

Name Disambiguation in AMiner:

Jing Zhang*, Jie Tang⁺

*Information School, Renmin University

⁺Computer Science Department, Tsinghua University



An Example in AMiner

All the researchers named “Jing Zhang”

Jing Zhang (张靖) 副教授
清华大学自动化系
+86 (10) 6279 7485
jzg_zhang@mails.tsinghua.edu.cn

65 Optomechanically induced stochastic resonance and chaos transfer between optical fields
Faraz Monifi, **Jing Zhang**, Şahin Kaya Ozdemir, Bo Peng, Fang Bo, Lan Yang, Jing Zhang, Jing Zhang
Nature Photonics (2016)
Cited by 66 Bibtext

64 Squeezed optomechanics with phase-matched amplification and dissipation.
Lü Xin-You, **Wu Ying**, Johansson J R, Jing Hui, Zhang Jing, Nori Franco
Physical review letters (2015)
Cited by 104 Bibtext

Jing Zhang (张经) 教授
华东师范大学
http://www.casad.cas.cn/aca/371/dxb-200906-1200906...
External Links
59 views

322 Sources of organic matter in Changjiang (Yangtze River) bed sediments: Preliminary insights from proxies
zhongqiao li, francien peterse, ying wu, hongyan bao, timothy i eglinton, **jing zhang**
Organic Geochemistry (2015)
Cited by 8 Bibtext

321 Association between JAK2 rs4495487 Polymorphism and Risk of Budd-Chiari Syndrome in China
peijin zhang, yanyan zhang, **jing zhang**, hui wang, he ma, wei wang, xiuyin gao, hao xu, zhaojun lu
Gastroenterology Research and Practice (2015)
Bibtext

Miner Data · Knowledge · Intelligence

Jing Zhang

People named “Jing Zhang”

- J. Zhang #-Index: 176
Institute of High Energy Physics
- Jing Zhang #-Index: 38
华东师范大学
- Jing Zhang #-Index: 37
Computer Science Department, Tsinghua University.
- Jing Zhang #-Index: 34
.Department of Pathophysiology, Hebei North University, Zhangjiakou, China
- Jing Zhang #-Index: 32
College of Animal Science and Veterinary, Jilin University, Changchun, China
- Victor S. Sheng #-Index: 22
Computer Science Department University of Central Arkansas
- Jing Zhang #-Index: 21
Northeast Agricultural University, Harbin Heilongjiang

Jing Zhang (张璟) 教授
西安理工大学计算机学院
029-82312231

120 社会资本对城镇减贫的影响——基于CGSS微观数据城镇样本的实证分析
王志刚, **张璟**, 蒋亚敏
Journal of Zhengzhou Institute of Aeronautical Industry Management (2013)
Bibtext

119 基于BIC与PSO的简约语音识别系统创建
包希日莫, 高光来, **张璟**
Computer Engineering and Applications (2013)
Bibtext

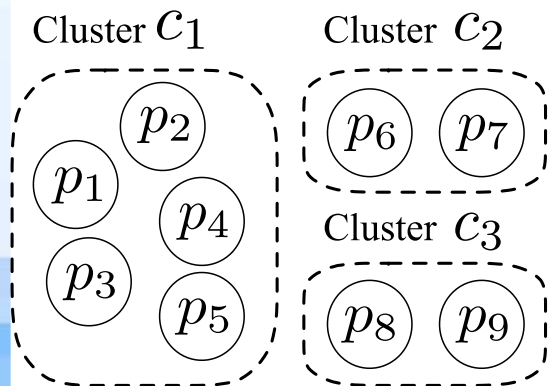
Jing Zhang (张静) Assistant Professor
Renmin University
zhang.jing@ruc.edu.cn
http://keg.cs.tsinghua.edu.cn/persons/zhangjing/
External Links

24 Fast and Flexible Top-k Similarity Search on Large Networks
Jing Zhang, Jie Tang, Cong Ma, Hanghang Tong, Yu Jing, Juanzi Li, Walter Luyten, Marie-Francine Moens
ACM Trans. Inf. Syst. (2017)
Cited by 5 Bibtext

23 Panther: Fast Top-k Similarity Search in Large Networks.
Jing Zhang, Jie Tang, Cong Ma, Hanghang Tong, Yu Jing, Juanzi Li
KDD (2015)
Cited by 22 Bibtext

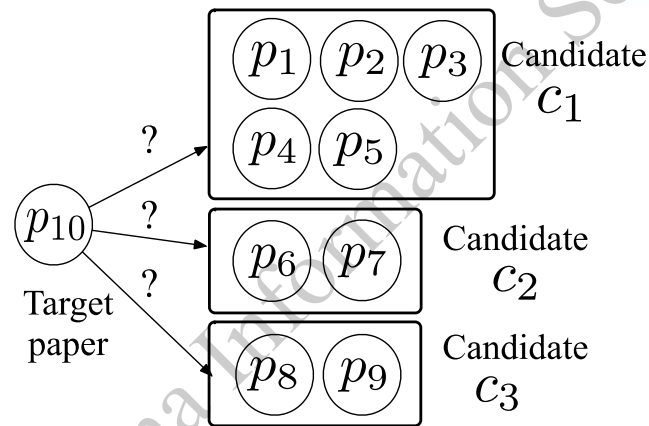
Assign papers with author name “Jing Zhang” to right authors

