H) \to L_V$. An **atom map** from G to H is a bijection $\alpha : V(G) \to V(H)$ which preserves vertex labels, i.e., $a_H(\alpha(x)) = a_G(x)$ for all $x \in V(G)$.

Introduction: Atom-to-Atom Maps

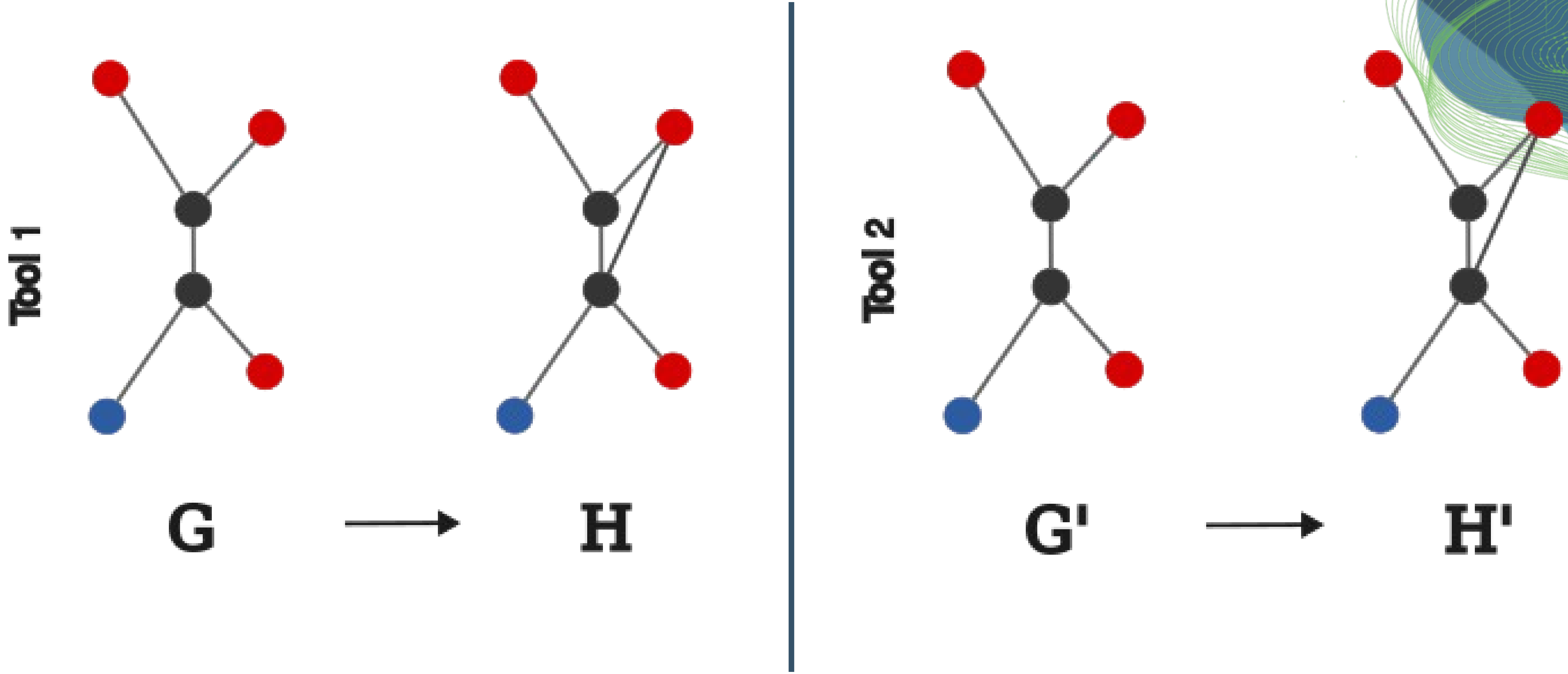


Intuitively speaking an **atom map** is a one-to-one correspondence describing how the atoms in the reactants have to be rearrange in order to obtain the products of a chemical reaction, thus *encoding* the mechanism of the reaction.

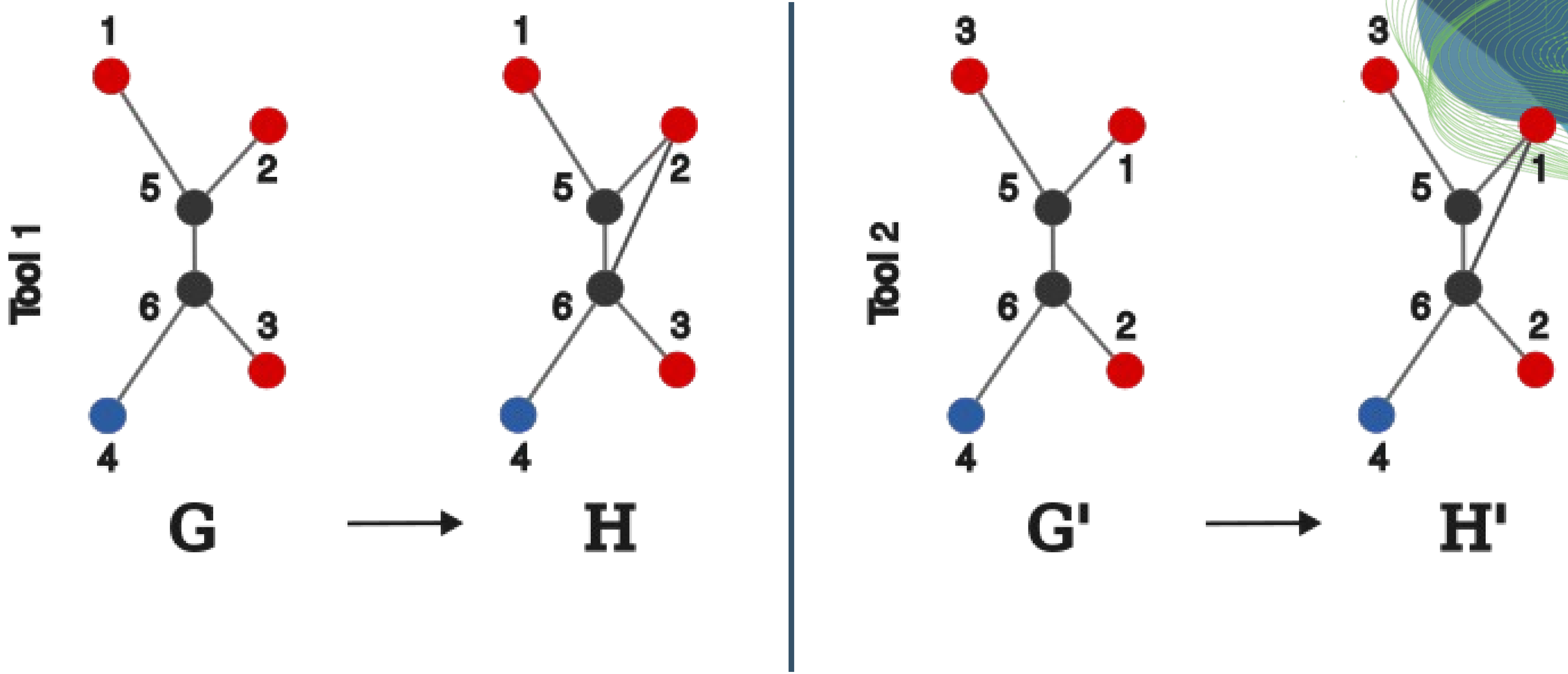
Definition

Let G and H be two simple graphs with associated vertex labelings $a_G : V(G) \to L_V$ and $a_H : V(H) \to L_V$. An **atom map** from G to H is a bijection $\alpha : V(G) \to V(H)$ which preserves vertex labels, i.e., $a_H(\alpha(x)) = a_G(x)$ for all $x \in V(G)$.

