

Automated Generation of Consistent, Diverse and Structurally Realistic Graph Models



Oszkár Semeráth¹, Aren Babikian², **Bogi Chen**², Chuning Li²,
Kristóf Marussy¹, Gábor Szárnyas¹, Dániel Varró^{1,2}

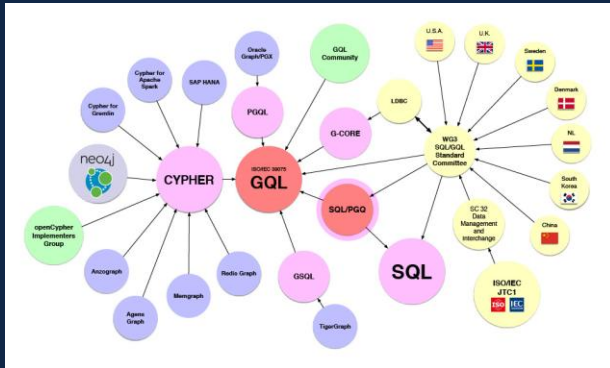
¹*Department of Measurement and Information Systems, Budapest University of Technology and Economics*

²*Department of Electrical & Computer Engineering, McGill University*

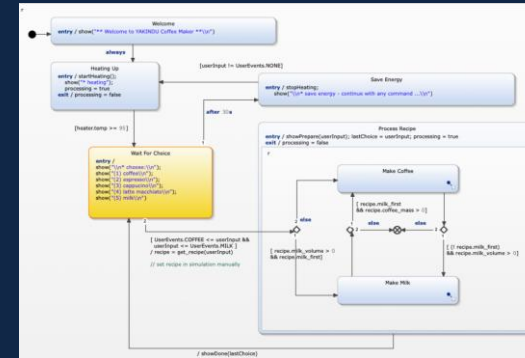
semerath@mit.bme.hu, aren.babikian@mail.mcgill.ca, bogi.chen@mail.mcgill.ca, chuning.li@mail.mcgill.ca, marussy@mit.bme.hu,
szarnyas@mit.bme.hu, varro@mit.bme.hu

Why Graph Models?

Graph Database Testing



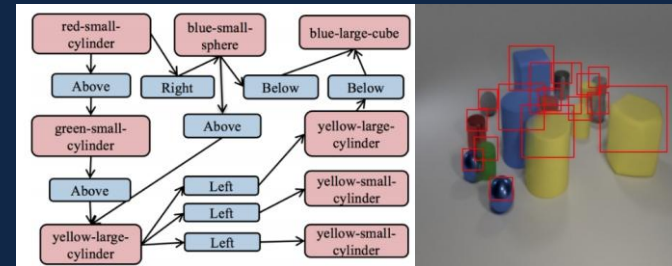
System Modeling



Safety-Critical Systems Testing



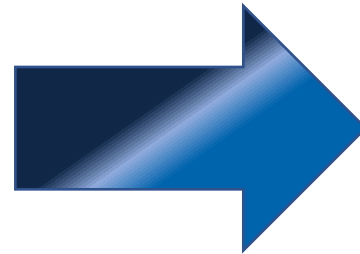
Data for Machine Learning



Why Graph Models Generation?

Limitation of Real (Manually- created) Models:

- **Various** models are needed for testing or training purposes
- Manual creation is **slow** and **expensive**
- Real models **may not be available** due to privacy concerns
- Critical edge cases may be **missing**



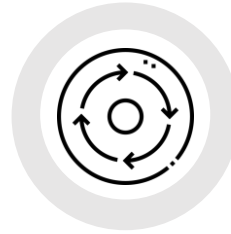
Automated
Graph
Model
Generation!

Motivation

The **structure** of the generated models should...

- Comply with domain-specific well-formedness constraints
- Cover critical edge cases / equivalence classes
- Be structurally similar to the real models in the domain

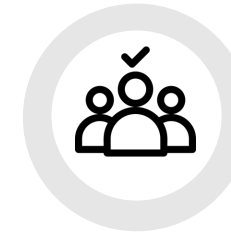
Varró, D., Semeráth, O., Szárnyas, G., & Horváth, Á. (2018).
Towards the Automated Generation of Consistent, Diverse, Scalable and Realistic Graph Models.
Graph Transformation, Specifications, and Nets.



Consistent



Diverse



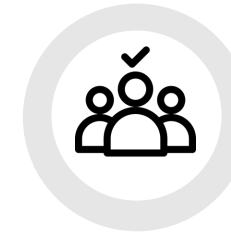
Realistic

Motivation

The generated models should...

- Comply with domain-specific well-formedness constraints
- Cover critical and representative scenarios
- Be realistic

No existing model generators satisfy all three properties!



Realistic

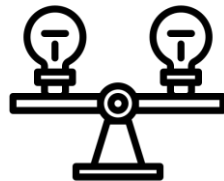
Varró, D., Semeráth, O., Szárnyas, G., & Horváth, Á. (2018).
Towards the Automated Generation of Consistent, Diverse, Scalable and Realistic Graph Models.
Graph Transformation, Specifications, and Nets.

Research Questions

RQ1 How to measure (structural) realisticness of generated graph models?

