



Meta Structure: Computing Relevance in Large Heterogeneous Information Networks

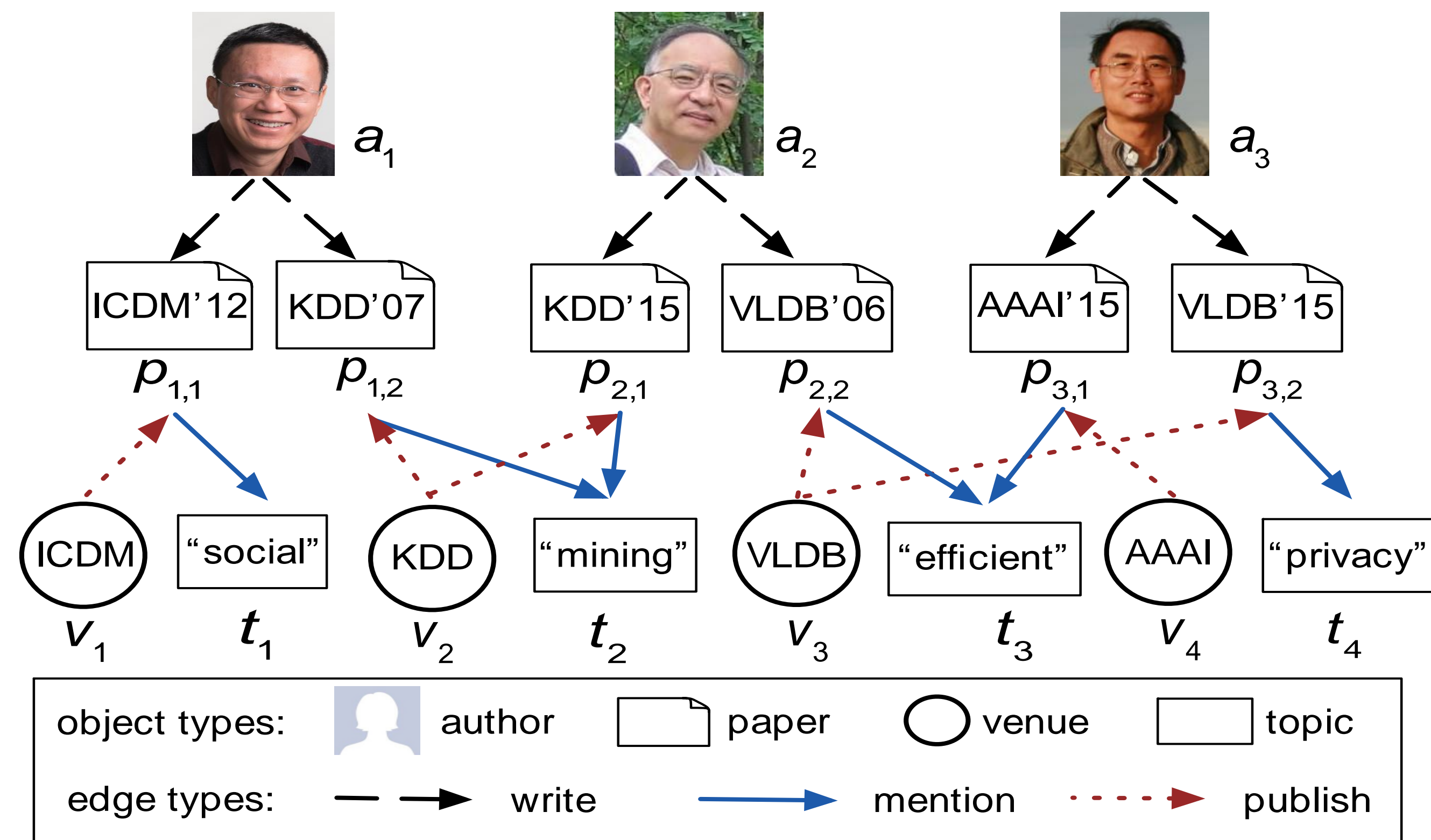


Zhipeng Huang*, Yudian Zheng*, Reynold Cheng*, Yizhou Sun#, Nikos Mamoulis*, Xiang Li*