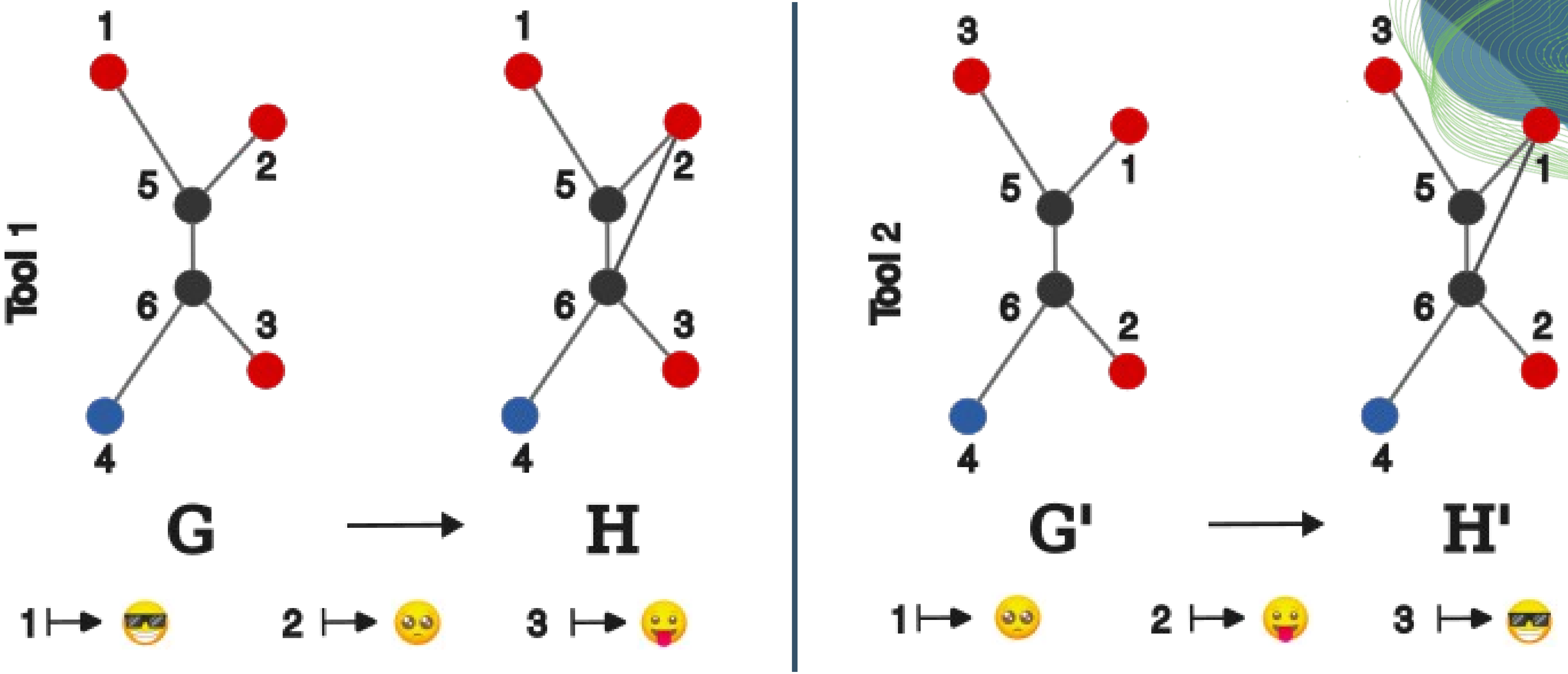
Problem: How do we know if two Atom Maps *encode* the same information or not?



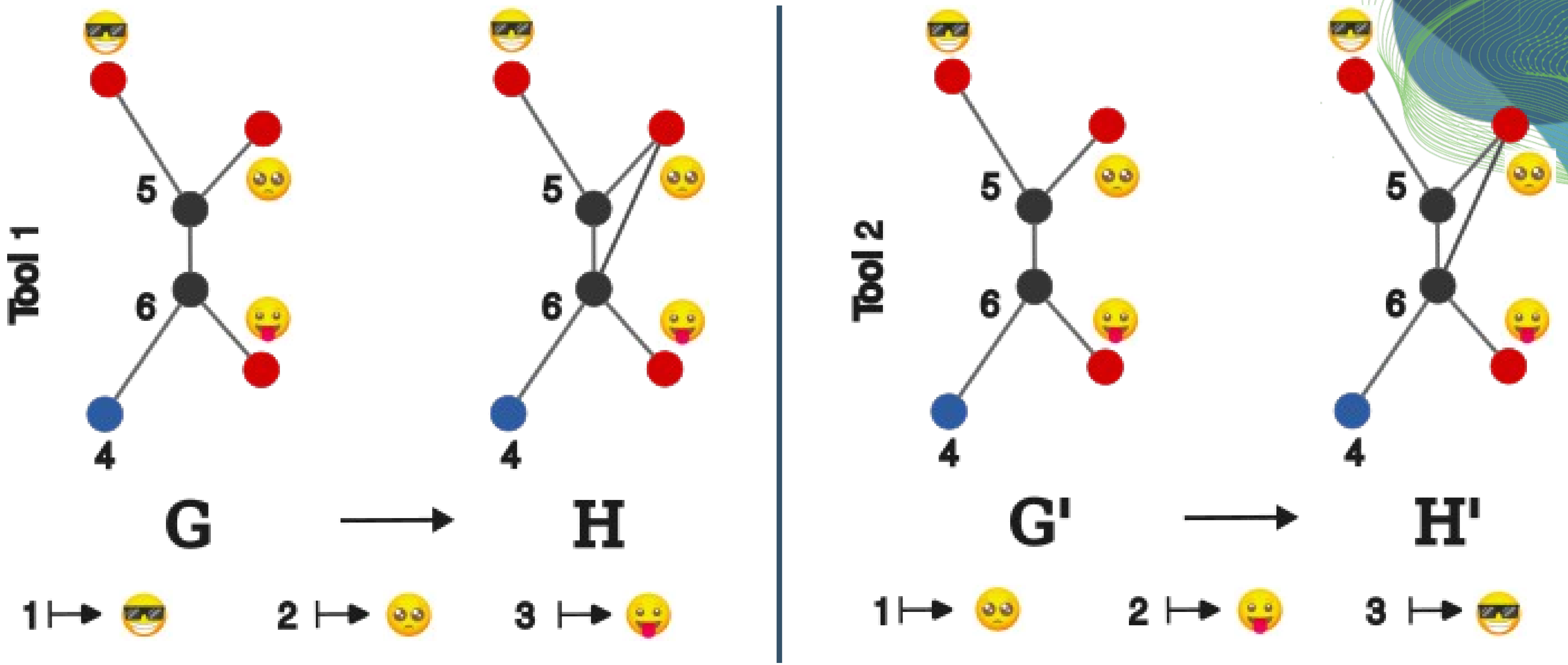
Problem: How do we know if two Atom Maps *encode* the same information or not?



Problem: How do we know if two Atom Maps *encode* the same information or not?

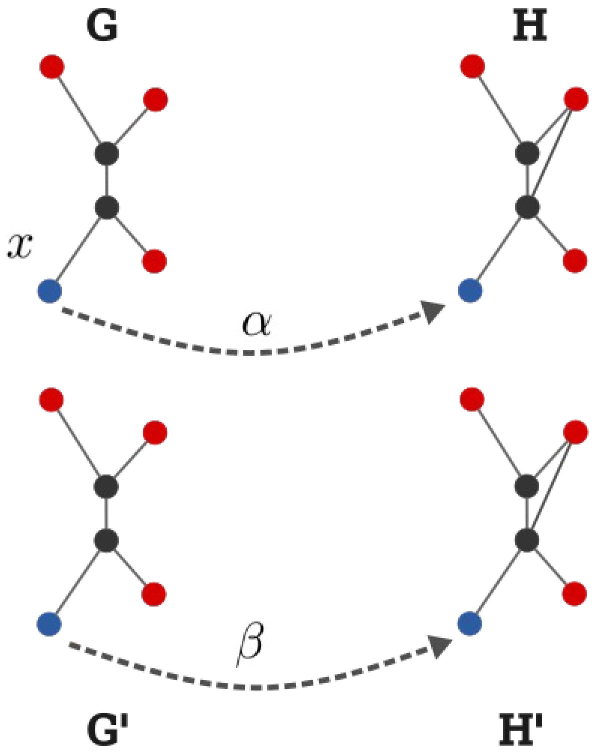


Problem: How do we know if two Atom Maps *encode* the same information or not?



