

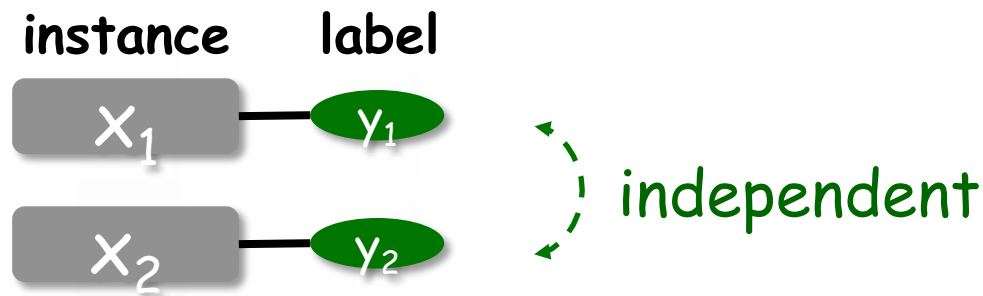
# Meta-path-based Collective Classification in Heterogeneous Information Networks

**Xiangnan Kong, Philip S. Yu, Ying Ding, David J. Wild**

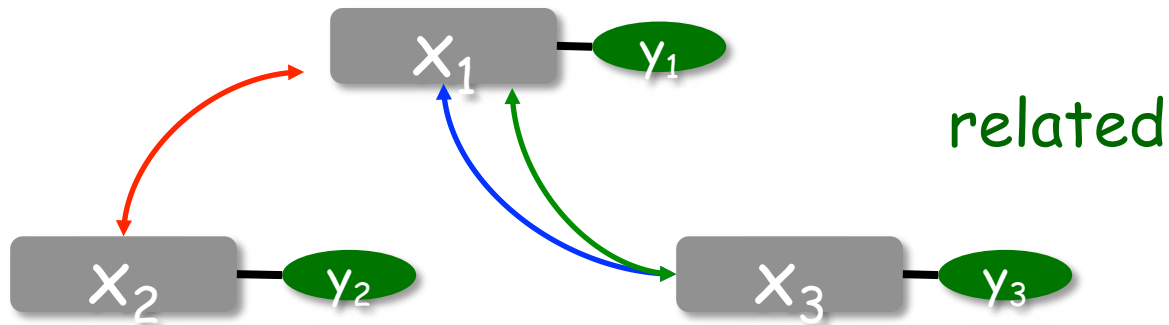
**University of Illinois at Chicago  
Indiana University of Bloomington**

# Collective Classification

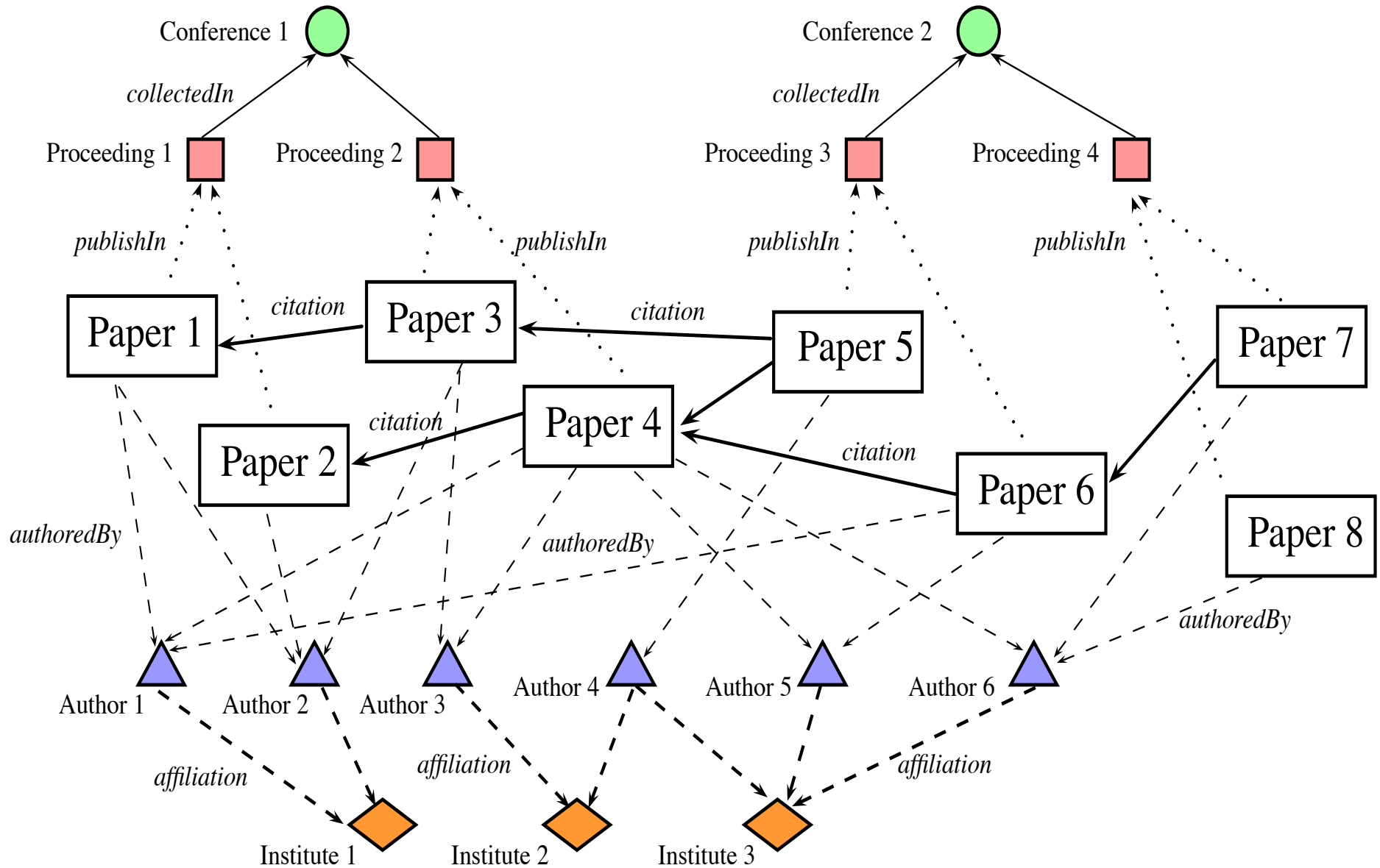
- Conventional classification approaches assume that instances are **independent** identically distributed (**i.i.d.**)