Three Scenarios



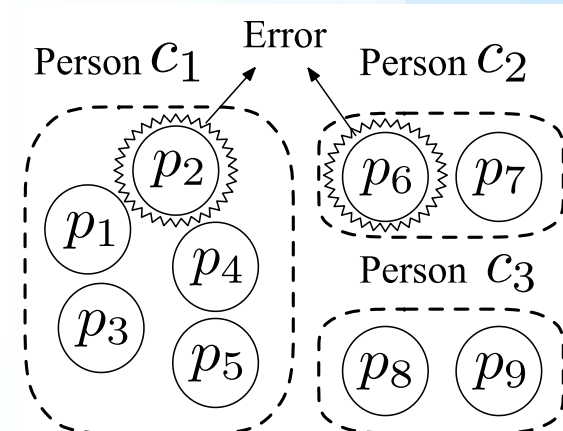
Full ND

- Name disambiguation when the system is built from scratch



Continuous ND

- Name disambiguation when persons' profiles are continuously updated

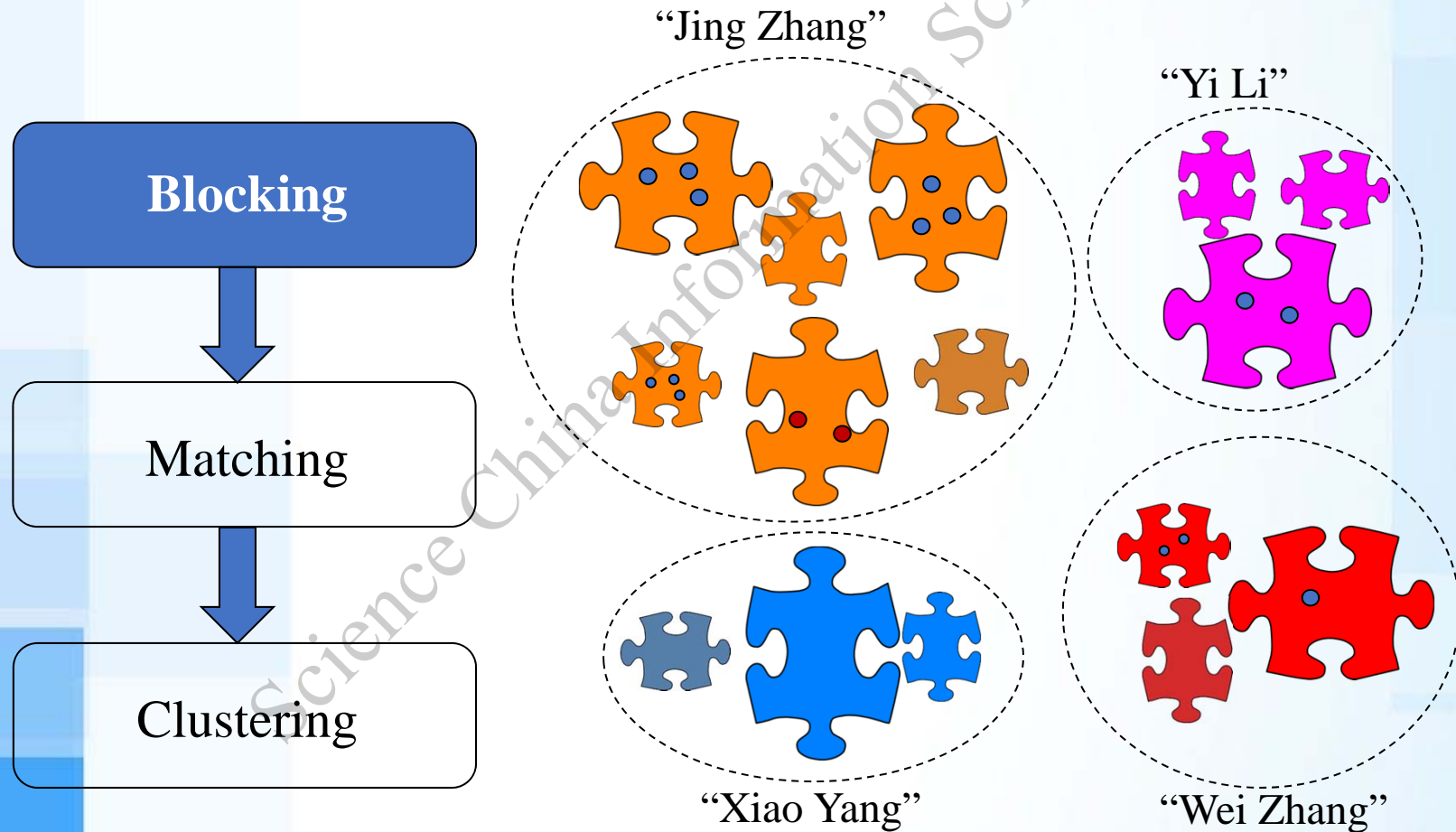


Error

- Error detection upon existing persons' profiles

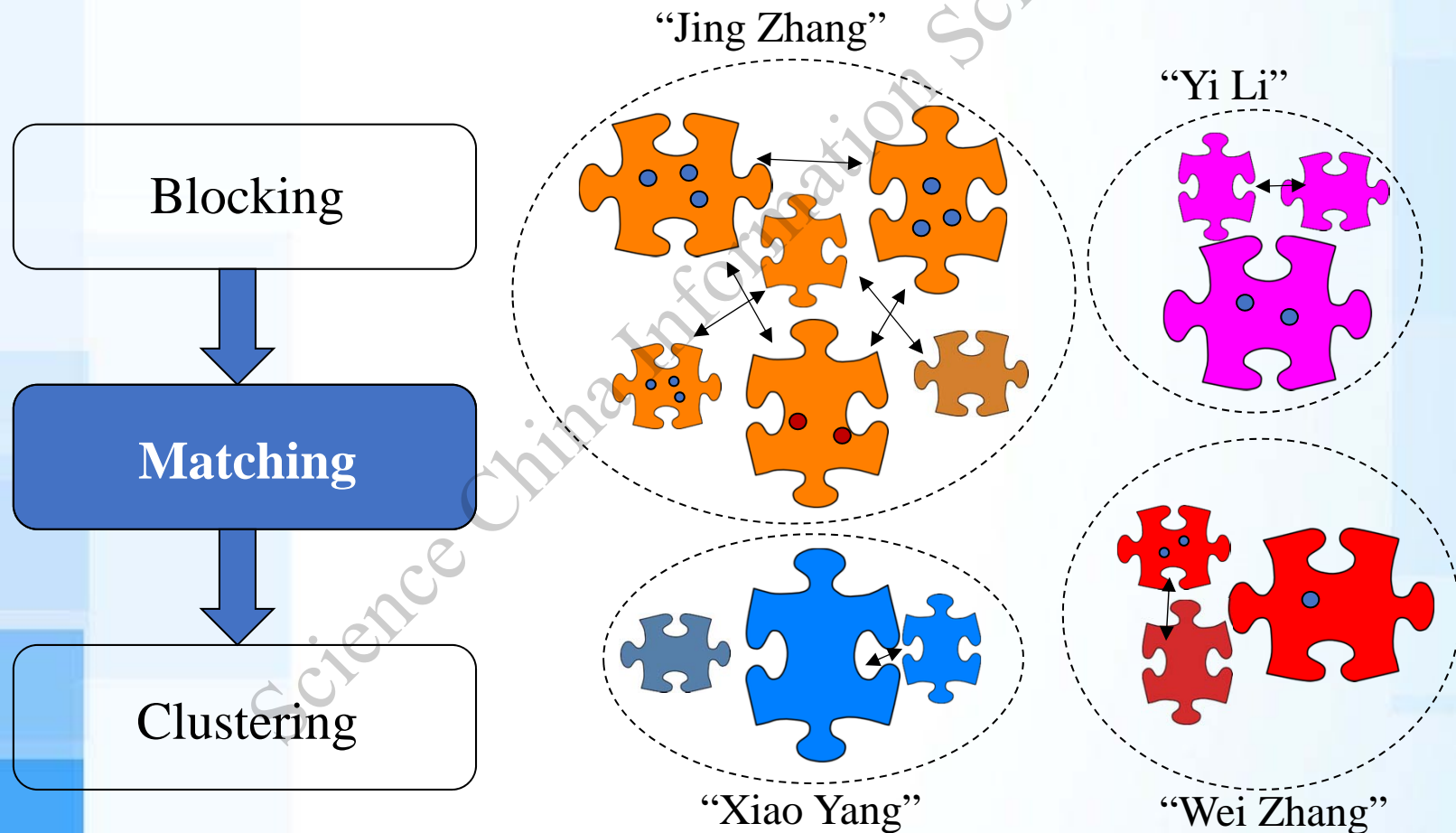
1st Scenario: Full ND

- Stage One: Constructing candidate sets of potential matchings.