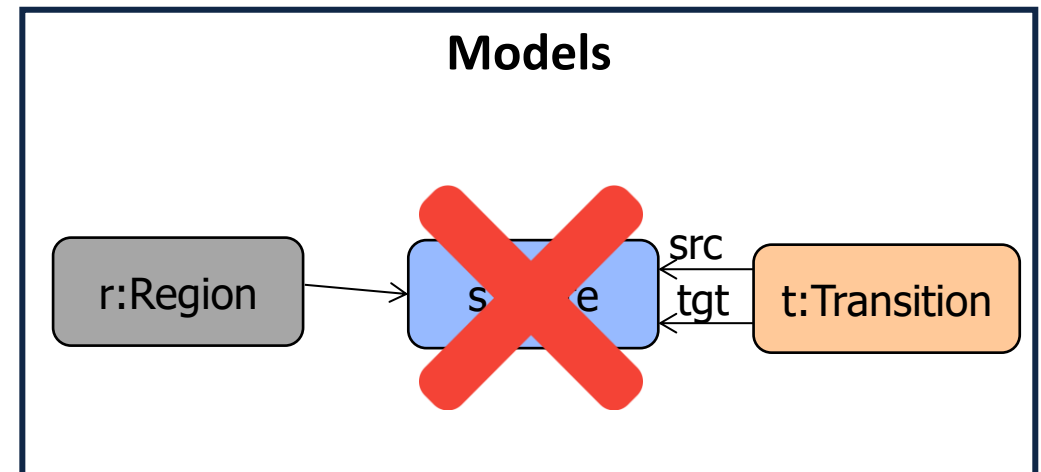
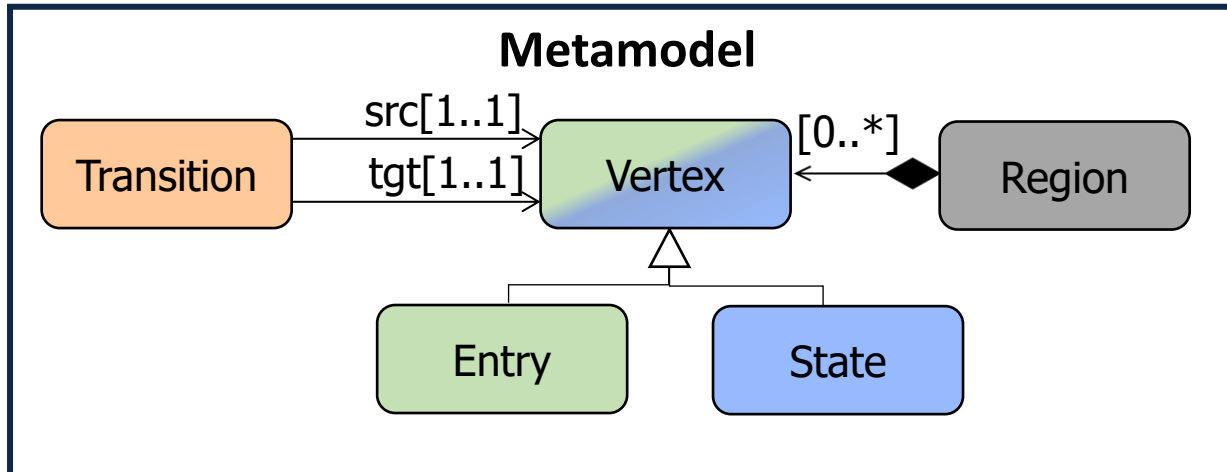


RQ2 How to generate realistic graph models while ensuring consistency and diversity?

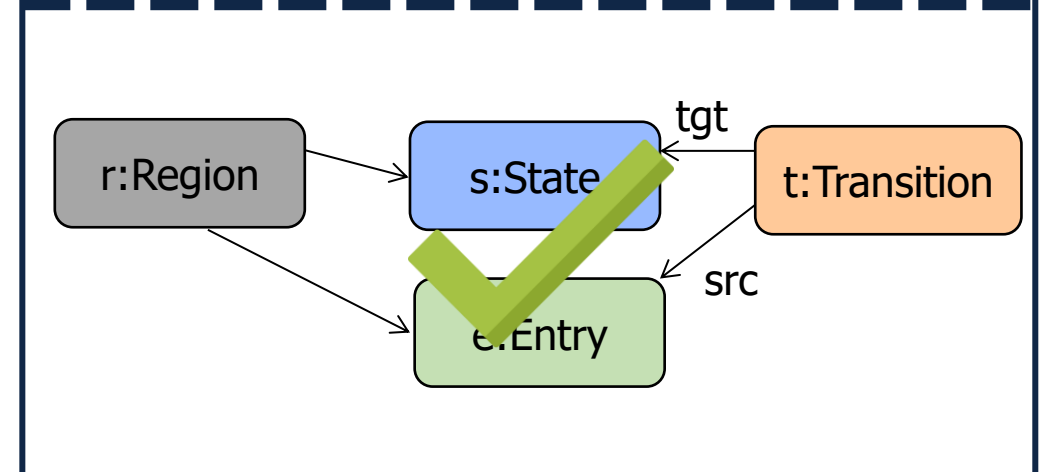


Consistent

The model complies the metamodel and does not violate any well-formedness constraints



- Well-formedness Constraints**
1. Region should have at least one Entry
 2. Entry should not have incoming Transition
 3. Entry should only have one outgoing Transition

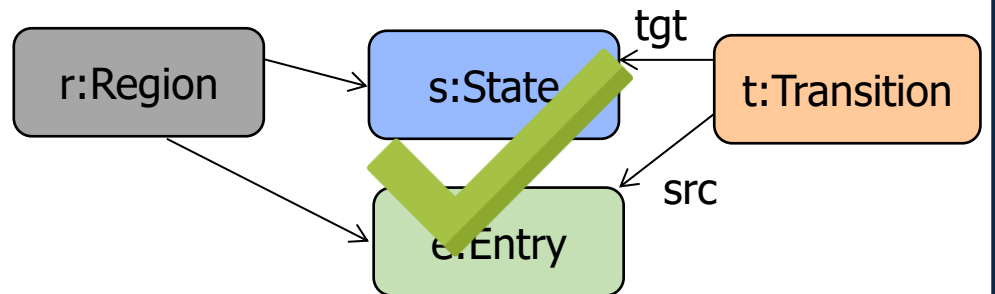
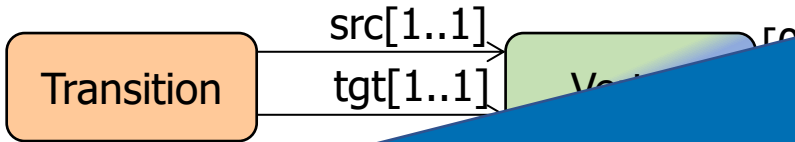