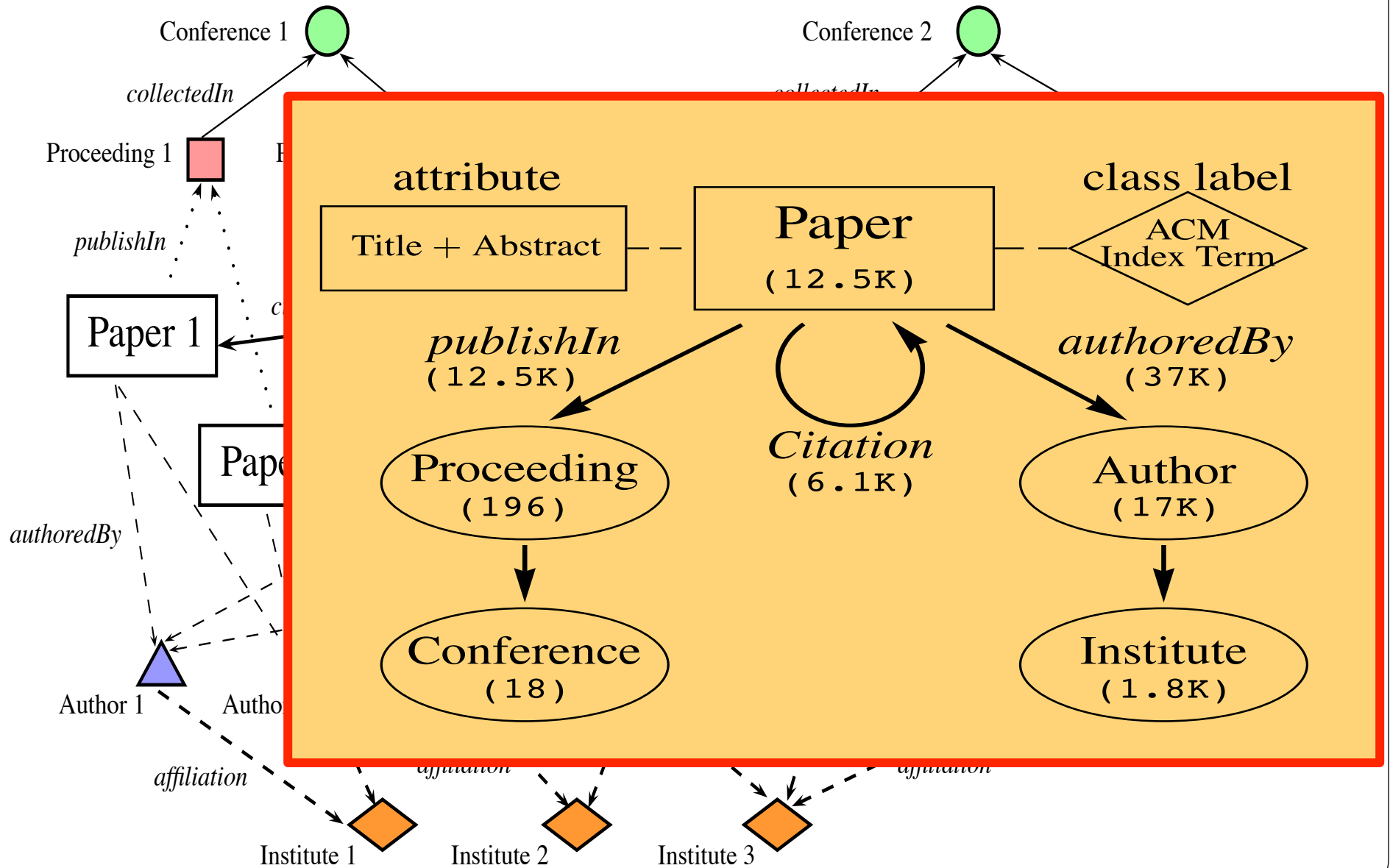
- In network data, instances are interconnected through links, thus are **correlated** with each other.