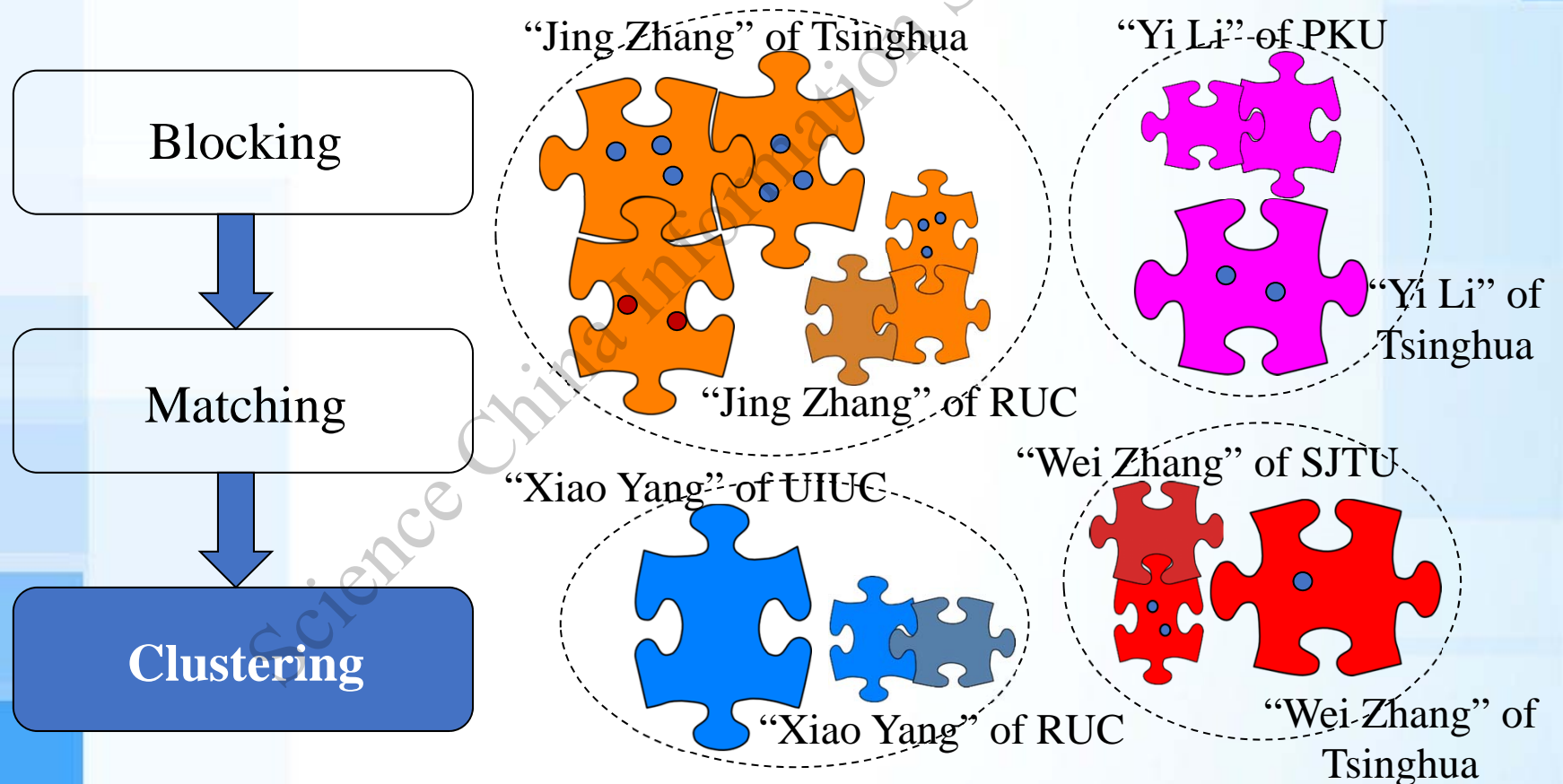
Stage Two: Similarity Matching

- Constructing similarity matrix for each candidate set.



Stage Three: Clustering

- Partitioning each candidate set based on similarity.



Feature-based Matching

TABLE 1
Attributes of Each Publication p_i

Attribute	Description
$p_i.title$	title of p_i
$p_i.pubvenue$	published conference/journal of p_i
$p_i.year$	published year of p_i
$p_i.abstract$	abstract of p_i
$p_i.authors$	authors name set of p_i $\{a_i^{(0)}, a_i^{(1)}, \dots, a_i^{(w)}\}$
$p_i.references$	references of p_i

TABLE 2
Relationships between Papers

R	W	Relation Name	Description
r_1	w_1	CoPubVenue	$p_i.pubvenue = p_j.pubvenue$
r_2	w_2	CoAuthor	$\exists r, s > 0, a_i^{(r)} = a_j^{(s)}$
r_3	w_3	Citation	p_i cites p_j or p_j cites p_i
r_4	w_4	Constraint	feedback supplied by users
r_5	w_5	τ -CoAuthor	τ -extension co-authorship ($\tau > 1$)

Local features

- Similarity between a paper and its cluster centroid according to papers' attributes

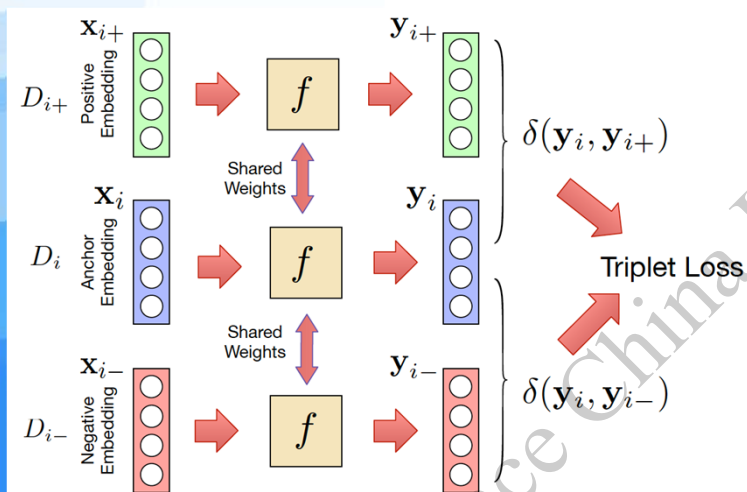
Correlation features

- Similarities between two papers according to their relationships.