Consistent

The model complies the metamodel and does not violate any well-formedness constraints

Real graph models are usually consistent
Random generated graphs are never consistent

Metamodel

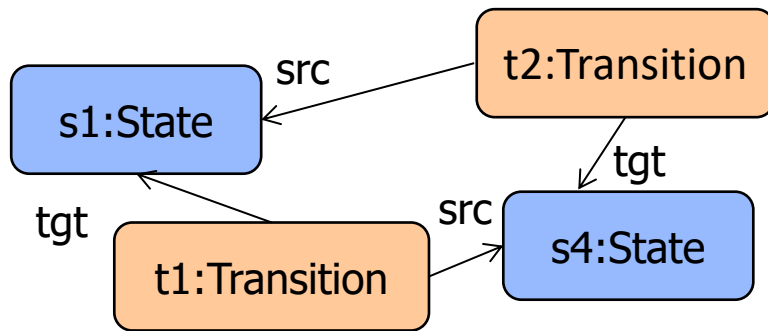


Diverse

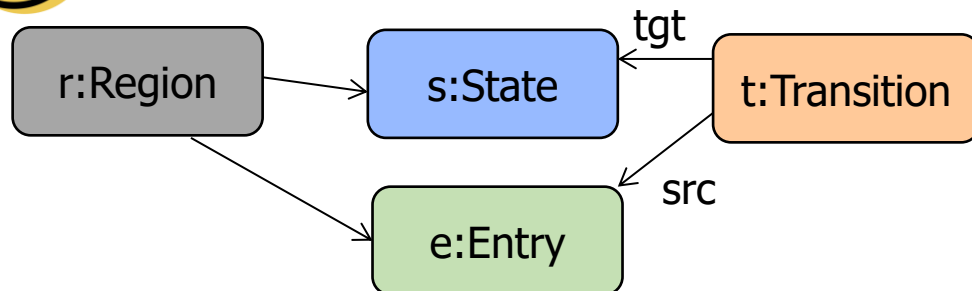
Internal Diversity: Different neighborhood structures in one model



Not Diverse M1



Diverse M2

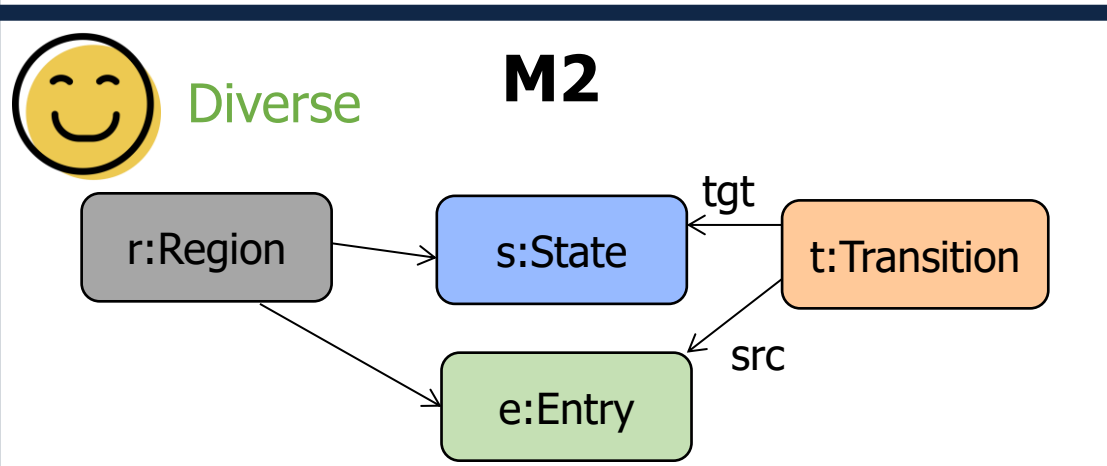
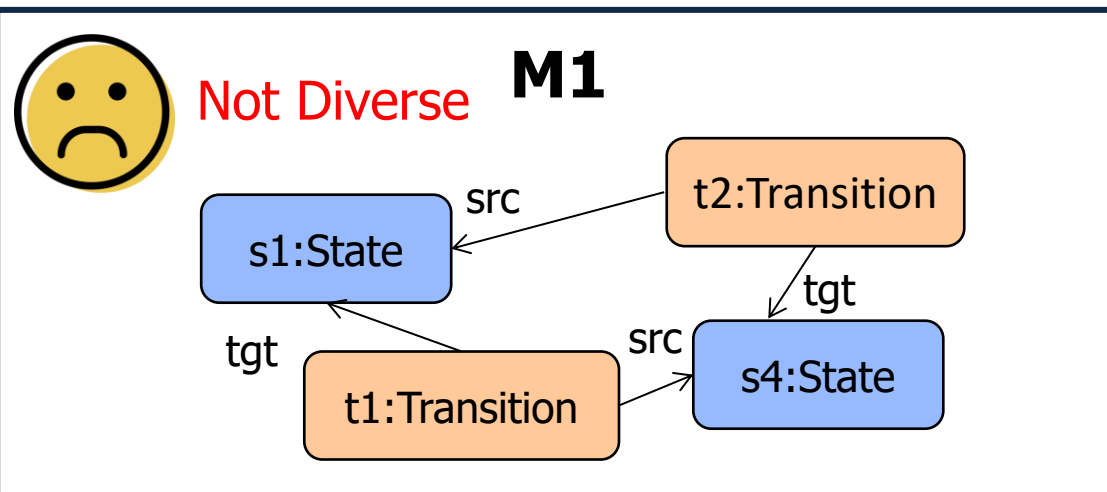