# Heterogeneous Information Networks



# Heterogeneous Information Networks



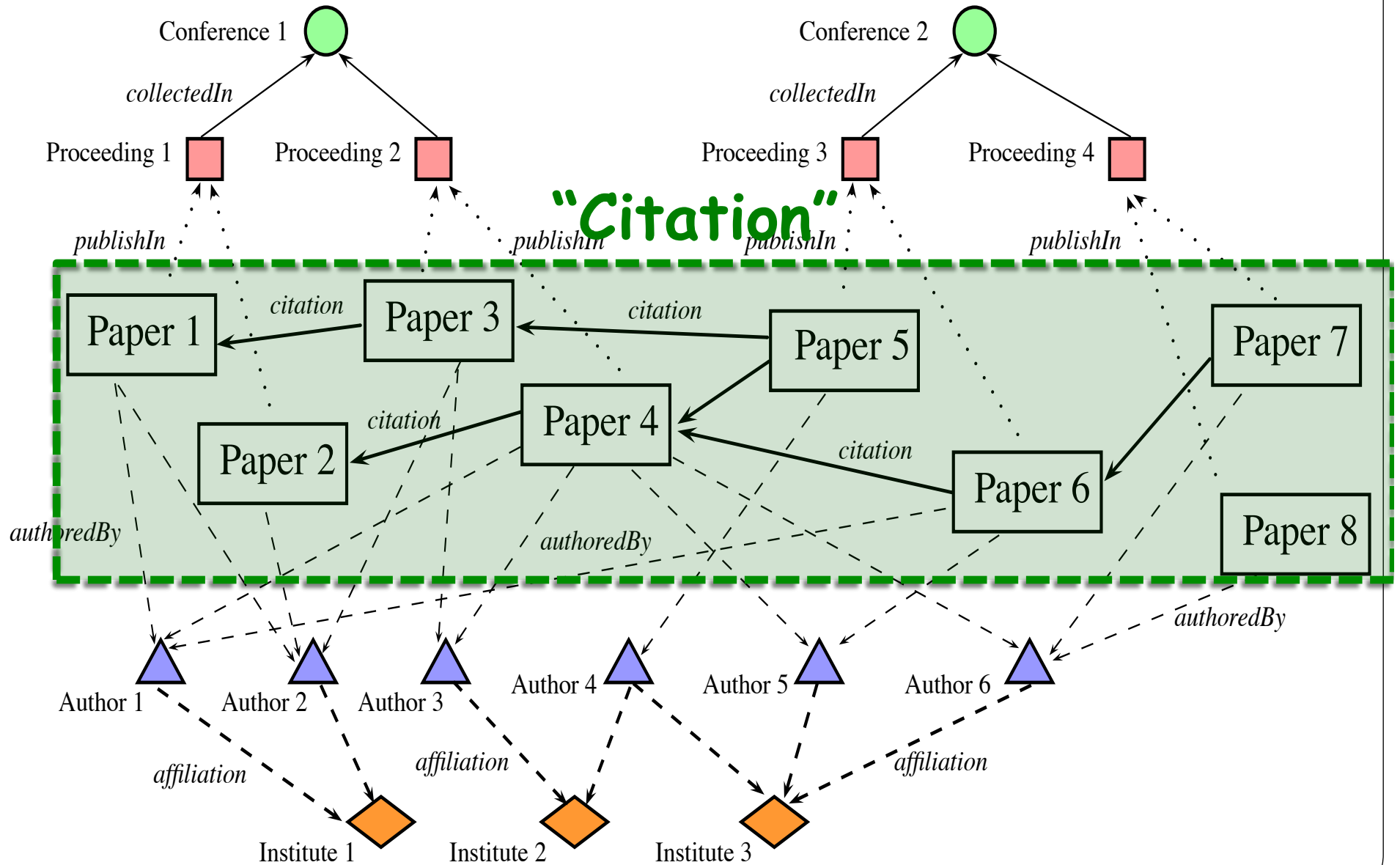
# The Problem

- **Multiple types of dependencies:**  
different types of dependencies have different semantic meanings

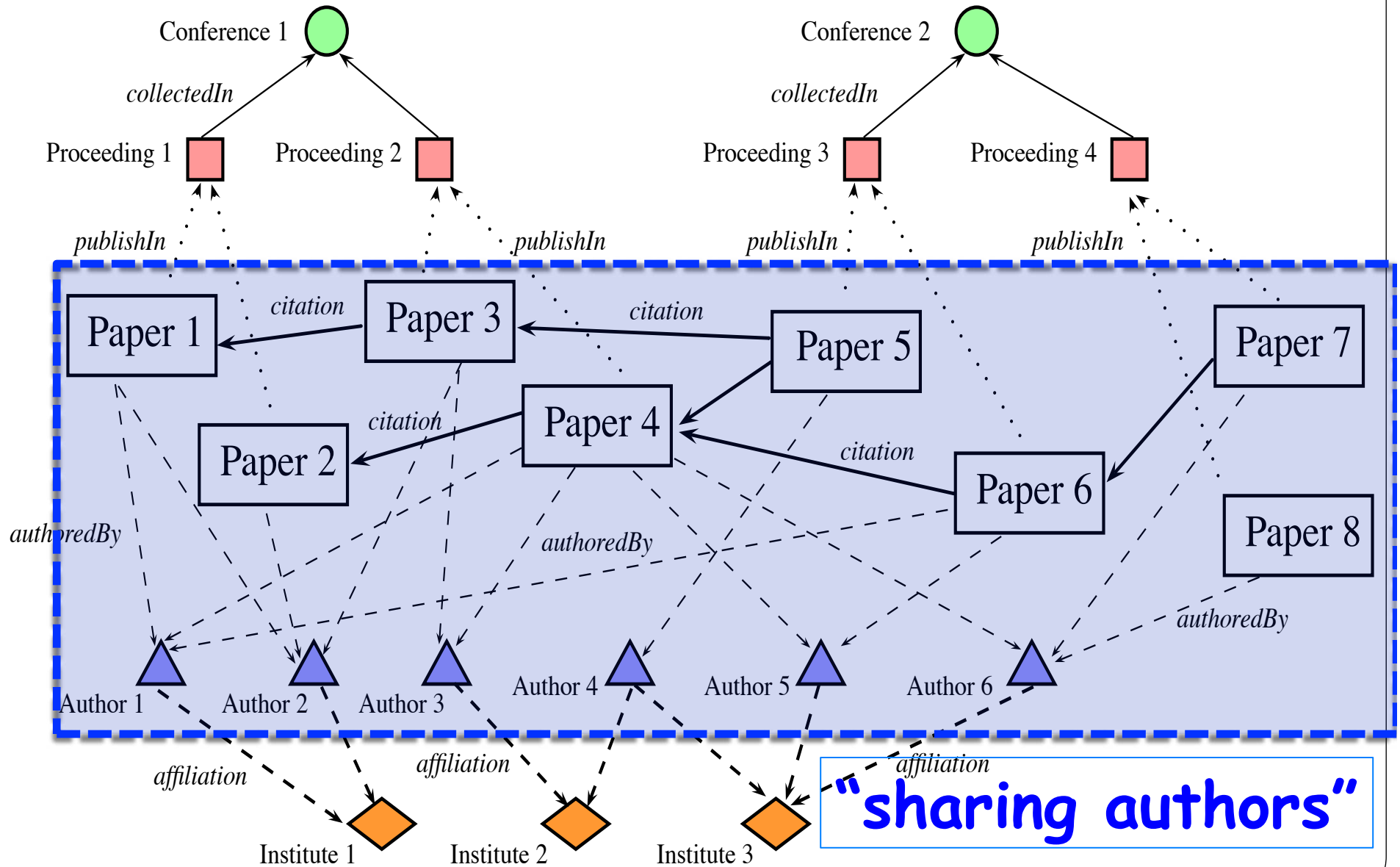
The key is to exploit the **heterogeneous correlations** among different instances