*The University of Hong Kong, #Northeastern University

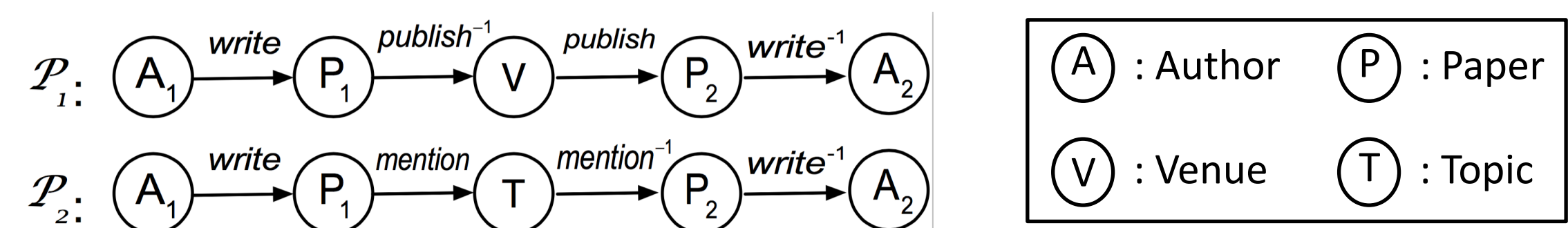
Heterogeneous Information Networks

HIN: directed graph with multiple node types and edge types.

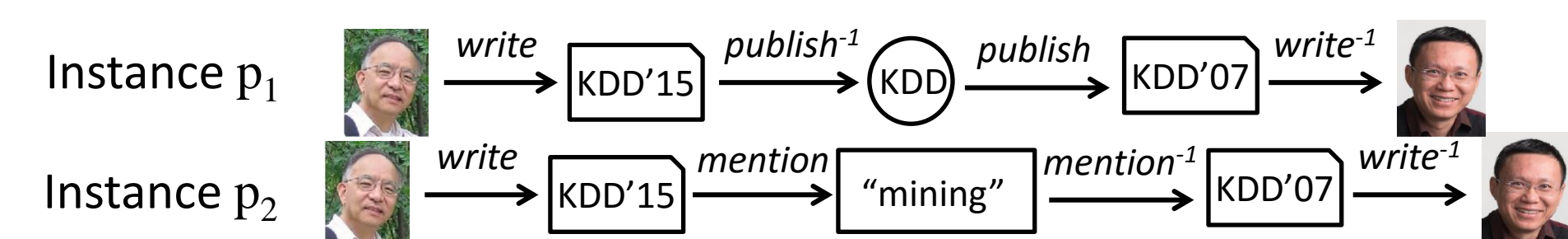


Relevance Computing on HIN

Meta Path [1]: a sequence of node types and edge types.



Meta Path Instance: a path of HIN conforming to the pattern.



Relevance Measures:

PathCount [1]: the number of meta path instances.

PathSim [1]: a normalized version of PathCount.

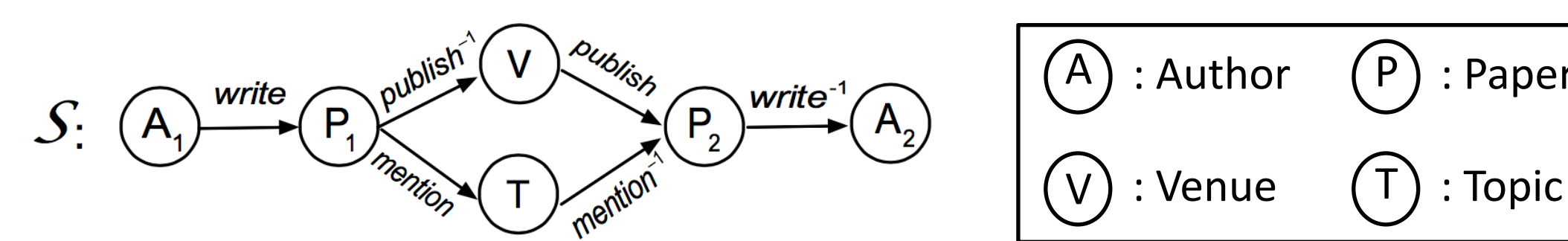
PCRW [2]: the probability of the random walk.

