Cross-domain Collaboration Recommendation

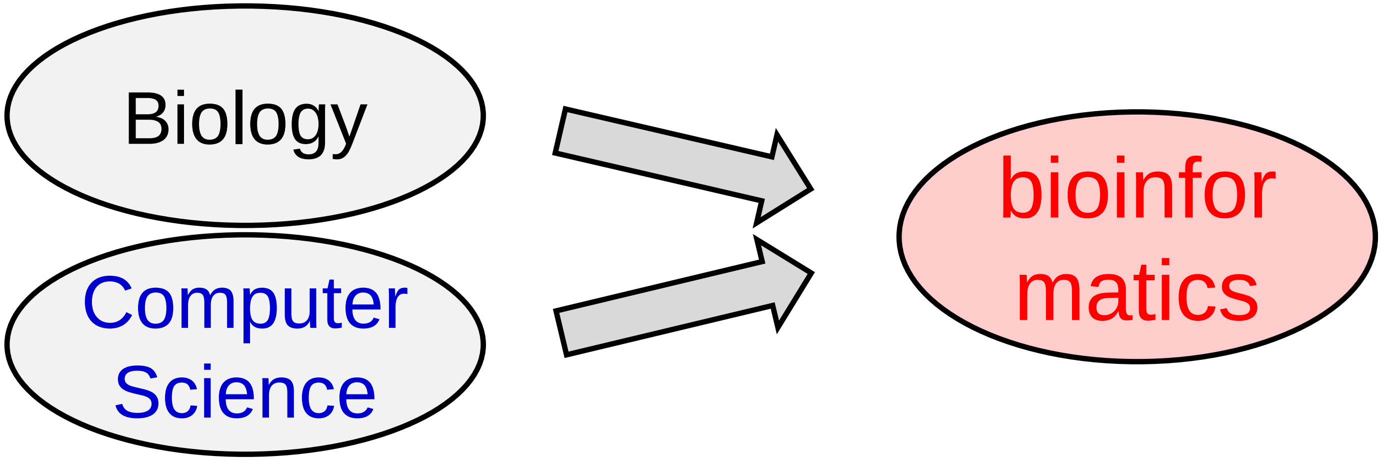
Jie Tang[1], Sen Wu[1], Jimeng Sun[2], Hang Su[1]

[1]Tsinghua University

[2]IBM TJ Watson Research Center

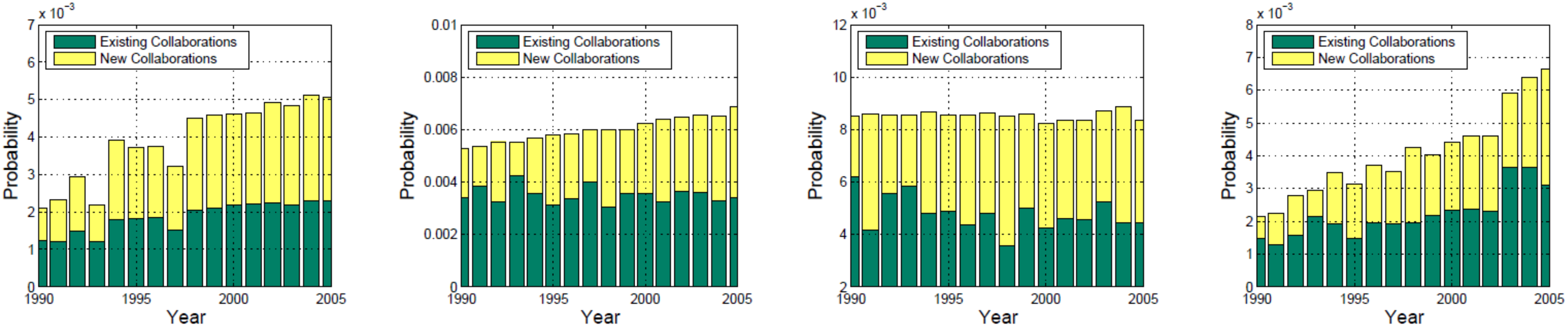
Cross-domain Collaboration

- Interdisciplinary collaborations have generated huge impact, for example,
  - 51 (>1/3) of the KDD 2012 papers are result of cross-domain collaborations between graph theory, visualization, economics, medical inf., DB, NLP, IR
  - Research field evolution



Cross-domain Collaboration (cont.)