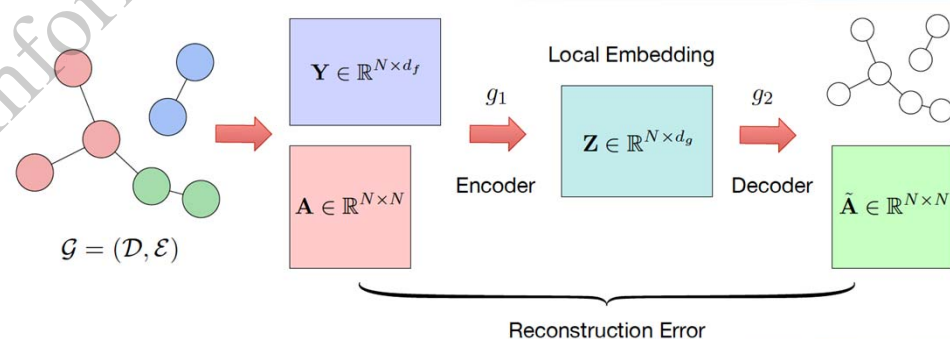
Embedding-based Matching

- Calculating similarities between embeddings.

Global embeddings



Local embeddings

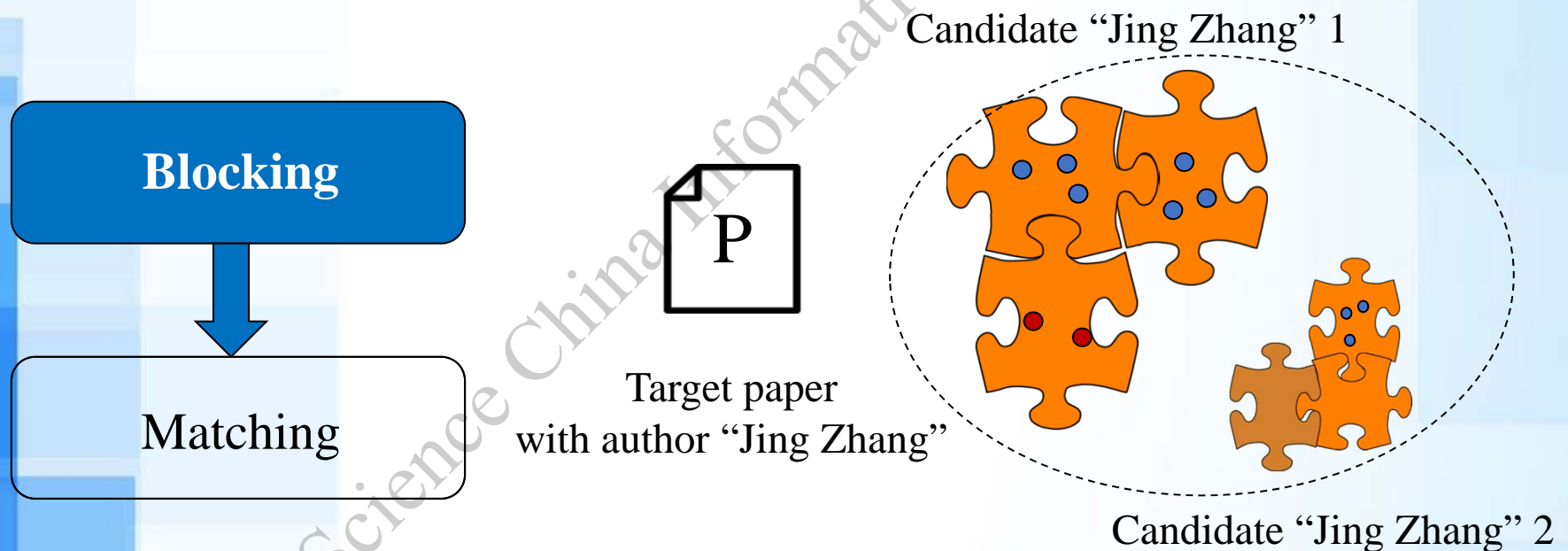


- Map every paper to a **unified representation space**.
- **Share supervision** across different candidate sets.

- For each candidate set: build a graph by linking each two similar papers.
- Train a **graph auto-encoder** to learn **separate representation space**.