	M1	M2
State -> Transition	1	0
Transition -> State	1	1
Transition -> Entry	0	1
Entry -> Transition	0	0
Region -> State	0	1
Region -> Entry	0	1
Total	2	4

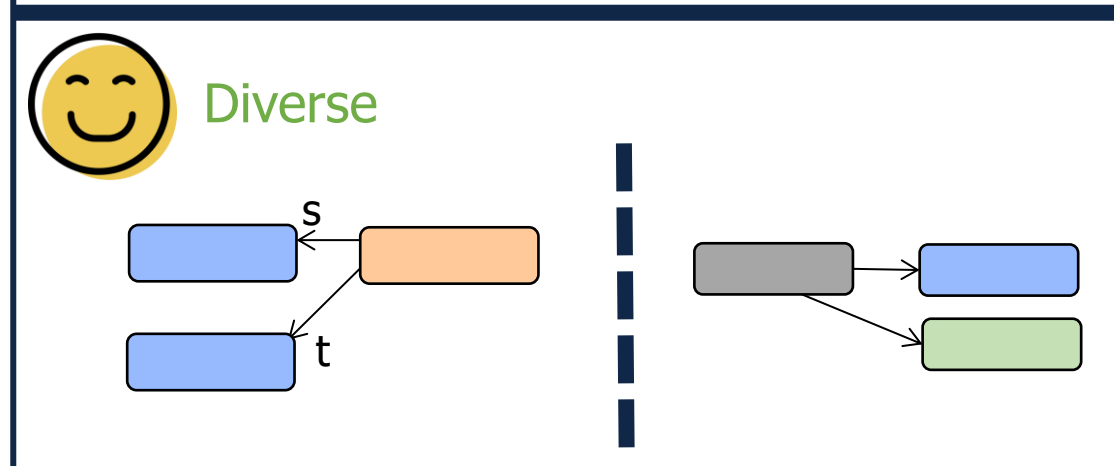
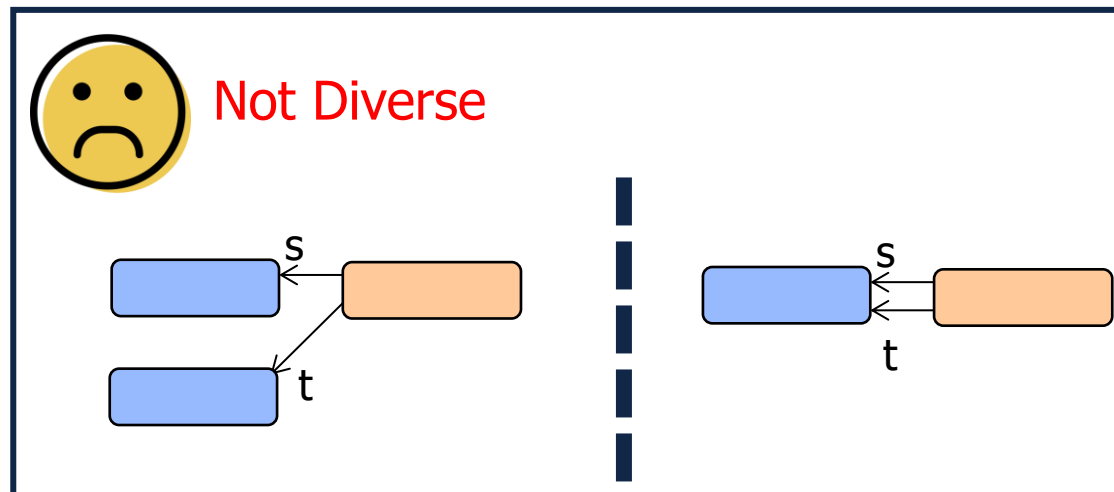
M2 is more diverse

Diverse

Internal Diversity: Different neighborhood structures in one model



External Diversity: Difference in structure for a pair of models



Diverse

Internal Diversity: Different neighborhood structures in one model

External Diversity: Different structure f

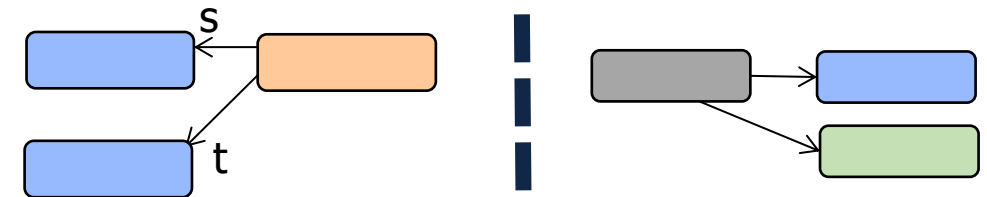
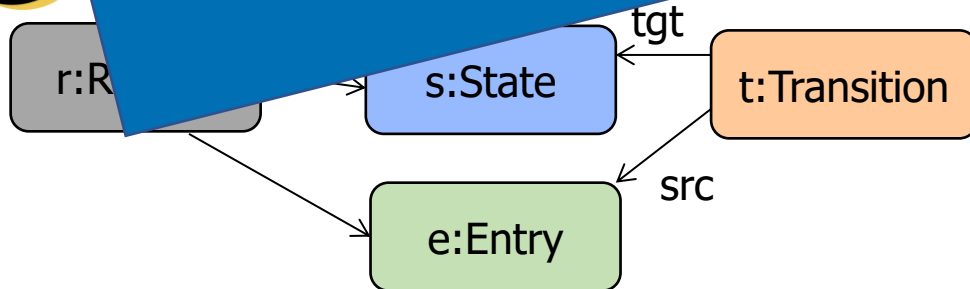