- Increasing trend of cross-domain collaborations



(a) DM - TH (b) DM - MI (c) DM - VIS (d) MI - DB

Data Mining(DM), Medical Informatics(MI), Theory(TH), Visualization(VIS)

Challenges

Related Work-Collaboration recommendation

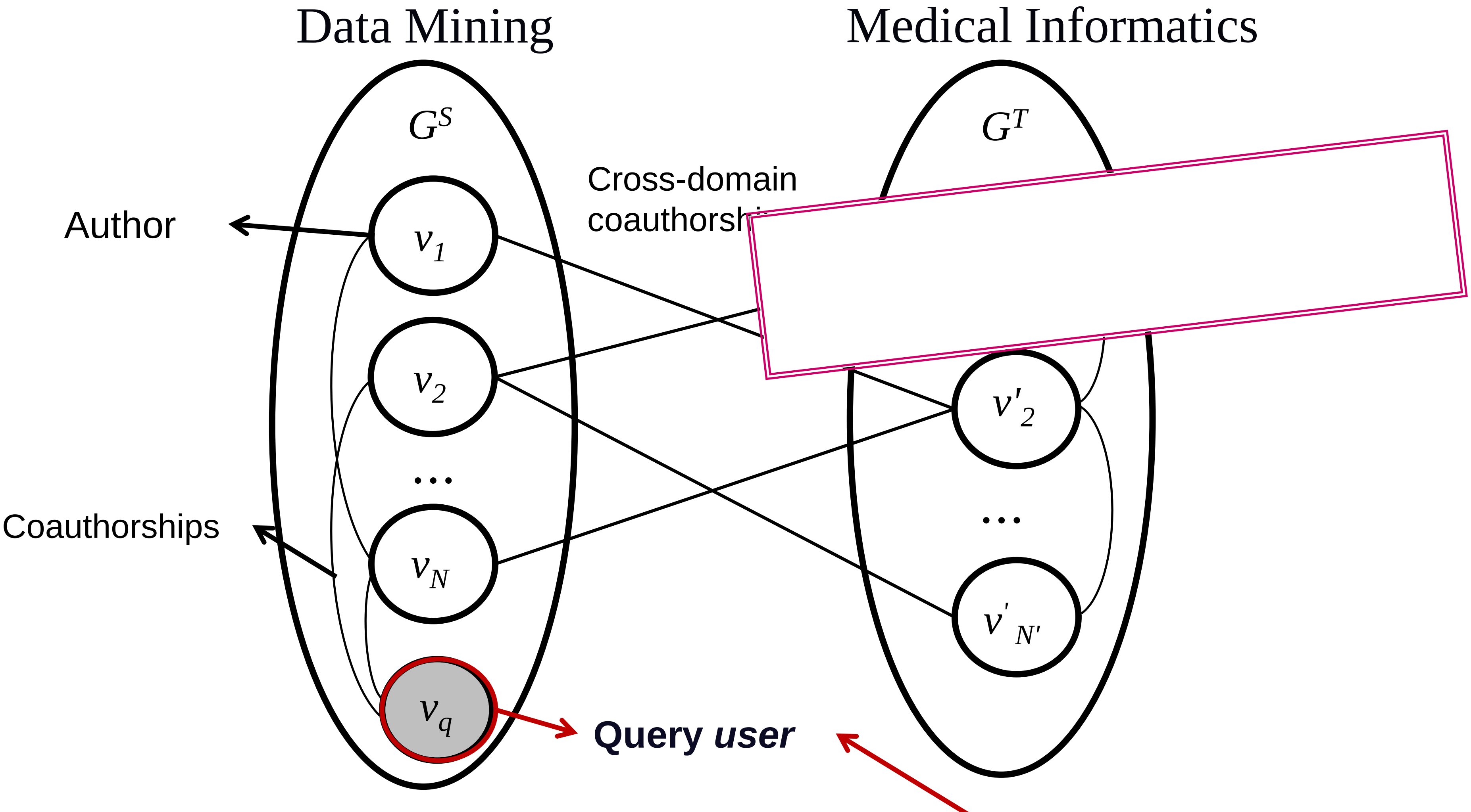
- Collaborative topic modeling for recommending papers
  - C. Wang and D.M. Blei. [2011]
- On social networks and collaborative recommendation
  - I. Konstas, V. Stathopoulos, and J. M. Jose. [2009]
- CollabSeer: a search engine for collaboration discovery
  - H.-H. Chen, L. Gou, X. Zhang, and C. L. Giles. [2007]
- Referral web: Combining social networks and collaborative filtering
  - H. Kautz, B. Selman, and M. Shah. [1997]
- Fab: content-based, collaborative recommendation
  - M. Balabanovi and Y. Shoham. [1997]