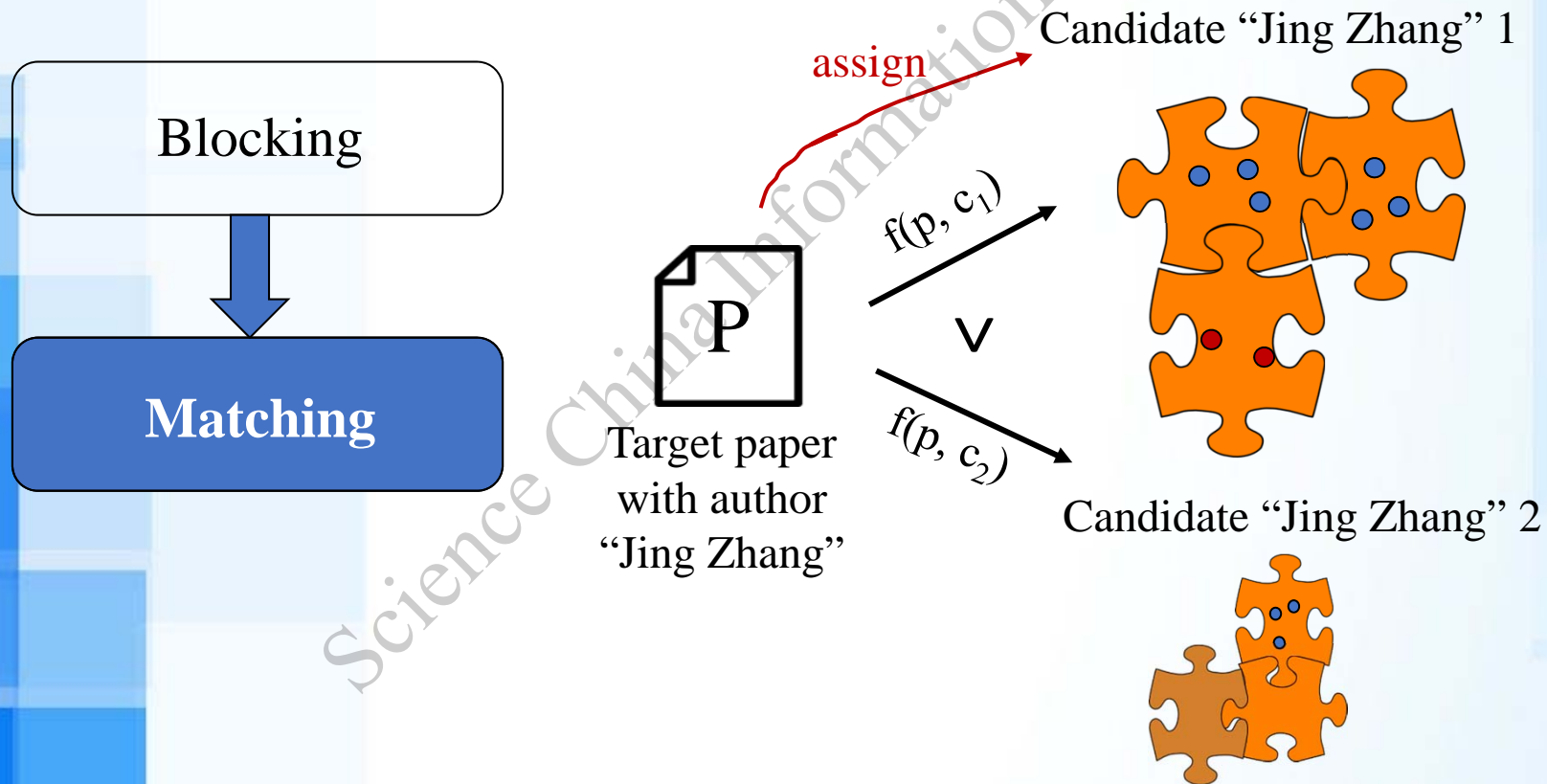
2nd Scenario: Continuous ND

- Papers come in a streaming fashion (**500,000/month**).
- Assigning new papers to right persons continuously.
- Stage One: Constructing candidates related to the author of the target paper.



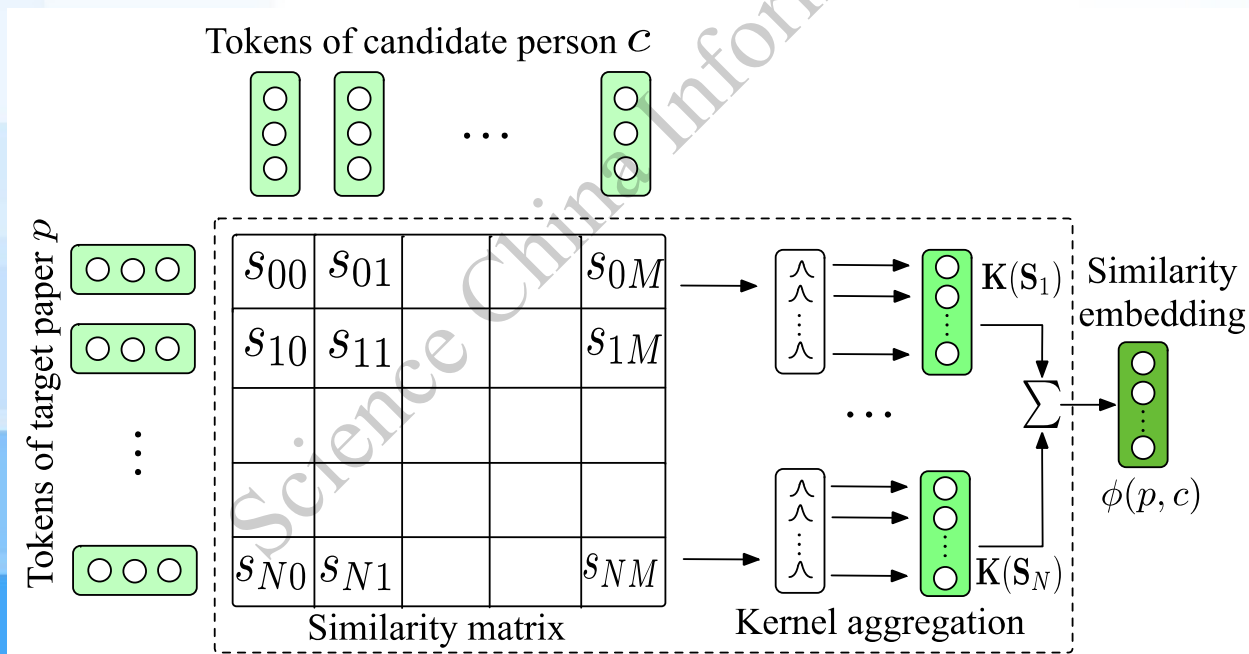
Stage Two: Similarity Matching

- Matching the target paper and each candidate.
- Assigning the paper to the candidate with the largest matching score.