Not Diverse

Manually created graph models are usually not diverse



Diverse

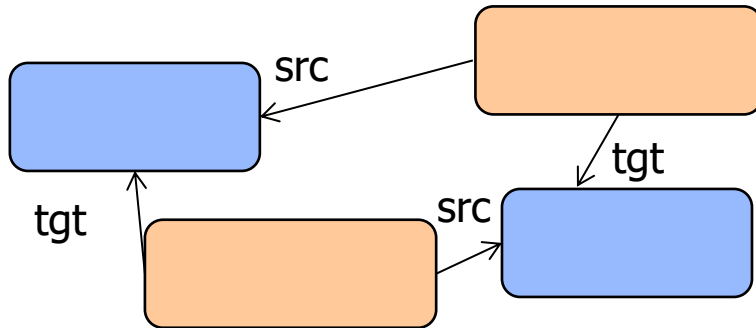


Realistic

Realistic Structure: The generated model structure should be indistinguishable from the manually created graphs



Not Realistic?



This definition is not easily to measure and optimize...

We need a more unambiguous measure of realisticness

New definition: A model is considered structurally realistic if the values of **desired *graph metrics*** are at close *distance* to the values of the same metric for a *representative*

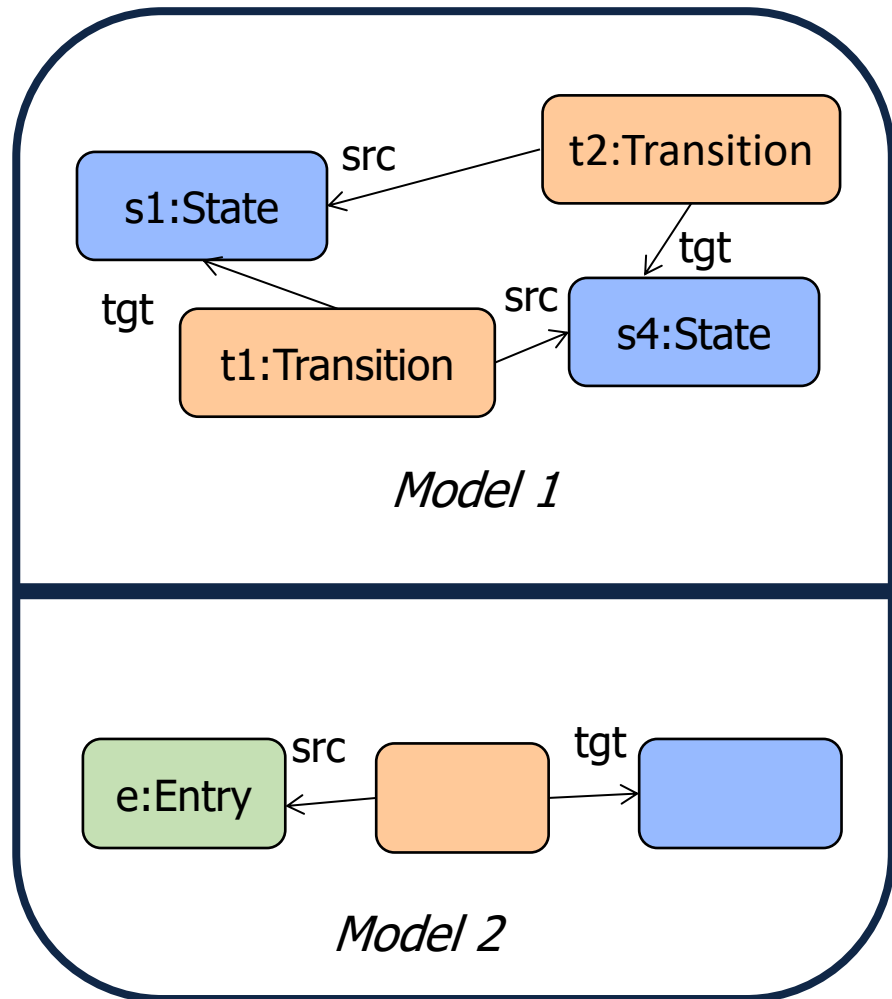


Realistic?



Measure of Structural Realisticness

New definition: A model is considered structurally realistic if the values of **desired graph metrics** are at close distance to the values of the same metric for a representative



Graph metrics: a graph level number or a distribution of node / edge level metric
E.g., *Node Type Distribution (NTD)*

$NTD(Model1) = [State: 2, Transition: 2, Region: 0, Entry: 0]$



Distance: measuring the distances between two metric distributions
E.g., *Manhattan (L1) Distance*