Defining Equivalent Atom Maps

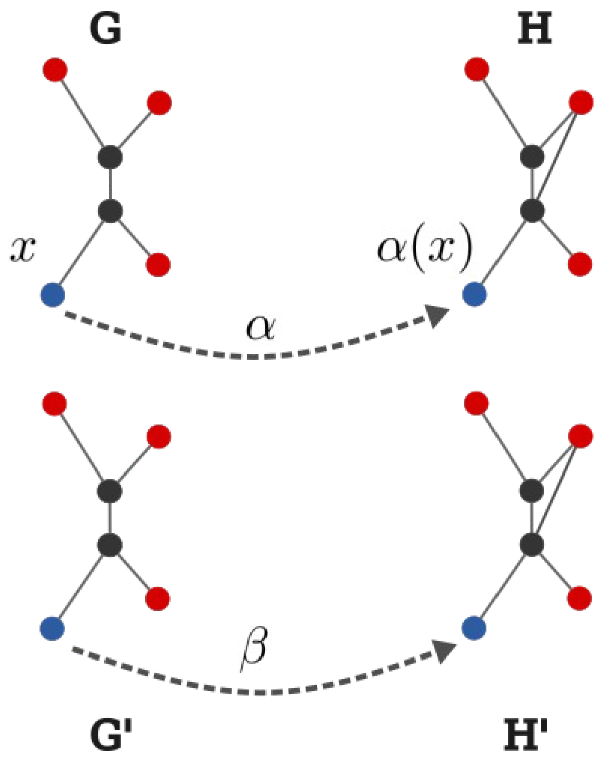
(Motivation)



ScaDS.AI DRESDEN LEIPZIG TECHNISCHE UNIVERSITÄT DRESDEN UNIVERSITÄT LEIPZIG

Defining Equivalent Atom Maps

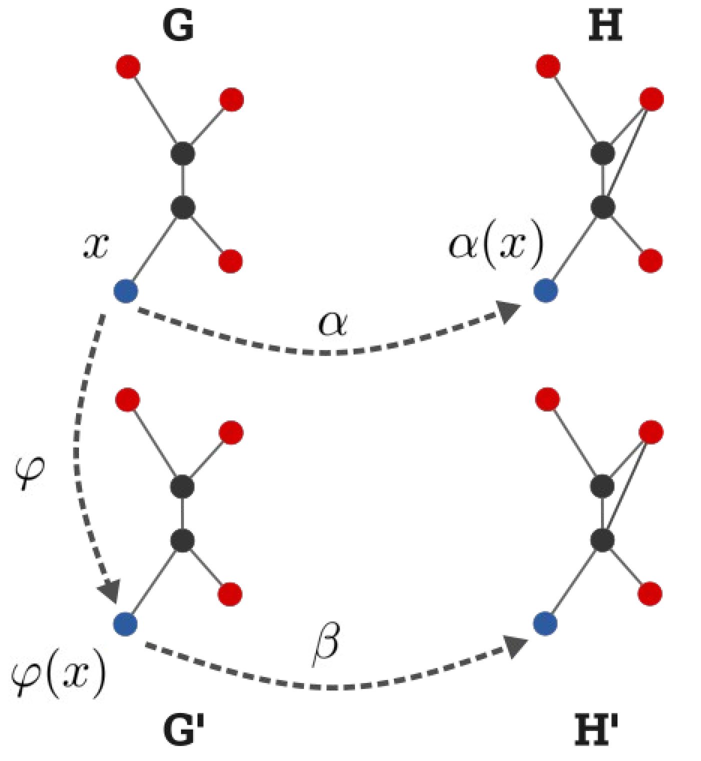
(Motivation)



$\alpha(x) =$

Defining Equivalent Atom Maps

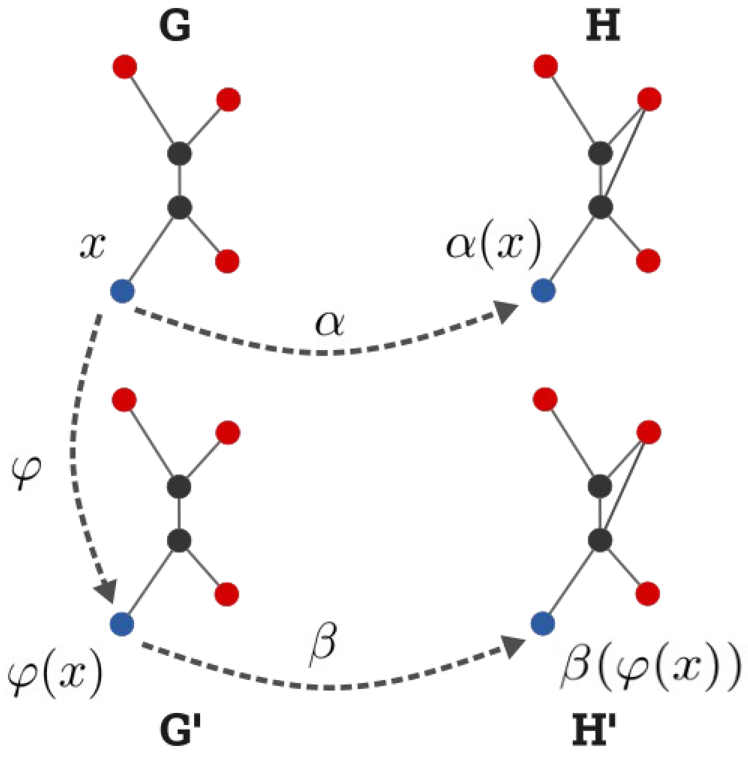
(Motivation)



$\alpha(x) = \quad \varphi(x) \quad \text{for } \varphi \in Iso^L(G, G')$

Defining Equivalent Atom Maps

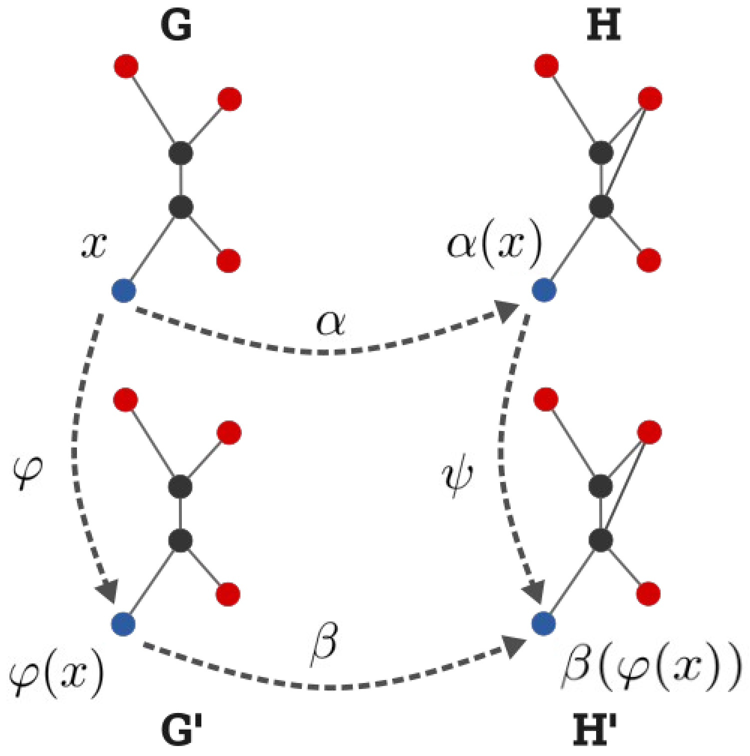
(Motivation)



$$\alpha(x) = \quad \beta(\varphi(x)) \quad \text{for } \varphi \in Iso^L(G, G')$$