- **Collective classification**

# Heterogeneous Information Networks



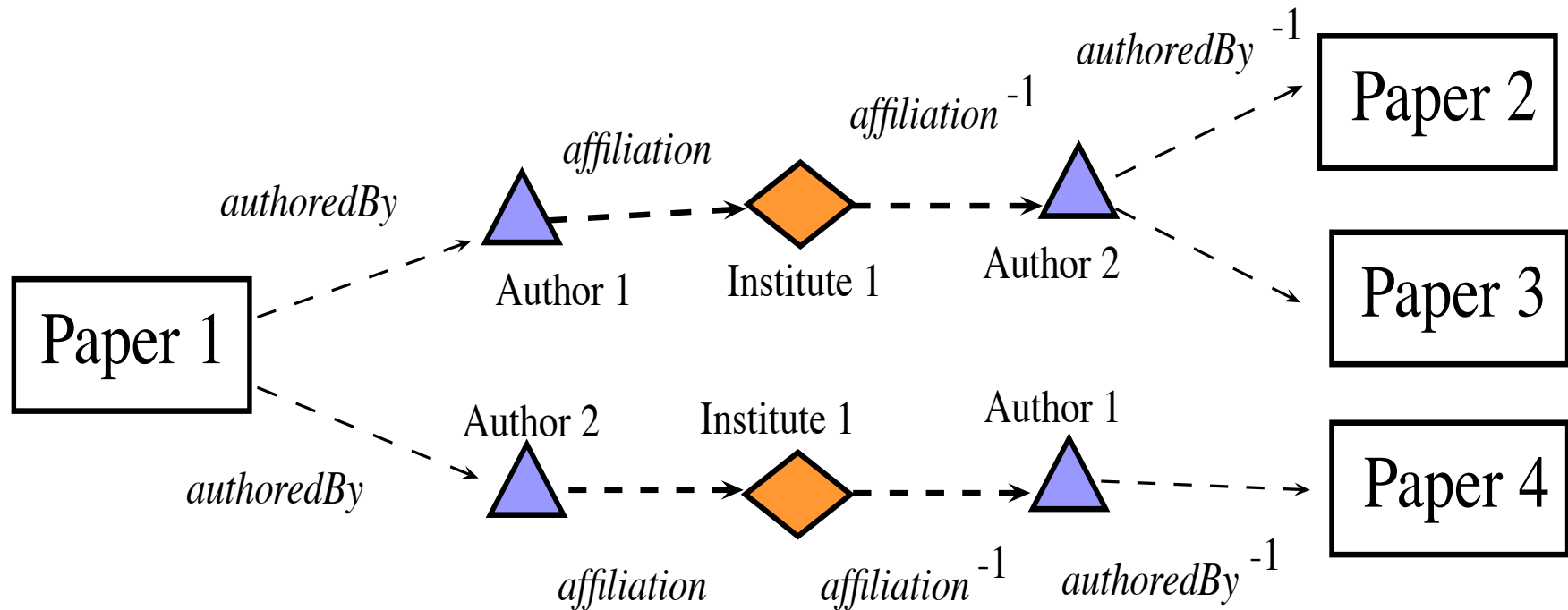
# Heterogeneous Information Networks



# Example III

“Same School”