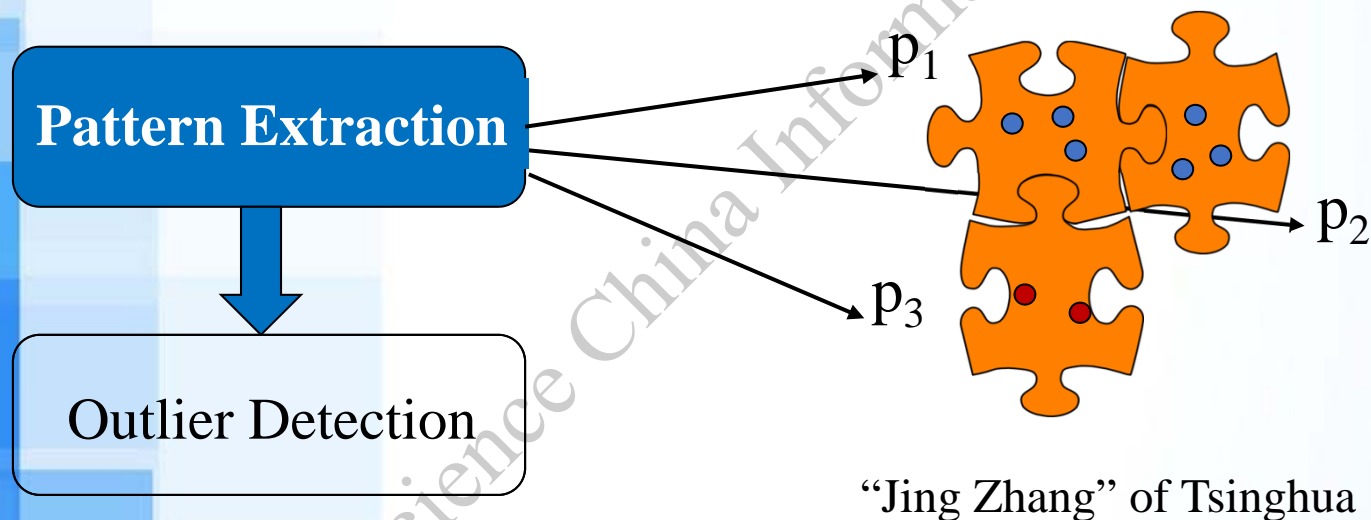
Interaction-based matching

- Calculating the similarities between **the embeddings of each pairs of tokens** in p and c .
- Capturing both **the exact and the soft matches**.



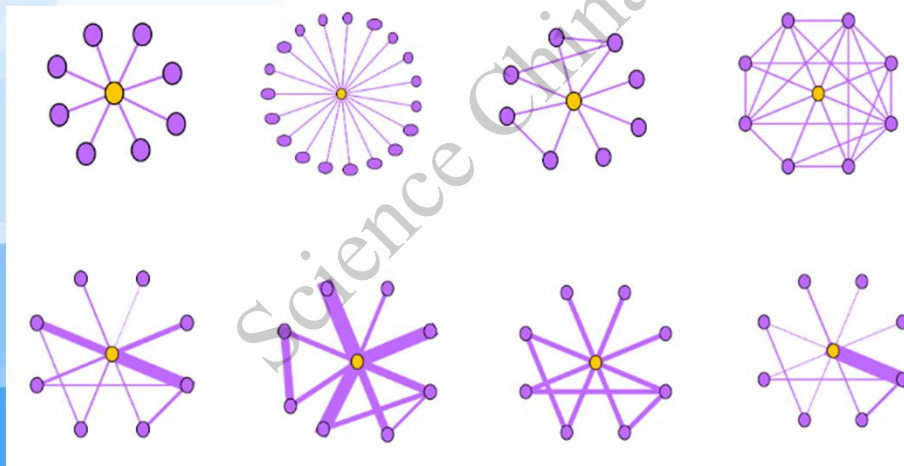
3rd Scenario: Error Detection

- The accuracy of ND algorithms cannot be 100%.
- An additional error detection function is needed.
- Stage One: Extract patterns that can distinguish the right (**normal**) and wrong (**abnormal**) assigned papers.