Our Main Contribution

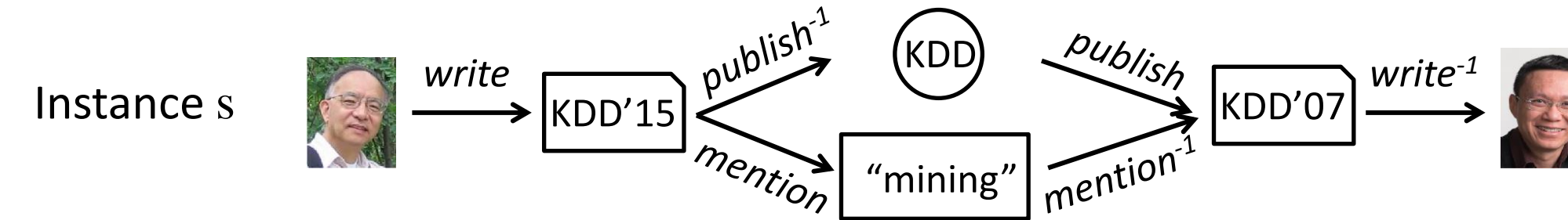
Meta Path \rightarrow **Meta Structure**

Meta Structure: an extension of meta path.

More Expressive!



Meta Structure Instance: a subgraph of HIN.



Advantages:

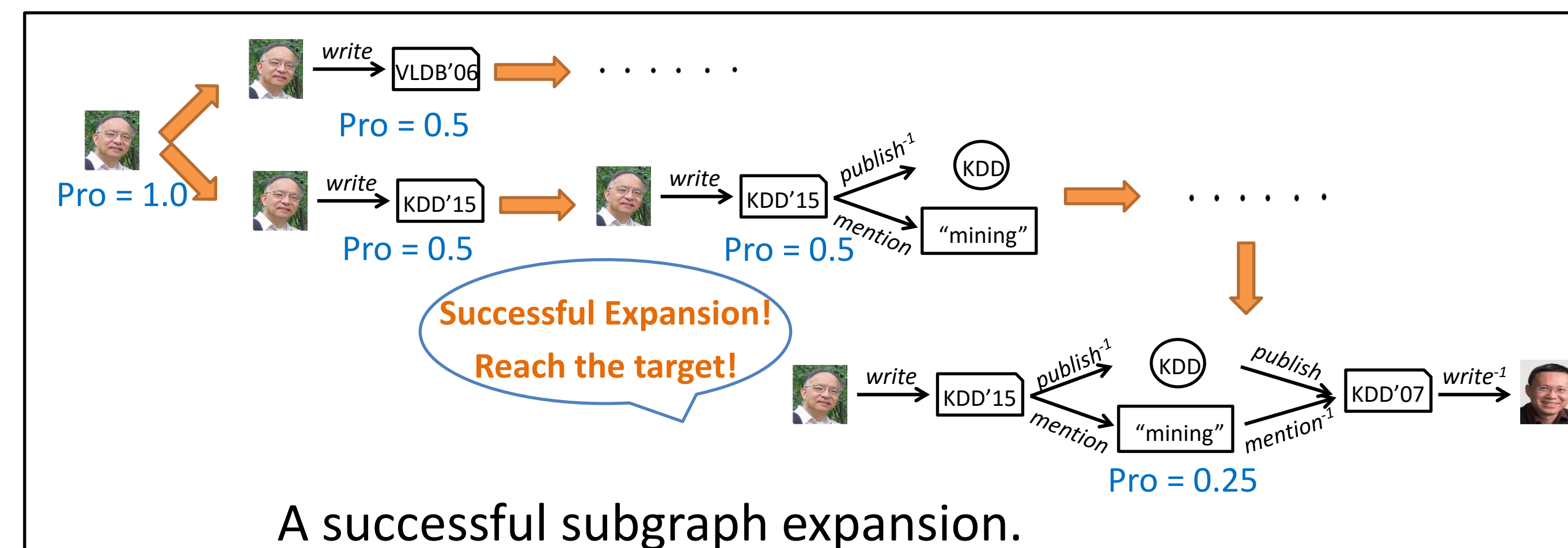
- More expressive and flexible than meta path.
- Support more complex relationships.