Defining Equivalent Atom Maps

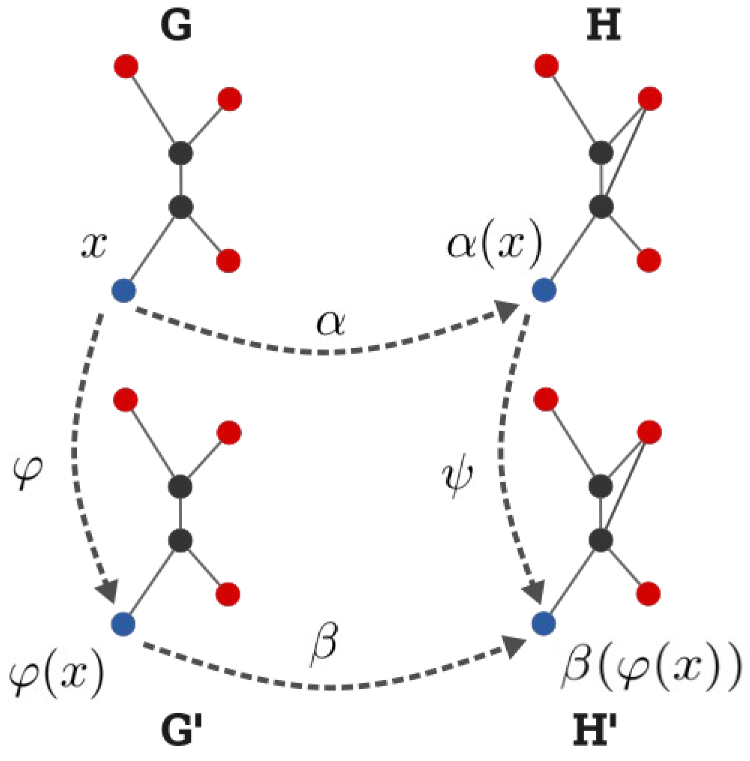
(Motivation)



$$\alpha(x) = \psi^{-1}(\beta(\varphi(x))), \text{ for } \varphi \in Iso^L(G, G'), \psi \in Iso^L(H, H')$$

Defining Equivalent Atom Maps

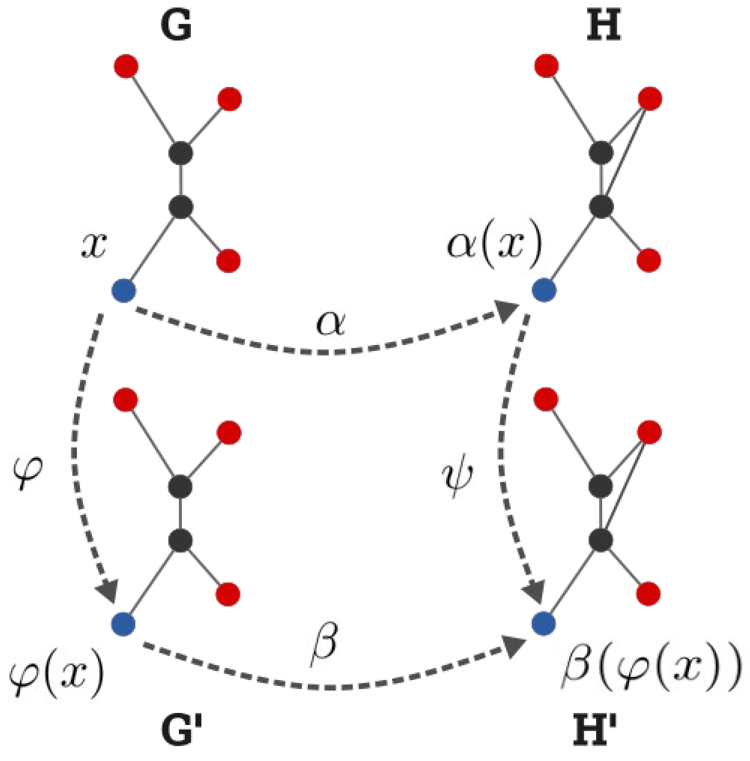
(Motivation)



$$\alpha(x) = \psi^{-1}(\beta(\varphi(x))), \text{ for } \varphi \in Iso^L(G, G'), \psi \in Iso^L(H, H')$$
$$\text{for all } x \in V(G)$$

Defining Equivalent Atom Maps

(Motivation)



$$\alpha(x) = \psi^{-1}(\beta(\varphi(x))), \text{ for } \varphi \in Iso^L(G, G'), \psi \in Iso^L(H, H')$$
$$\text{for all } x \in V(G)$$

Definition

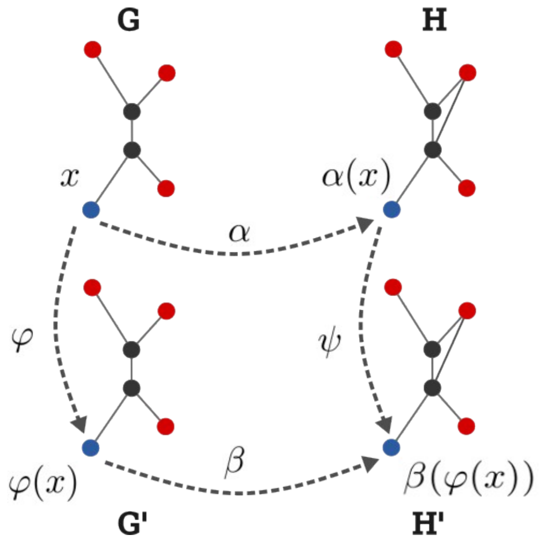
We say that two atom maps $\alpha : V(G) \to V(H)$ and $\beta : V(G') \to V(H')$ are **equivalent**, in symbols $\alpha \equiv \beta$, if there are $\varphi \in Iso^L(G, G')$ and $\psi \in Iso^L(H, H')$ such that $\psi \circ \alpha = \beta \circ \varphi$.

Lemma

The equivalence of atom maps, $\equiv$, is an equivalence relation.

Idea of the proof. Follows since the identity, and the inverses and the compositions of isomorphisms, are all isomorphisms.

Comparing Atom Maps
Method 1: [ISO-$\equiv$] Exhaustive Enumeration of Isomorphisms

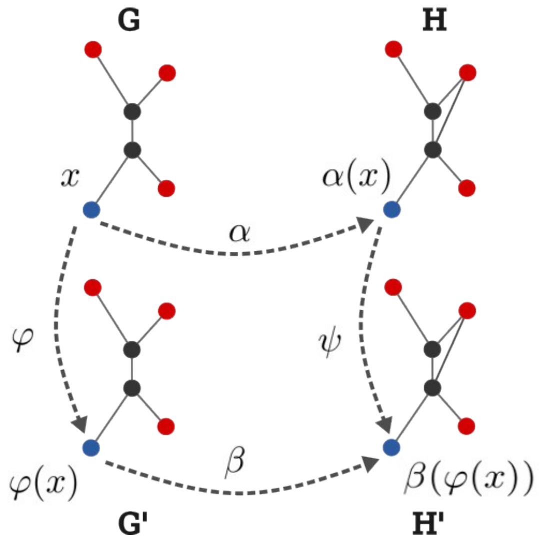


forall $\varphi \in Iso^L(G, G')$ **do**
  forall $\psi \in Iso^L(H, H')$ **do**
    if ($\psi = \beta \circ \varphi \circ \alpha^{-1}$) **then**
      return ($"\alpha \equiv \beta", \varphi, \psi$)
    end
  end
end

Definition

We say that two atom maps $\alpha : V(G) \to V(H)$ and $\beta : V(G') \to V(H')$ are **equivalent**, in symbols $\alpha \equiv \beta$, if there are $\varphi \in Iso^L(G, G')$ and $\psi \in Iso^L(H, H')$ such that $\psi \circ \alpha = \beta \circ \varphi$.