$Manhattan(NTD(Model1), NTD(Model2)) = 3$

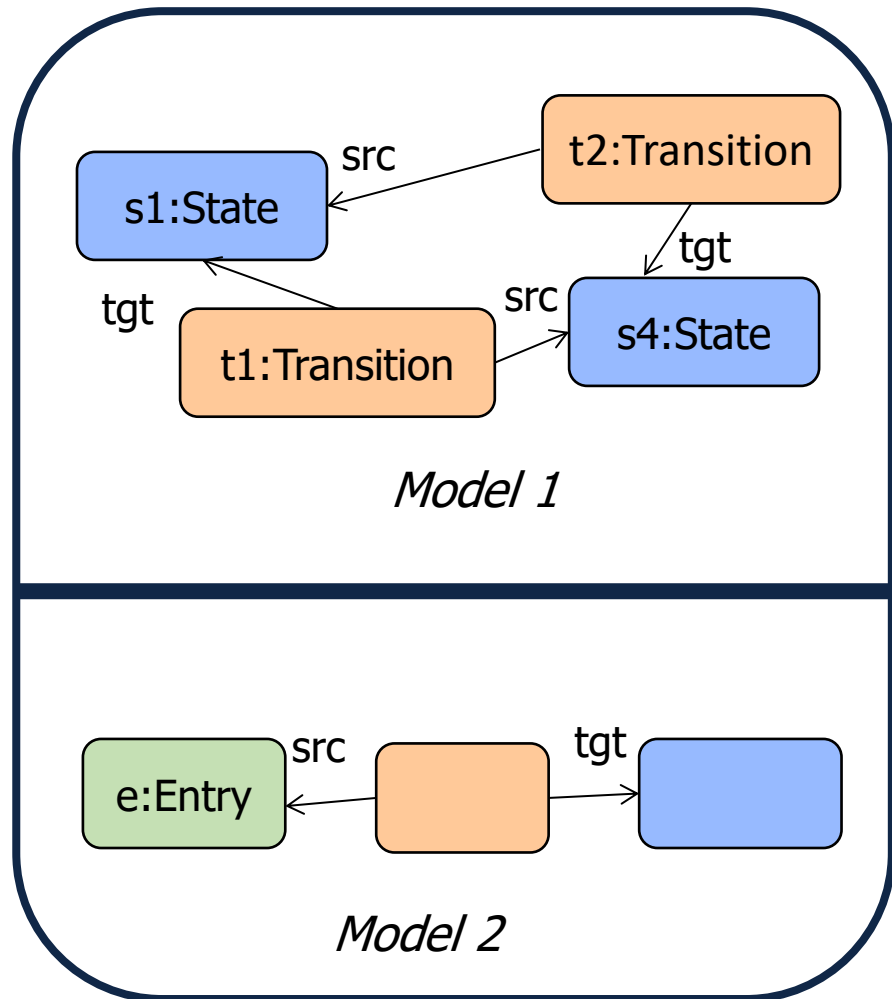


Representative: A metric value that can represent the properties of the real models.

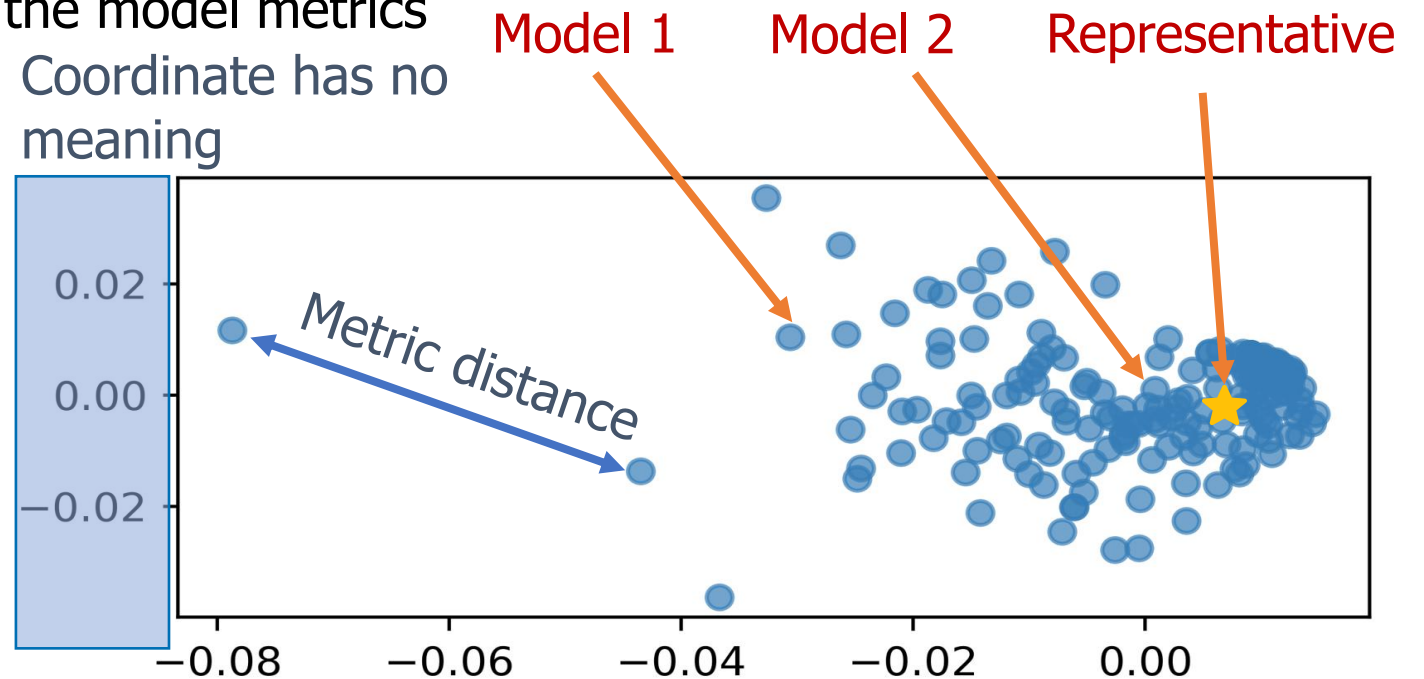
- *Provided by Experts*
- *Select the centroid of the metrics cluster on a set of human models*
- *Select by any other strategies*