Structure-based Relevance

StructCount: no. of meta structure instances.

SCSE: prob. of a "successful" subgraph expansion.

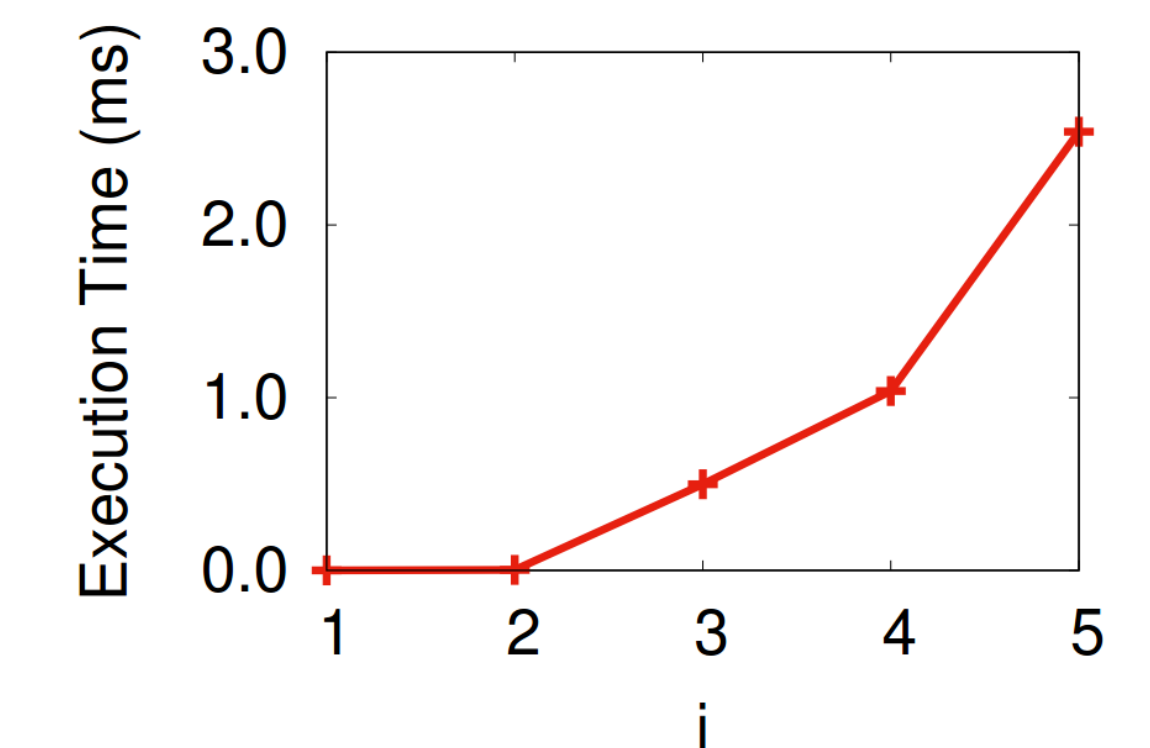
BSCSE: StructCount + SCSE.