Comparing Atom Maps
Method 1: [ISO-≡] Exhaustive Enumeration of Isomorphisms




```
forall φ ∈ Iso^L(G, G') do
    forall ψ ∈ Iso^L(H, H') do
        if ( ψ = β ∘ φ ∘ α^{-1} ) then
            return ( "α ≡ β", φ, ψ )
        end
    end
end
```

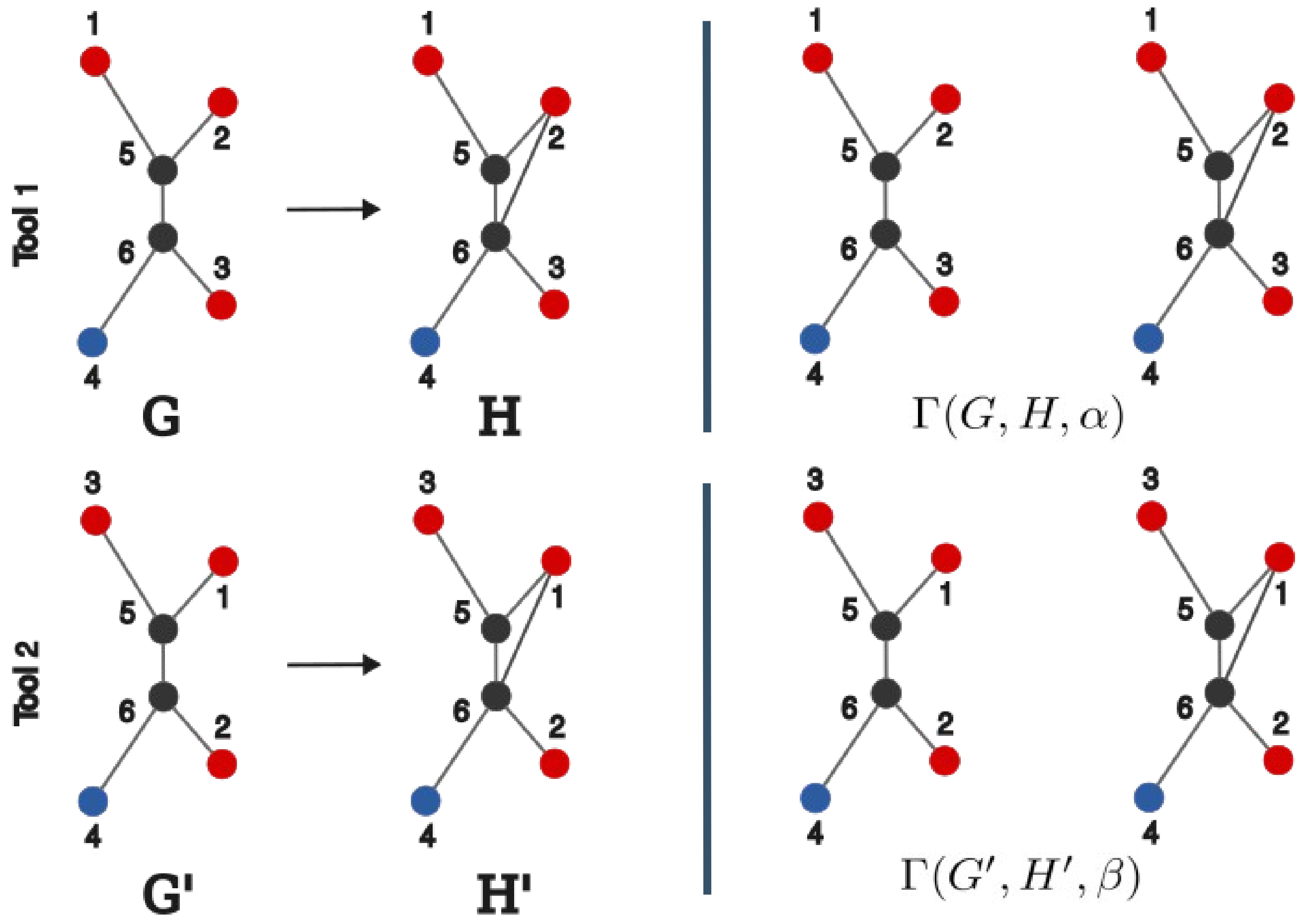
Definition

We say that two atom maps $\alpha : V(G) \to V(H)$ and $\beta : V(G') \to V(H')$ are **equivalent**, in symbols $\alpha \equiv \beta$, if there are $\varphi \in Iso^L(G, G')$ and $\psi \in Iso^L(H, H')$ such that $\psi \circ \alpha = \beta \circ \varphi$.

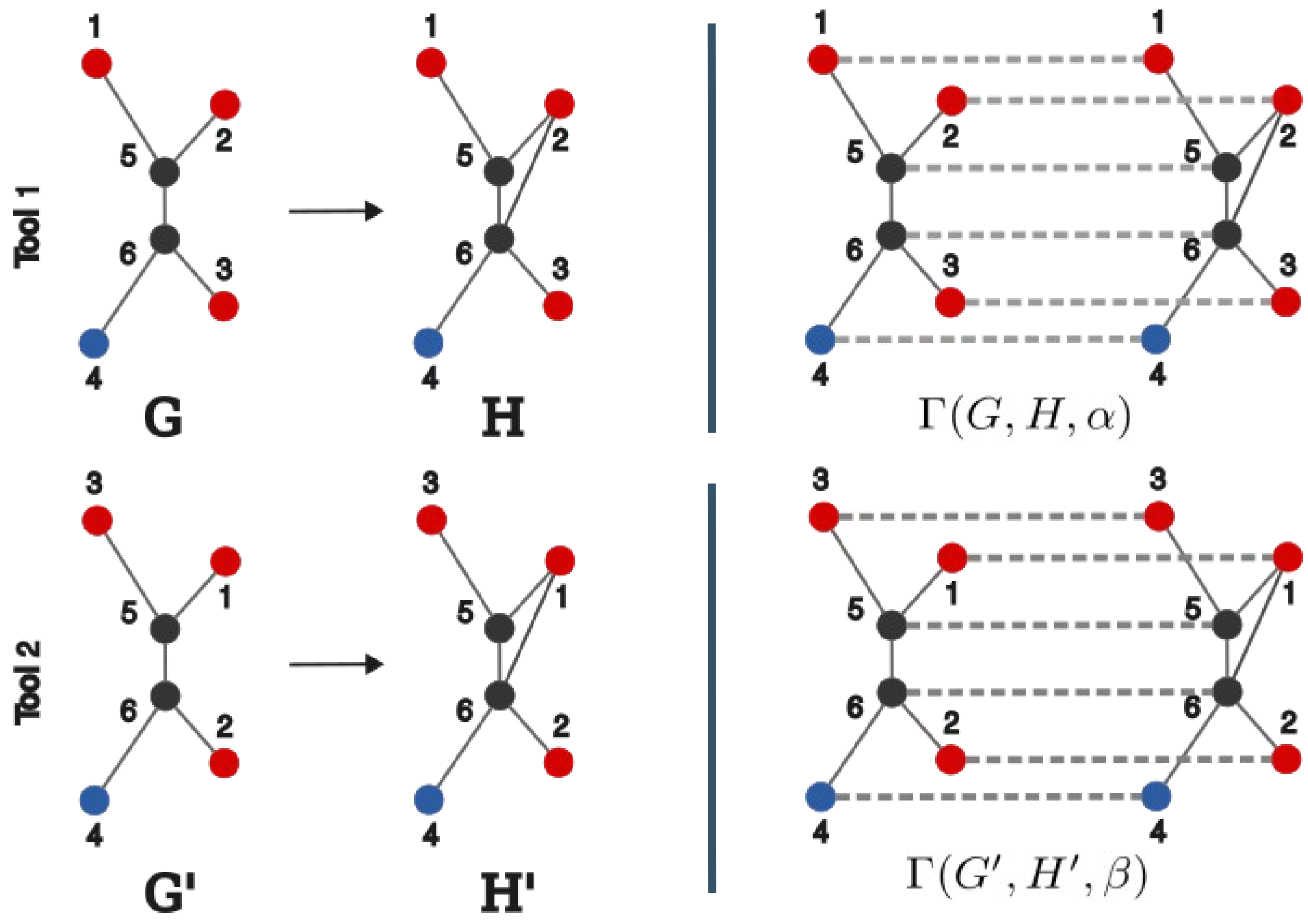
Lemma

Two atom maps $\alpha : V(G) \to V(H)$ and $\beta : V(G') \to V(H')$ are equivalent iff there is $\varphi \in \mathrm{Iso}^L(G, G')$ such that $\beta \circ \varphi \circ \alpha^{-1} \in \mathrm{Iso}^L(H, H')$.