Measure of Structural Realisticness

New definition: A model is considered structurally realistic if the values of **desired *graph metrics*** are at close *distance* to the values of the same metric for a *representative*



Metric distances of a set of models plotted by MDS (Multi-Dimensional Scaling). Distance represent the closeness of the model metrics
Coordinate has no meaning



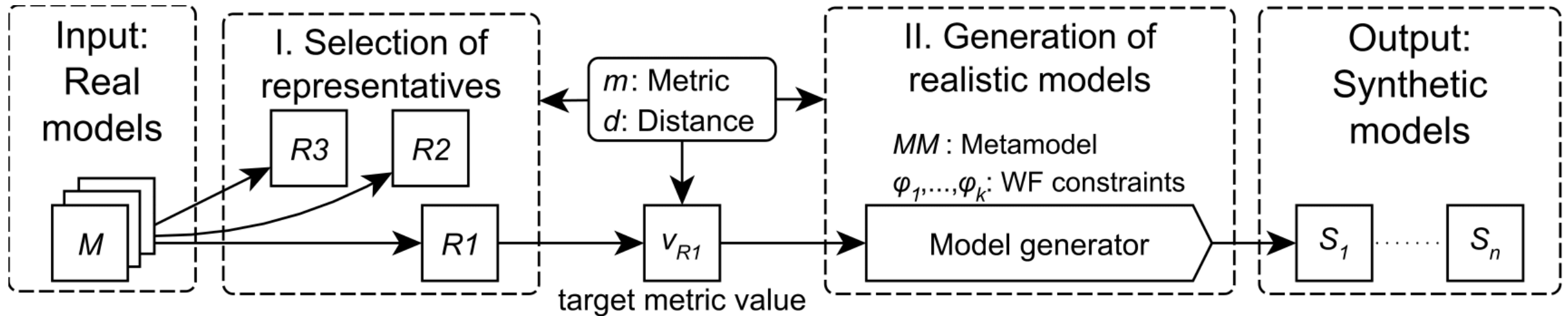
With the new definition, we can conclude Model 2 is more realistic than Model 1

Approach Overview

Objectives: Generate graph models that are close to the human representatives w.r.t. different metrics, while ensuring consistency and diversity

Approach Overview

Objectives: Generate graph models that are close to the human representatives w.r.t. different metrics, while ensuring consistency and diversity



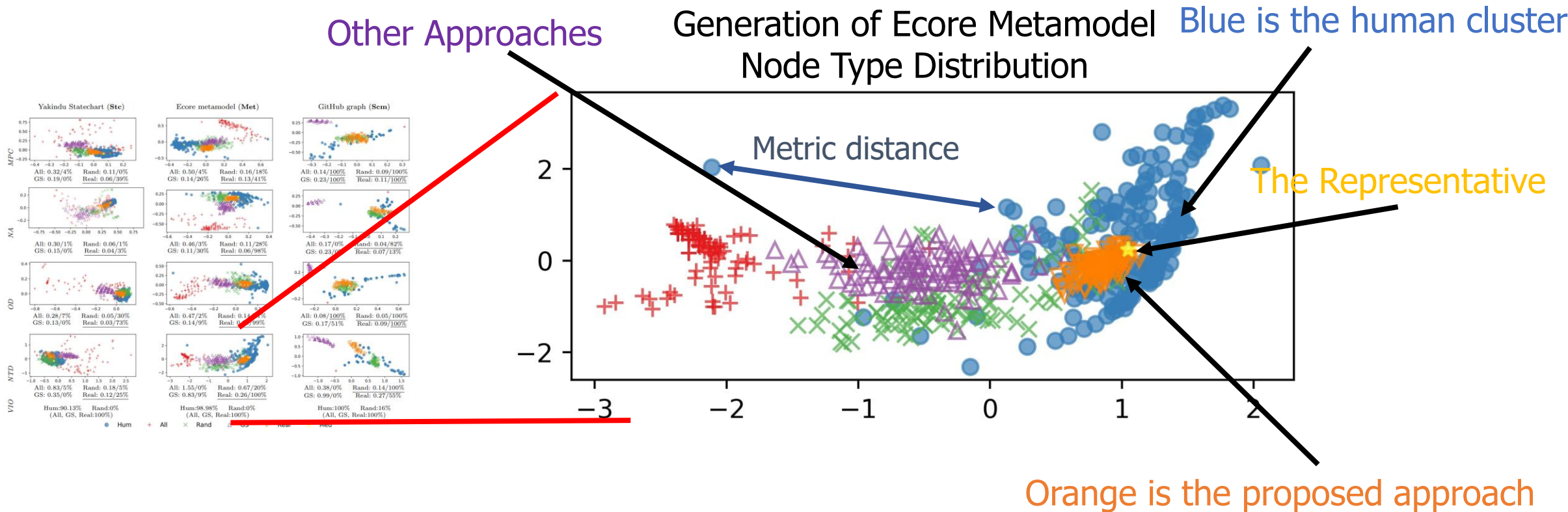
- Selection of representatives: can use any method suggested by experts
- Search strategy: *hill-climbing* to select the action leading to the best next state
- Objective function: linear combination of metrics
- State coding (*graph isomorphism*): partial model with neighborhood shapes

Selection of Graph Generator

- Random Model Generators: *RandomEMF*
- Consistent Model Generators: ***VIATRA Solver, Alloy***
- Diverse Model Generators: ***VIATRA Solver***