P - A - I - A - P

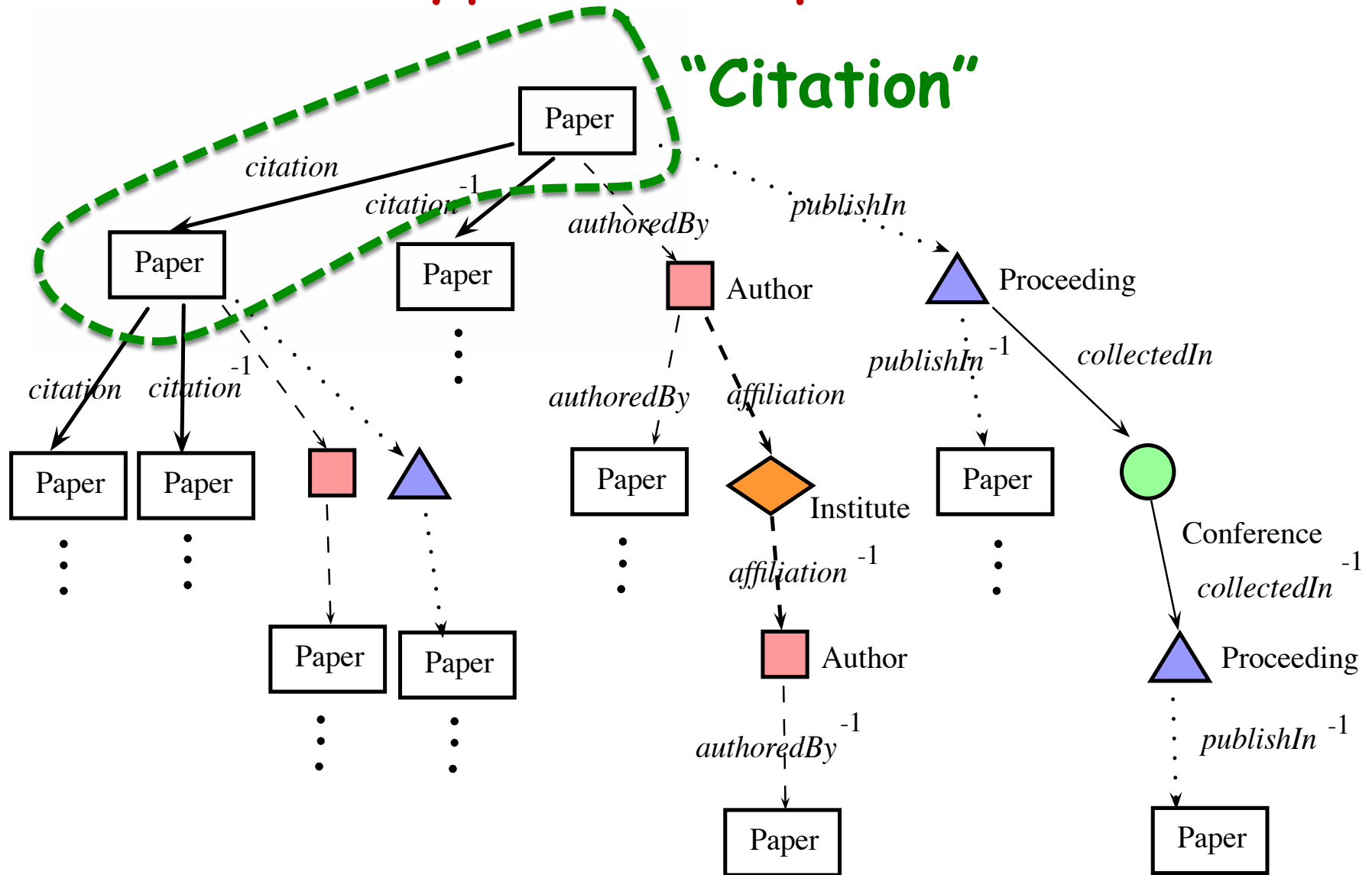




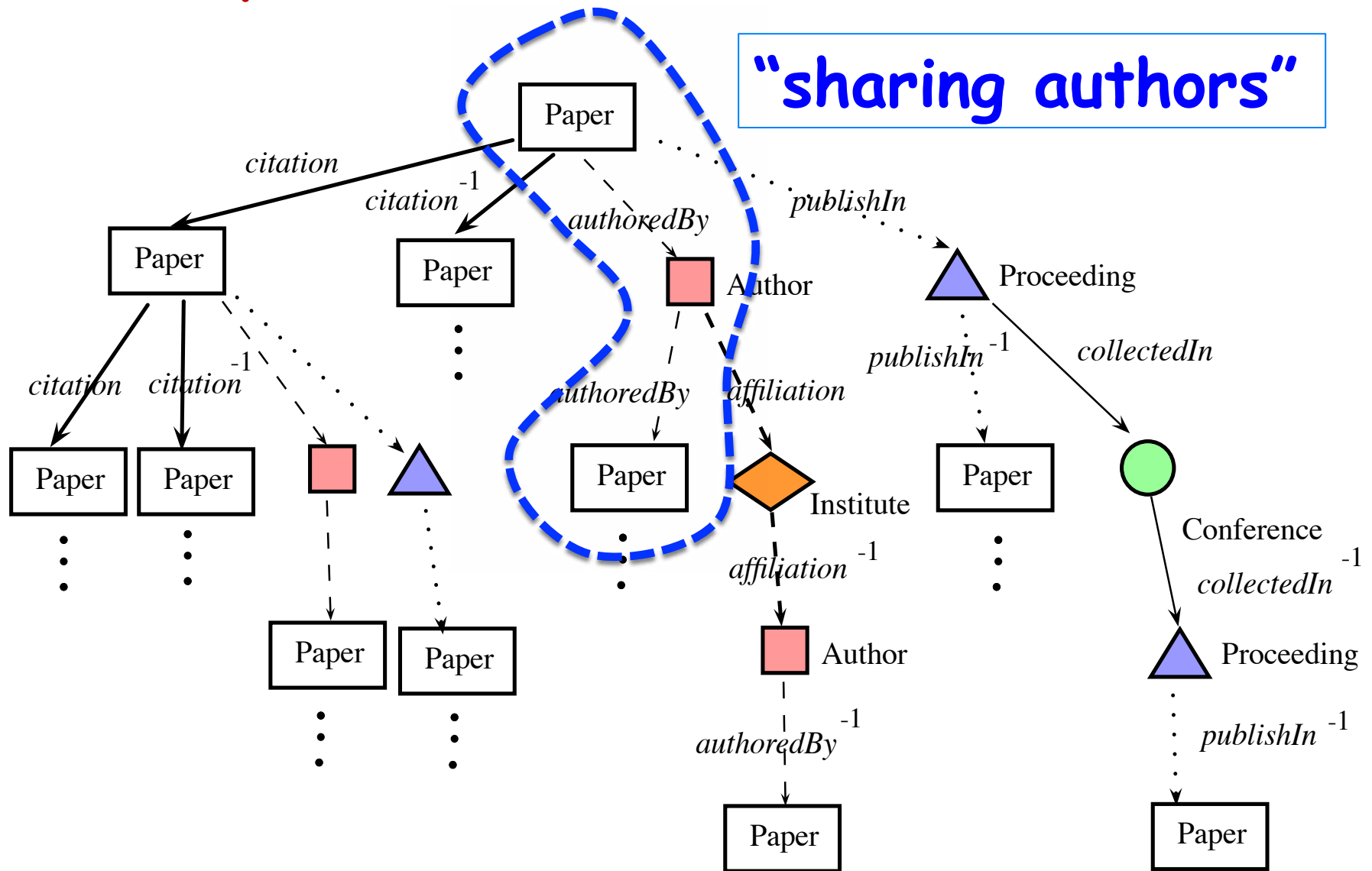
# Different Types of Dependencies

| Notation | Meta Path   | Semantics of the Dependency    |
|----------|---|--------------------------------|
| 1 P→P    | Paper $\xrightarrow{\text{cite}}$ Paper   | Citation                       |
| 2 P←P→P  | Paper $\xrightarrow{\text{cite}^{-1}}$ Paper $\xrightarrow{\text{cite}}$ Paper  | Co-citation                    |
| 3 P→P←P  | Paper $\xrightarrow{\text{cite}}$ Paper $\xrightarrow{\text{cite}^{-1}}$ Paper  | Bibliographic coupling         |
| 4 PVP    | Paper $\xrightarrow{\text{publishIn}}$ Proceeding $\xrightarrow{\text{publishIn}^{-1}}$ Paper   | Papers in the same proceeding  |
| 5 PVCVP  | Paper $\xrightarrow{\text{publishIn}}$ Proceeding $\xrightarrow{\text{collectIn}}$ Conference<br>$\xrightarrow{\text{collectIn}^{-1}}$ Proceeding $\xrightarrow{\text{publishIn}^{-1}}$ Paper | Papers in the same conference  |
| 6 PAP    | Paper $\xrightarrow{\text{write}^{-1}}$ Author $\xrightarrow{\text{write}}$ Paper   | Papers sharing authors         |
| 7 PAFAP  | Paper $\xrightarrow{\text{write}^{-1}}$ Author $\xrightarrow{\text{affiliation}}$ Institute<br>$\xrightarrow{\text{affiliation}^{-1}}$ Author $\xrightarrow{\text{write}}$ Paper              | Papers from the same institute |