Comparing Atom Maps
Method 2: [AUX-Γ] Comparison of Auxiliary Graphs



Comparing Atom Maps
Method 2: [AUX-Γ] Comparison of Auxiliary Graphs



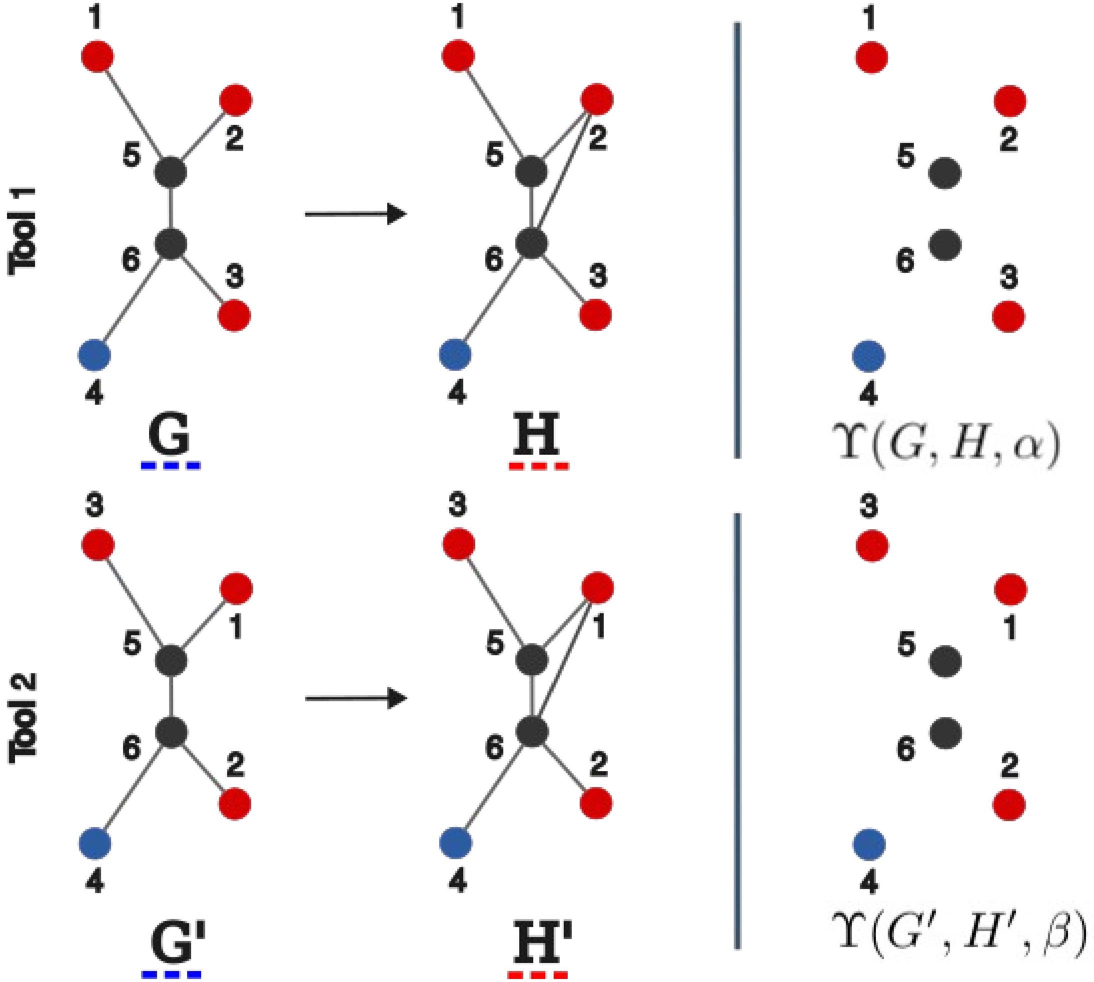
Theorem

Two atom maps $\alpha : V(G) \to V(H)$ and $\beta : V(G') \to V(H')$ are equivalent **iff** their labeled auxiliary graphs are isomorphic, that is, $\Gamma(G, H, \alpha) \simeq \Gamma(G', H', \beta)$.

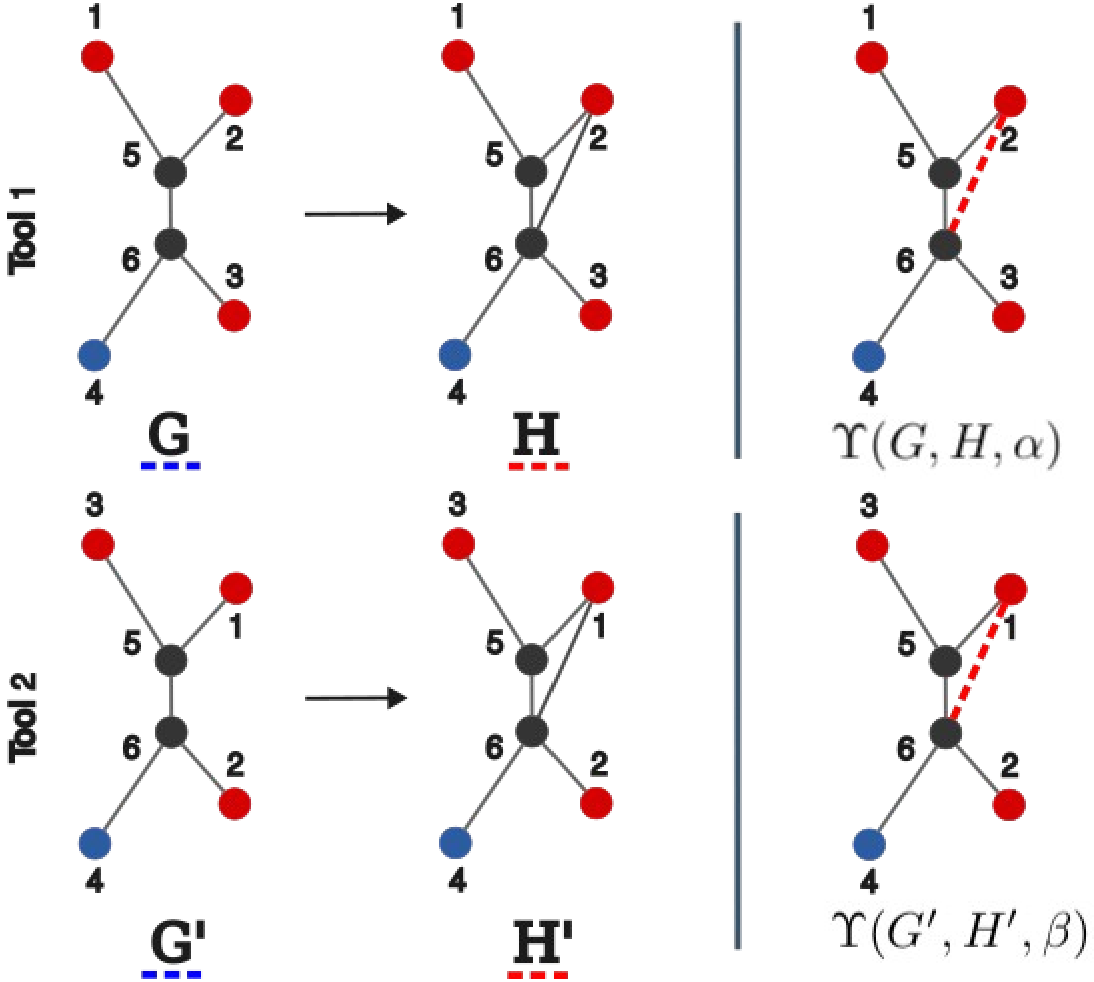
Idea of the proof.