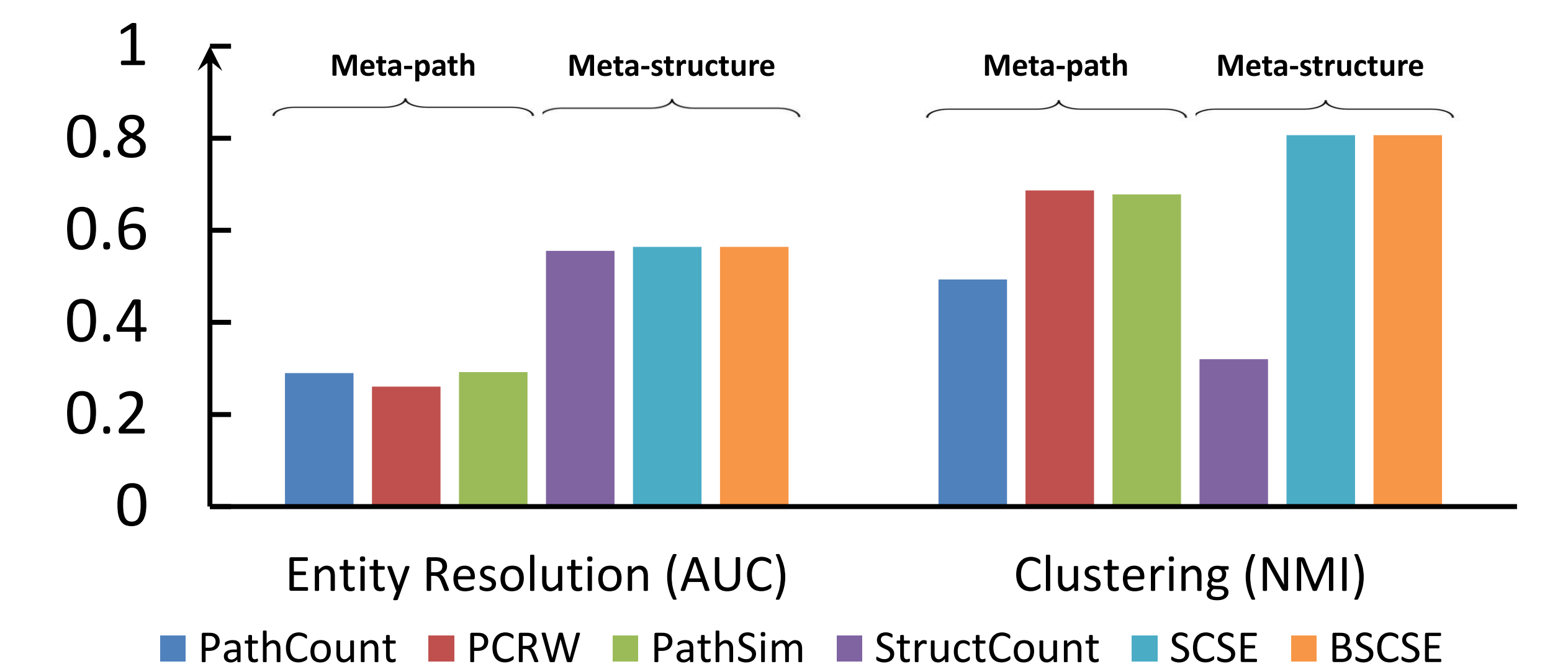
Efficiency

i-Table: a lookup table for storing the search results of a given meta structure at the i -th layer.

key	Value
<ICDM, social>	<Pei, 1.0>
<KDD, mining>	<Pei, 0.5>
<VLDB, efficient>	<Han, 1.0>
<VLDB, privacy>	<Yang, 1.0>
<AAAI, efficient>	<Yang, 1.0>



Effectiveness



Conclusions

- We propose the meta structure, which expresses complex relations between two objects in HINs.
- We design 3 relevance measures based on meta structures, which are more effective than those based on meta paths.

References

- Yizhou Sun, et al. "Pathsim: Meta path-based top-k similarity search in heterogeneous information networks." VLDB'11 (2011).
- Changping Meng, et al. "Discovering Meta-Paths in Large Heterogeneous Information Networks" WWW'15 (2015).