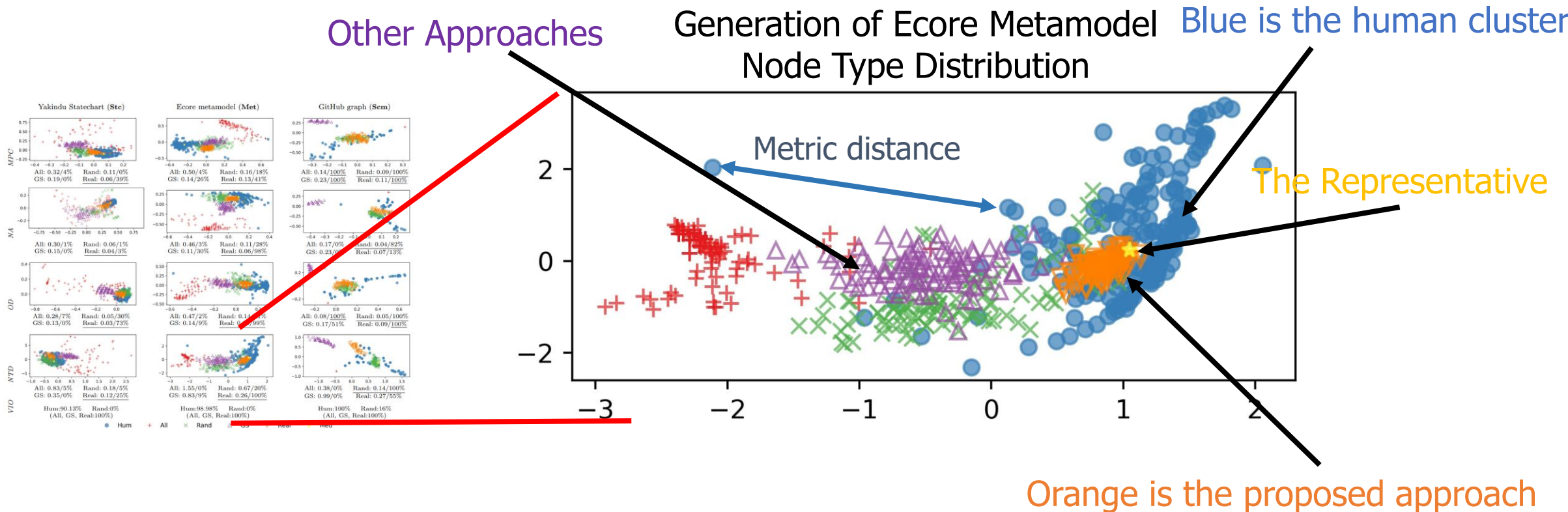
We choose *VIATRA Solver* as the *Model Generator* component of our approach as it is able to generate both **consistent** and **diverse** graph models

Structure Realisticness Compared to Existing Generators



The proposed approach is closer to the representative and the human cluster than other approaches

Structure Realisticness Compared to Existing Generators



Concerning the *average minimal distance* to the **human cluster**, the **proposed approach** provide more realistic models in most setups of case studies in the paper.

Conclusion



Consistent



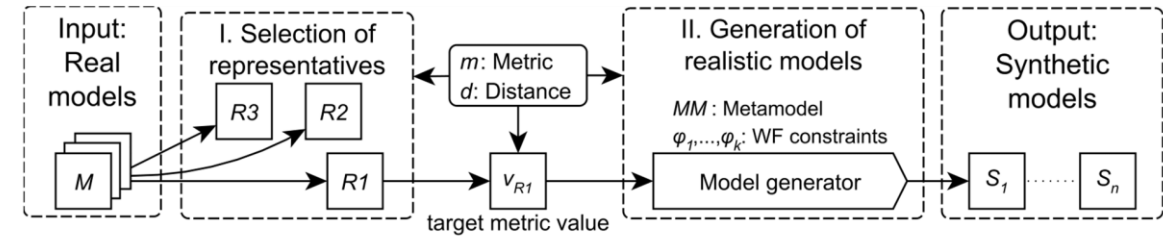
Diverse



Realistic

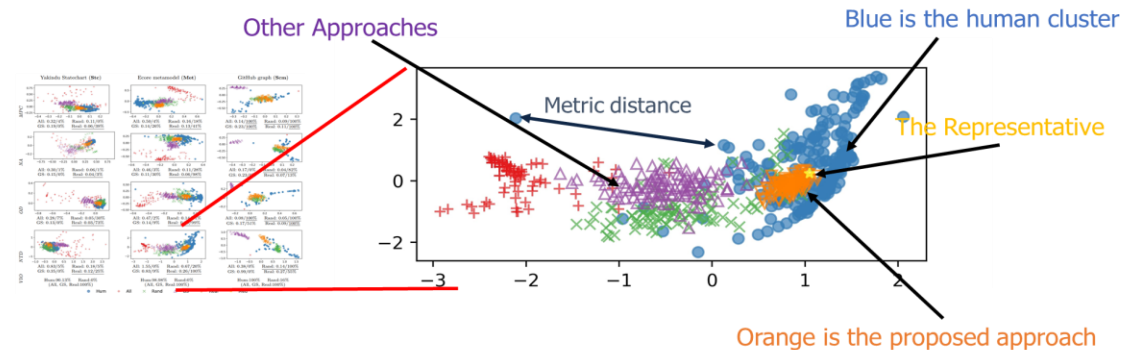
Approach Overview

Objectives: Generate graph models that are close to the human representatives w.r.t. different metrics, while ensuring consistency and diversity



- Selection of representatives: can use any method suggested by experts
- Search strategy: *hill-climbing* to select the action leading to the best next state
- Objective function: linear combination of metrics
- State coding (*graph isomorphism*): partial model with neighborhood shapes

Structure Realisticness Compared to Existing Generators



Concerning the *average minimal distance*, the proposed approach (**Real**) provide more realistic in many cases of the case studies in the paper.