If $\alpha \equiv \beta$ there are $\varphi \in \mathrm{Iso}^L(G, G')$ and $\psi \in \mathrm{Iso}^L(H, H')$ such that $\psi \circ \alpha = \beta \circ \varphi$ and so $\zeta := \varphi \dot\cup \psi$ can be proven to be an isomorphism from $\Gamma(G, H, \alpha)$ to $\Gamma(G', H', \beta)$. On the other hand, given an isomorphism ζ from $\Gamma(G, H, \alpha)$ to $\Gamma(G', H', \beta)$ one can prove that the restrictions ζ_G and ζ_H of ζ to, respectively, $V(G)$ and $V(H)$, are isomorphisms $\zeta_G \in \mathrm{Iso}^L(G, G')$ and $\zeta_H \in \mathrm{Iso}^L(H, H')$, satisfying $\zeta_H \circ \alpha = \beta \circ \zeta_G$, provided that the vertices in $\Gamma(G, H, \alpha)$ have labels of the form $(a_G(x), 1)$ if $x \in V(G)$ and $(a_H(x), 2)$ if $x \in V(H)$, and similarly for $\Gamma(G', H', \beta)$.

Comparing Atom Maps
Method 3: [ITS-ϒ] Comparison of Imaginary Transition Structures



Comparing Atom Maps
Method 3: [ITS-Υ] Comparison of Imaginary Transition Structures



Corollary

Two atom maps $\alpha : V(G) \to V(H)$ and $\beta : V(G') \to V(H')$ are equivalent **iff** their ITS graphs are isomorphic, that is, $\Upsilon(G, H, \alpha) \simeq \Upsilon(G', H', \beta)$.

Idea behind the main theorem.

The main theorem makes use of (a version of) the labeled auxiliary graphs $\Gamma(G, H, \alpha)$ and $\Gamma(G', H', \beta)$. Intuitively speaking the restriction to $V(G)$ of an isomorphism ζ from $\Gamma(G, H, \alpha)$ to $\Gamma(G', H', \beta)$ is an isomorphism from $\Upsilon(G, H, \alpha)$ to $\Upsilon(G', H', \beta)$. On the other hand, an isomorphism ξ from $\Upsilon(G, H, \alpha)$ to $\Upsilon(G', H', \beta)$ can be *extended* into an isomorphism θ from $\Gamma(G, H, \alpha)$ to $\Gamma(G', H', \beta)$ by setting $\theta(x) = \xi(x)$ for $x \in V(G)$ and $\theta(x) = \beta(\xi(\alpha^{-1}(x)))$ if $x \in V(H)$.

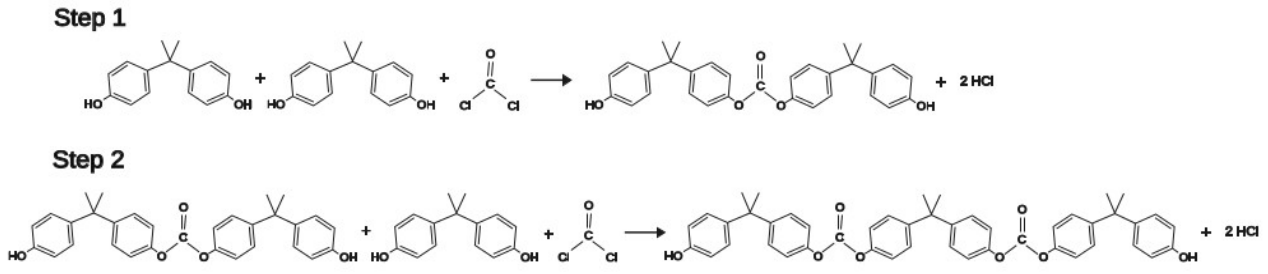
The ITS graphs were first introduced by Shinsaku Fujita in **Description of organic reactions based on imaginary transition structures. 1. introduction of new concepts**, J. Chem. Inf. Comput. Sci. 26 **(1986)** 205–212.

Later another version of them was introduced as *Condensed Graph of a Reaction* (or CGR) by Hoonakker et.al., in **A representation to apply usual data mining techniques to chemical reactions – illustration on the rate constant of SN2 reactions in water**, Int. J. Artif. Intelligence Tools 20 **(2011)** 253–270.

Computational Implementation and Running Time Comparison of the three Methods

We implemented the three methods mentioned before and analyzed the output of three existing Atom-Mapping Tools: RXNMapper, RDT and GraphormerMapper.

Polycondensation of bisphenol A (BPA) and phosgene

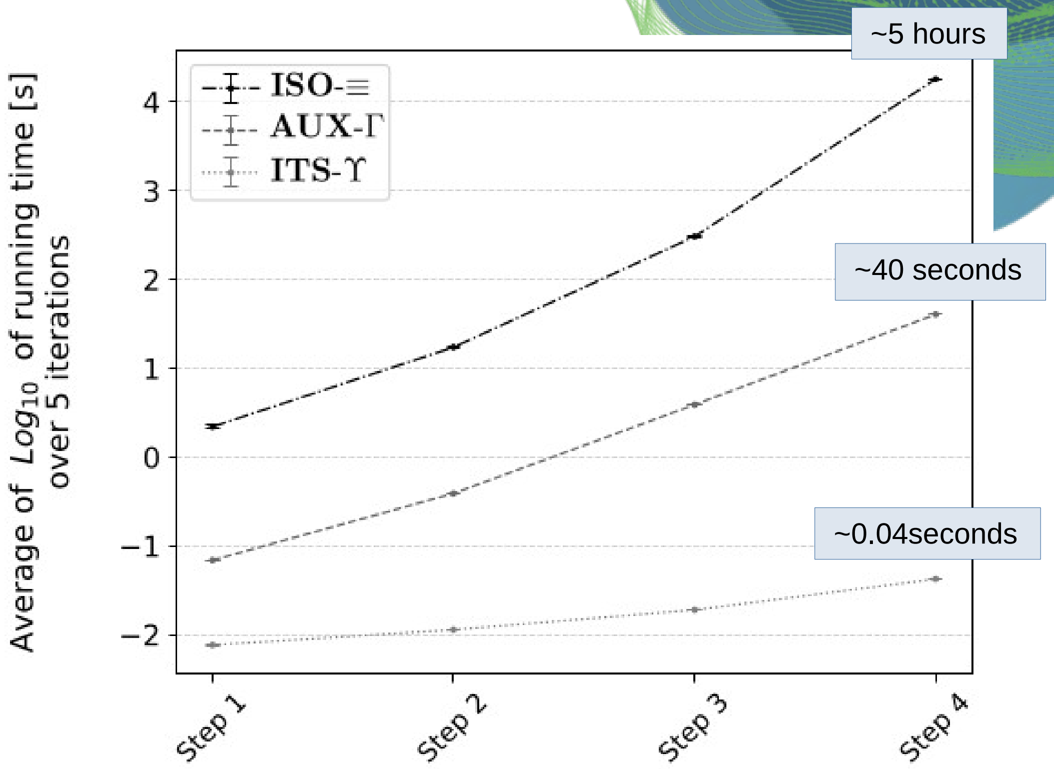


Step 1 $|\mathrm{Aut}(G)| = 1024$ $|\mathrm{Aut}(H)| = 256$

Step 2 $|\mathrm{Aut}(G)| = 4096$ $|\mathrm{Aut}(H)| = 2048$

Step 3 $|\mathrm{Aut}(G)| = 32,768$ $|\mathrm{Aut}(H)| = 16,384$

Step 4 $|\mathrm{Aut}(G)| = 262,144$ $|\mathrm{Aut}(H)| = 131,072$



Manuscript Accepted for Publication

MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. 00 (0000) 1–28
ISSN: 0340–6253
doi: 10.46793/match.xx-x.xxxX