# Different types of Dependencies

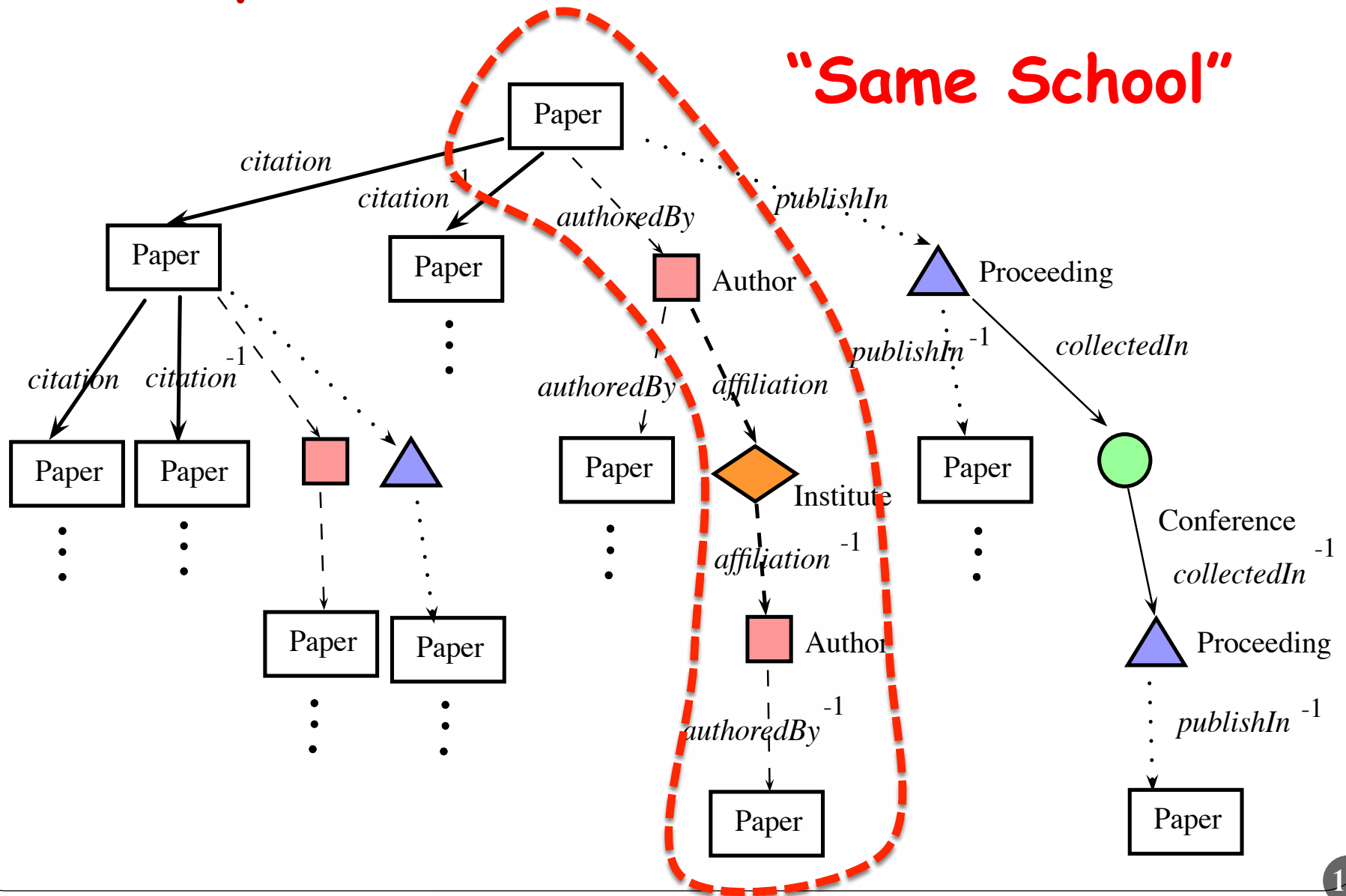


# Example II

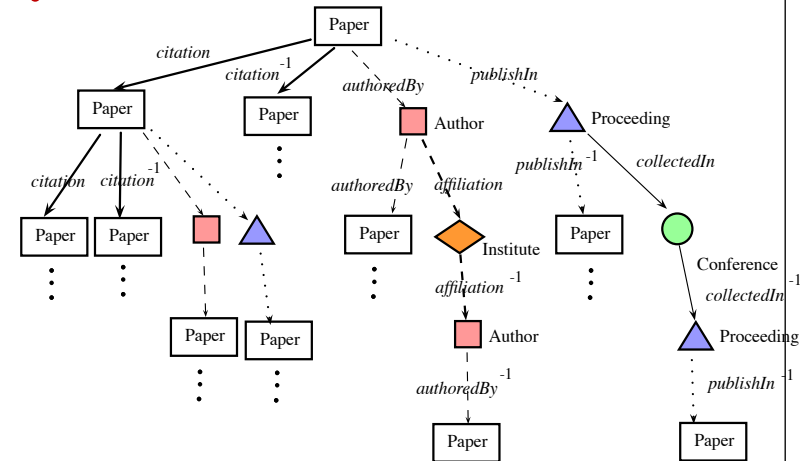
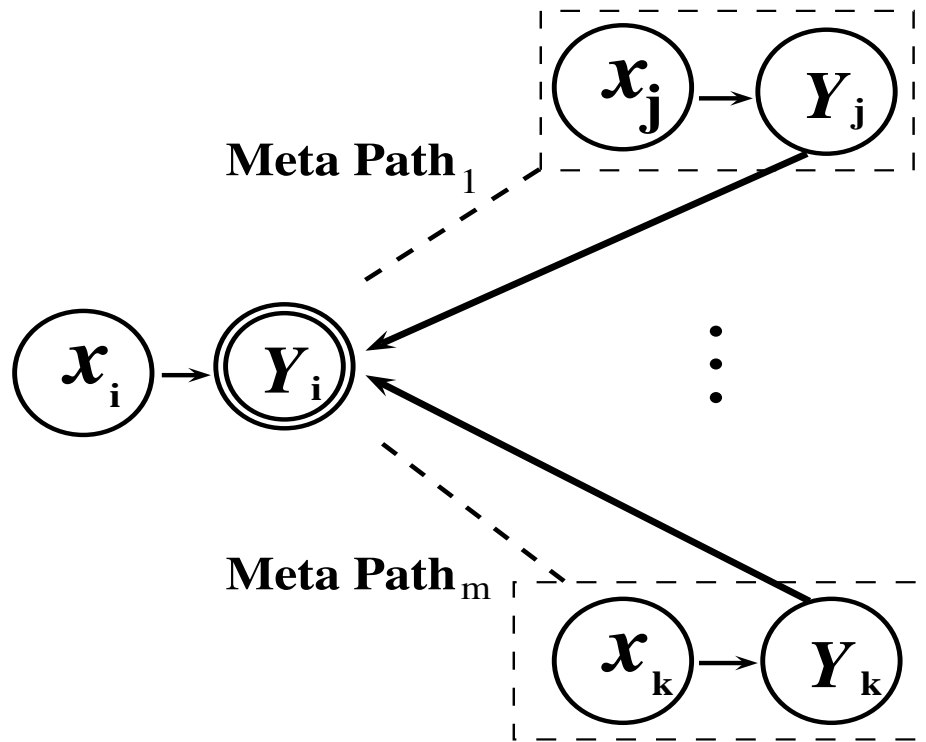