Related Work-Expert finding and matching

- Topic level expertise search over heterogeneous networks
  - J. Tang, J. Zhang, R. Jin, Z. Yang, K. Cai, L. Zhang, and Z. Su. [2011]
- Formal models for expert finding in enterprise corpora
  - K. Balog, L. Azzopardi, and M.de Rijke. [2006]
- Expertise modeling for matching papers with reviewers
  - D. Mimno and A. McCallum. [2007]
- On optimization of expertise matching with various constraints
  - W. Tang, J. Tang, T. Lei, C. Tan, B. Gao, and T. Li. [2012]

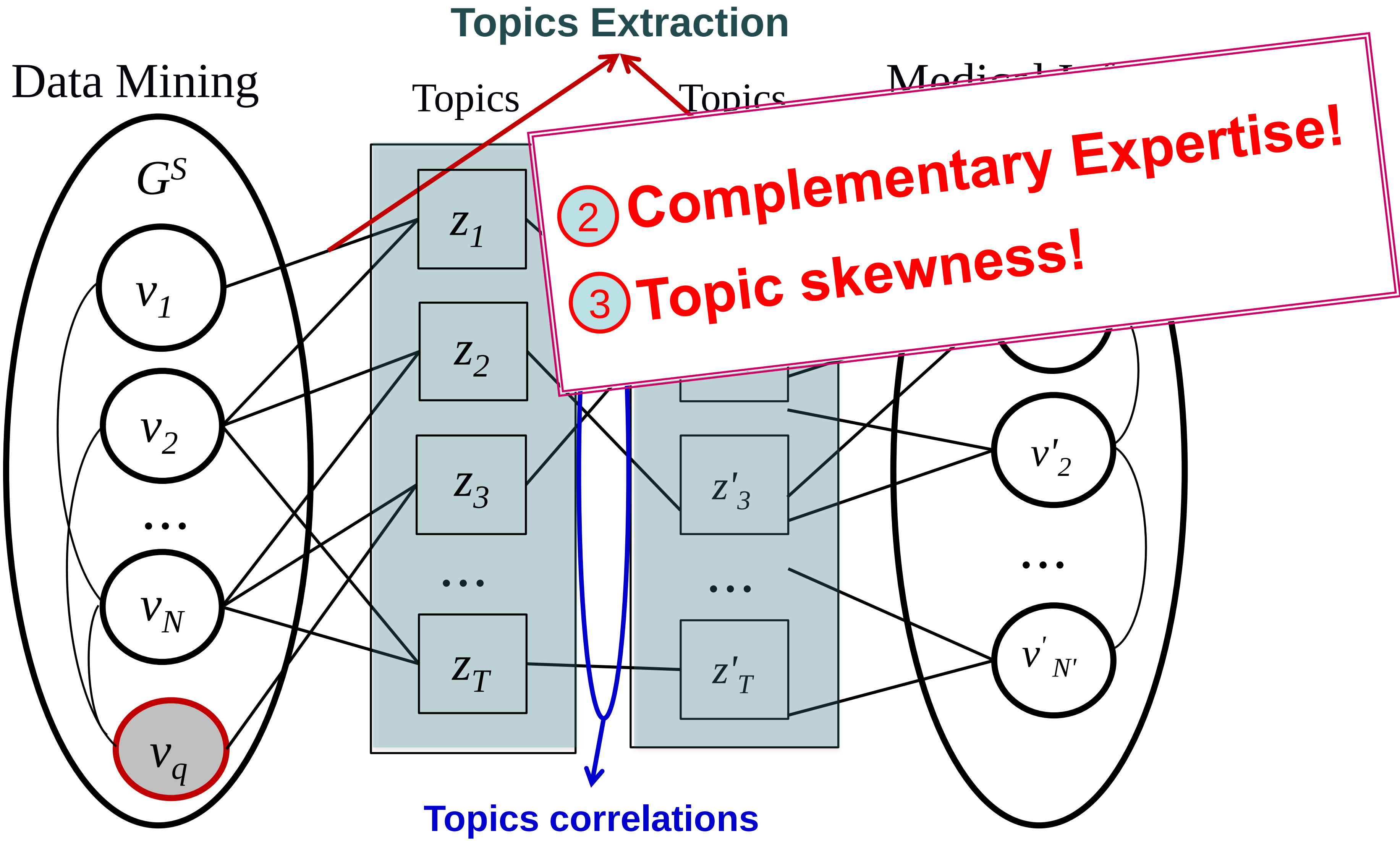
Approach Framework
—Cross-domain Topic Learning