Comparison of Atom Maps

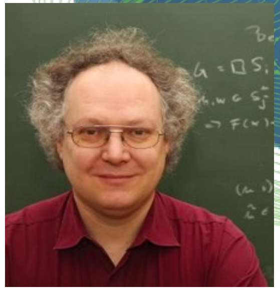
Marcos E. González Laffitte[1,2], Nora Beier[1], Nico Domschke[1], Peter F. Stadler[1−7,*]

[1]Bioinformatics Group, Department of Computer Science & Interdisciplinary Center for Bioinformatics & Leipzig University, D-04107 Leipzig, Germany

[2]Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Leipzig University, D-04103 Leipzig, Germany

Nora Beier

Nico Domschke



Peter F. Stadler

EEquAAM - Github Repository
Evaluation of the Equivalence of Atom-to-Atom Maps

- A. Lin et.al., **Atom-to-atom mapping: A benchmarking study of popular mapping algorithms and consensus strategies**, Molecular Inf. 41 **(2021)** 2100138.

- R. Nugmanov et.al, **Bidirectional graphormer for reactivity understanding: Neural network trained to reaction atom-to-atom mapping task**, Journal of Chemical Information and Modeling 62 **(2022)** 3307–3315.

Thank you for your attention

Funding Institution

Federal Ministry of Education and Research of Germany and by the Sächsische Staatsministerium für Wissenschaft Kultur und Tourismus in the program Center of Excellence for AI-research “Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig”, **project identification number: ScaDS.AI.**

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

SACHSEN

Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Contact

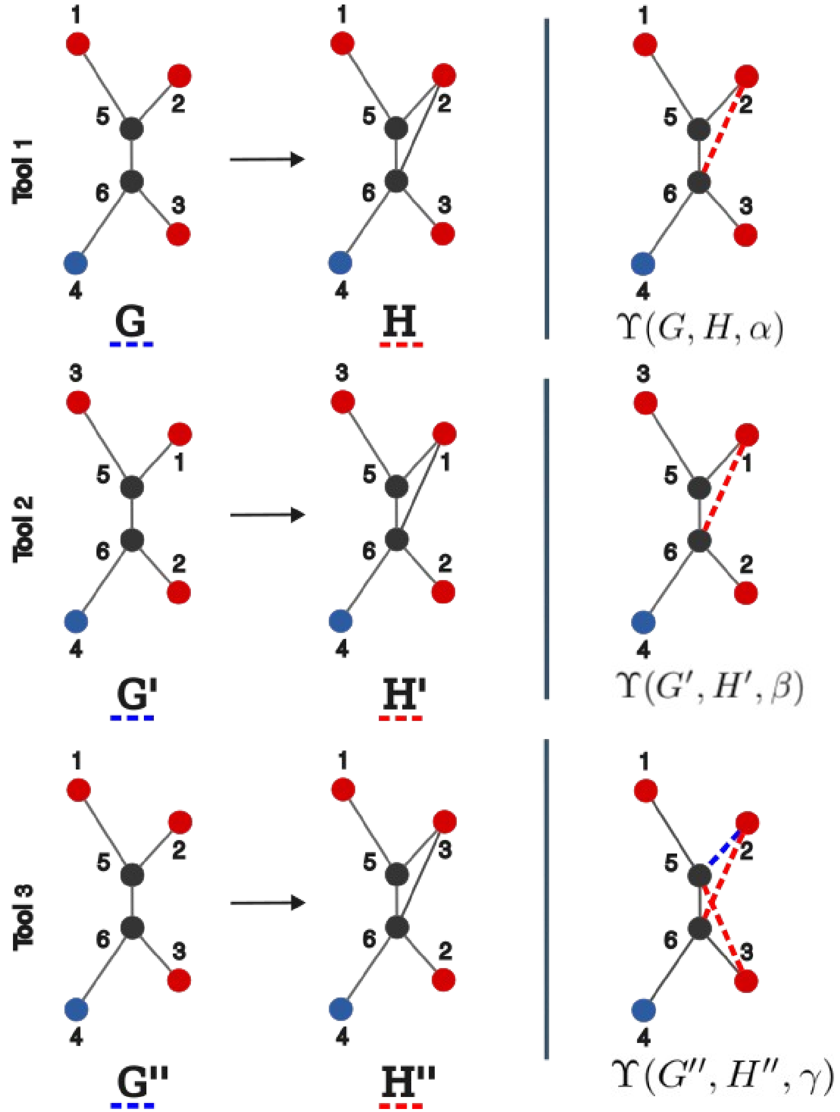
marcos.laffitte@informatik.uni-leipzig.de

ScaDS.AI DRESDEN LEIPZIG

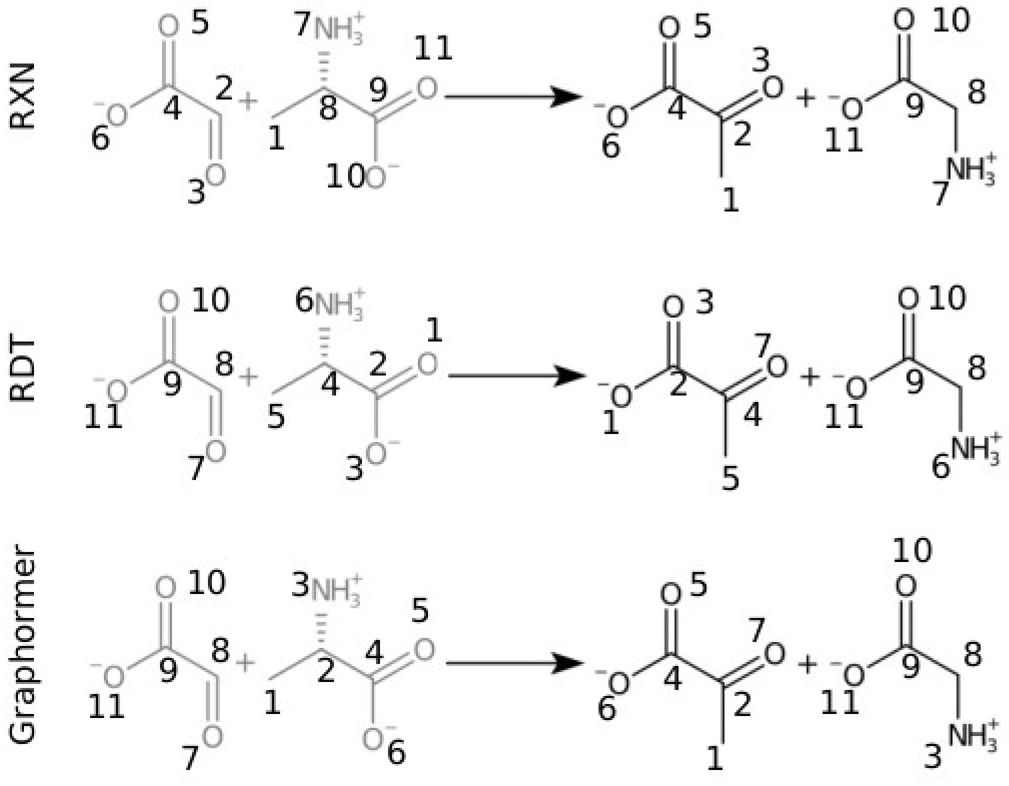
TECHNISCHE UNIVERSITÄT DRESDEN

UNIVERSITÄT LEIPZIG

(Toy-) Example of not-equivalent atom maps...



(Real) Example of not-equivalent atom maps...



Non-equivalent atom maps for the reaction of L-alanine + glyoxylate to pyruvate + glycine.