# Example III

"Same School"



# Meta-path-based Dependencies



$$Pr(\mathcal{Y}|\mathcal{X}) = \prod_{i \in U} Pr(Y_i | x_i, \mathbf{Y}_{P_1(i)}, \mathbf{Y}_{P_2(i)}, \dots, \mathbf{Y}_{P_m(i)})$$

# Experiments: Compared Methods

## Independent classification

**SVM**

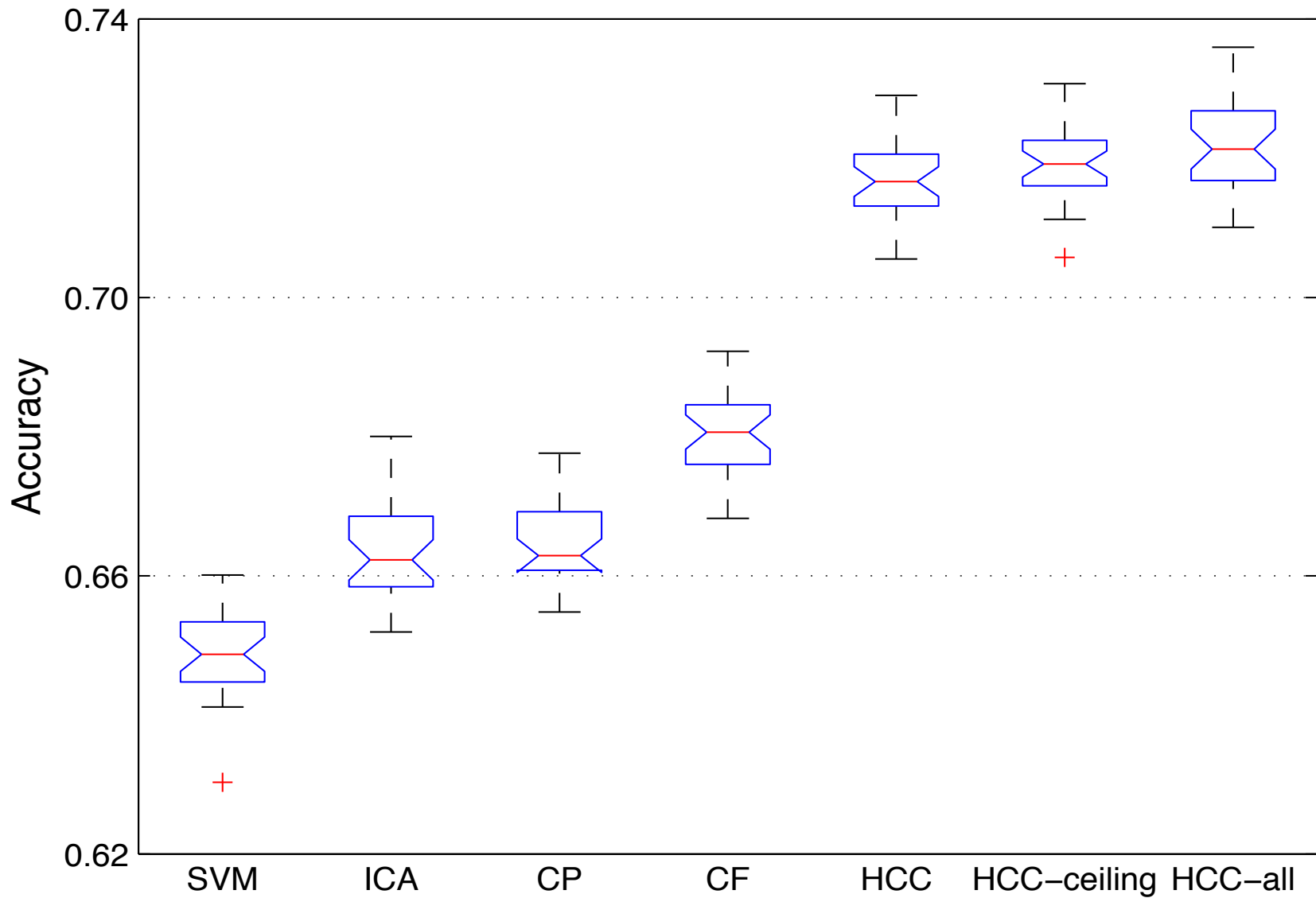
## Collective classification

|            |                                    |                              |
|------------|------------------------------------|------------------------------|
| <b>ICA</b> | iterative classification algorithm | [Lu&Getoor, ICML'03]         |
| <b>CP</b>  | Combined Path Relations            |                              |
| <b>CF</b>  | Ensemble of path relations         | [Eldardiry&Neville, AAAI'11] |

## Meta-path-based collective classification

|                    |                       |              |
|--------------------|-----------------------|--------------|
| <b>HCC</b>         | the proposed approach | [this paper] |
| <b>HCC-ceiling</b> | a ceiling analysis    | [this paper] |

# Experiment Results, ACM dataset



# Conclusions



- **Meta-path-based Collective Classification** in Heterogeneous Information Networks
- Propose an algorithm to exploit the dependencies among heterogeneous dependencies among related instances
  - ❖ Meta-path-based Dependency
- Classification performances can be improved by considering the heterogeneous dependencies among instances.



# Thank you!