Author Matching



$$\mathbf{r}^{(t+1)} = (1-\tau)\mathbf{S}\cdot\mathbf{r}^{(t)} + \tau\boxed{\mathbf{q}}$$

Topic Matching



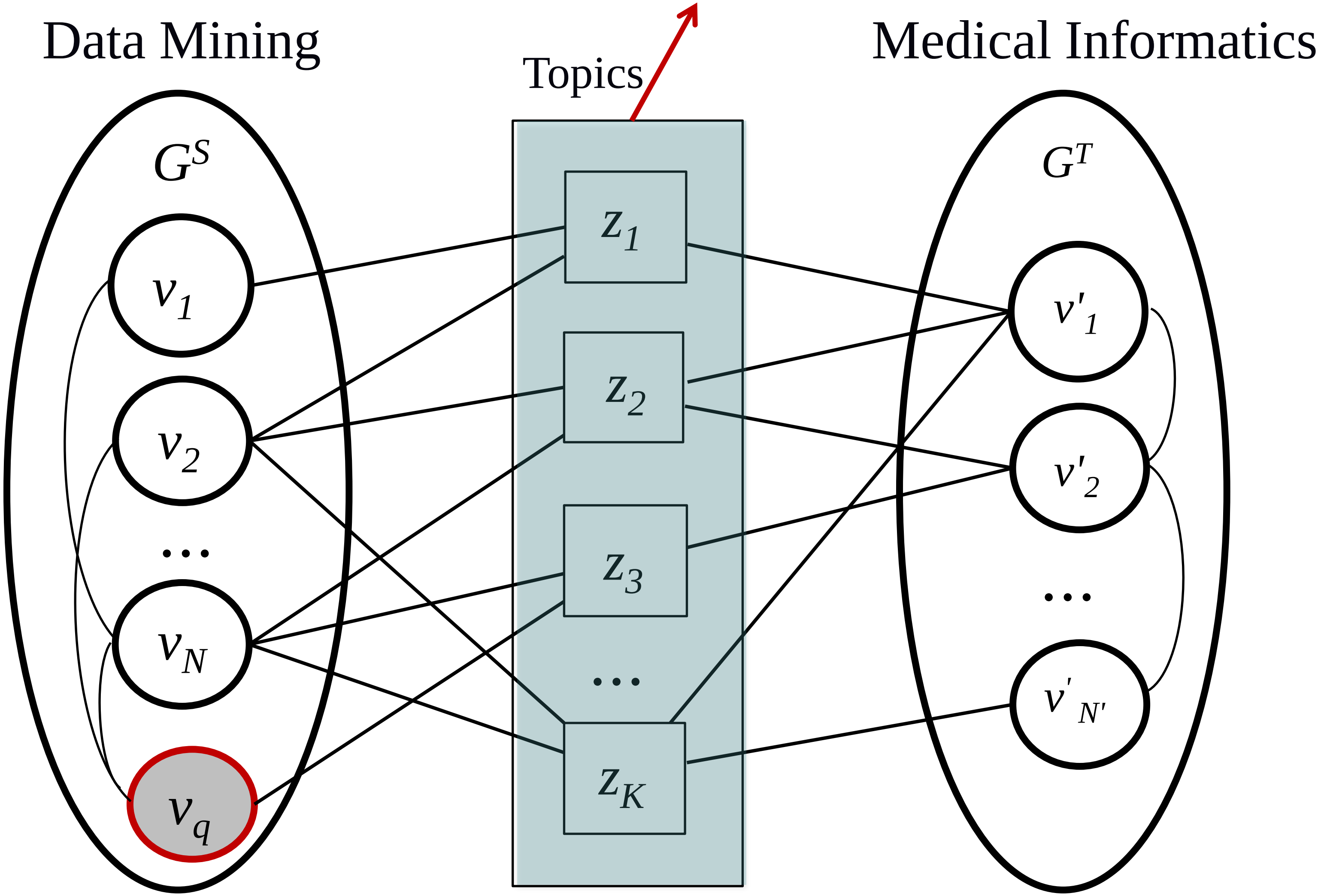
② Complementary Expertise!