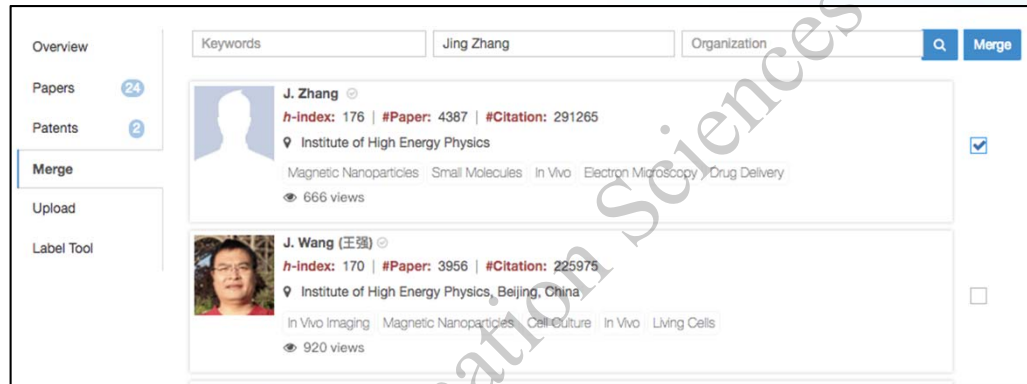
Pattern Extraction

- Construct a **multi-relation egonet** for each candidate c .
 - Co-author, co-venue, citation relationships
- For each relation, extract:
 - Number of neighbors of ego c
 - Number of edges in c 's egonet
 - Total weight of c 's egonet
 - Principal eigenvalue of the weighted adjacency matrix of c 's egonet

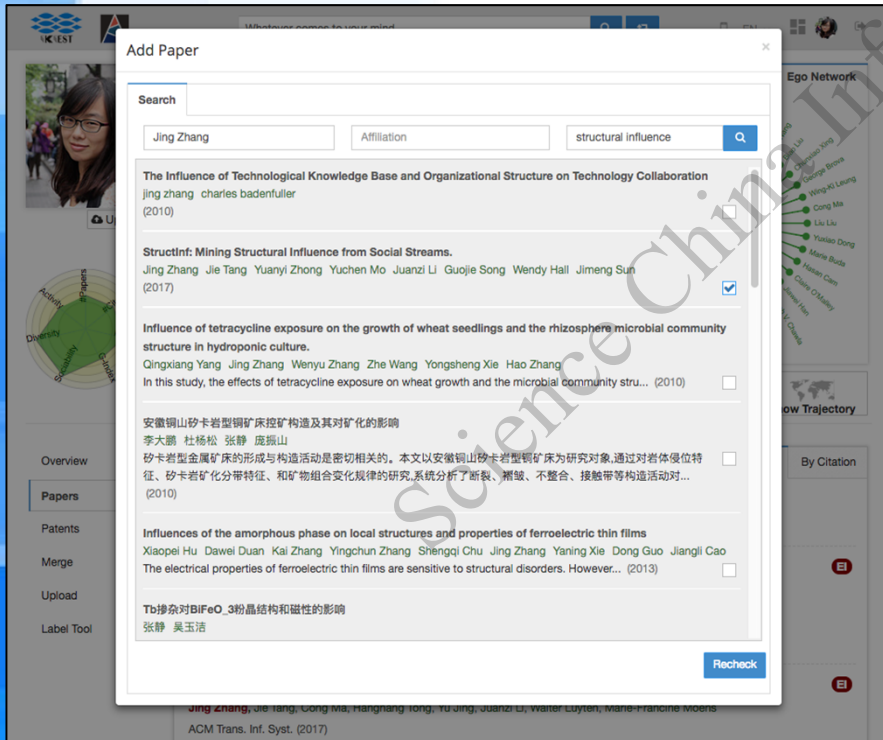


User Feedbacks

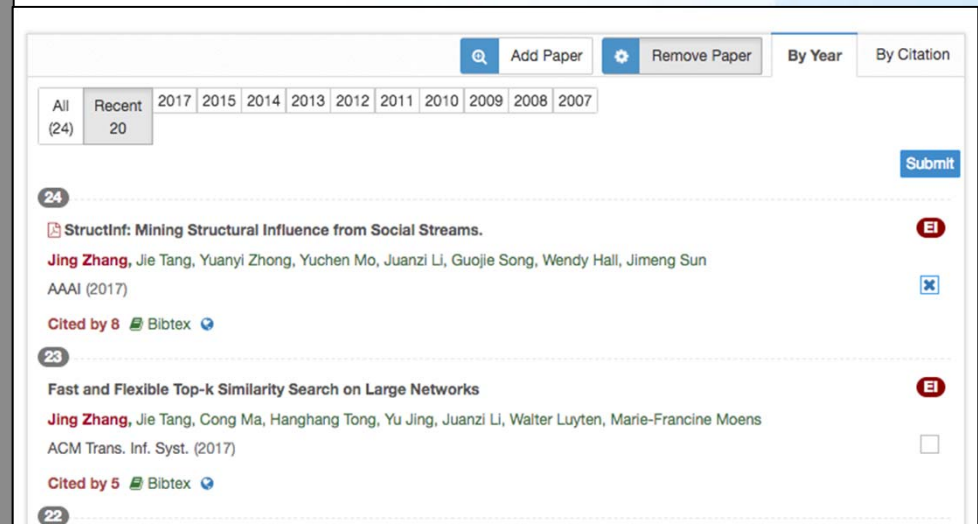
Merge Function:
Merge person profiles



Add function:
Assign new papers to persons



Remove function:
Remove the wrongly assigned papers



Name Disambiguation in AMiner:

Jing Zhang*, Jie Tang⁺

*Information School, Renmin University

⁺Computer Science Department, Tsinghua University

Dataset: <https://www.aminer.cn/na-data>

AMiner: <https://www.aminer.cn/>