③ Topic skewness!

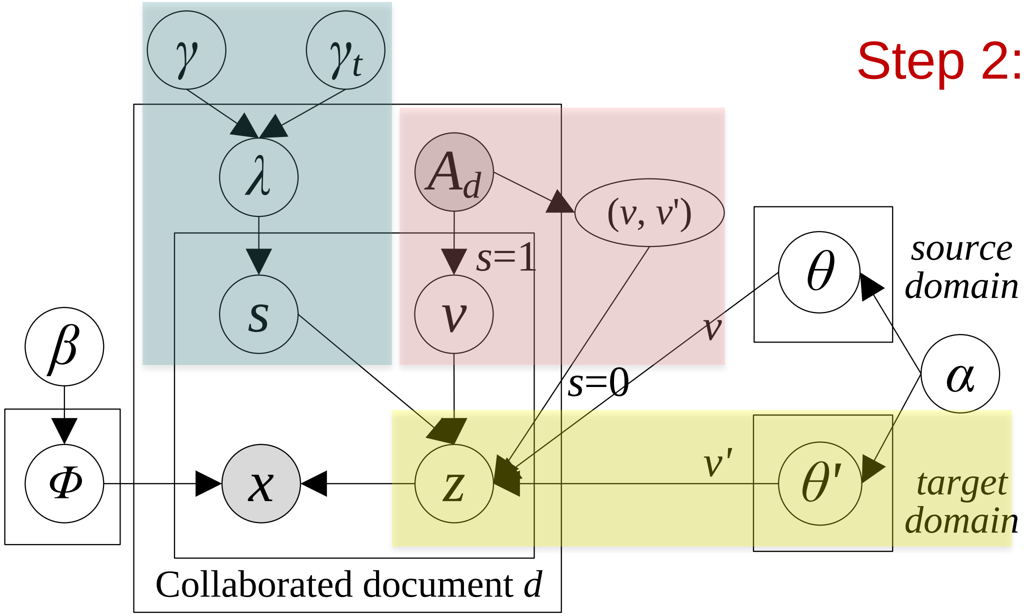
Cross-domain Topic Learning

-

Topics

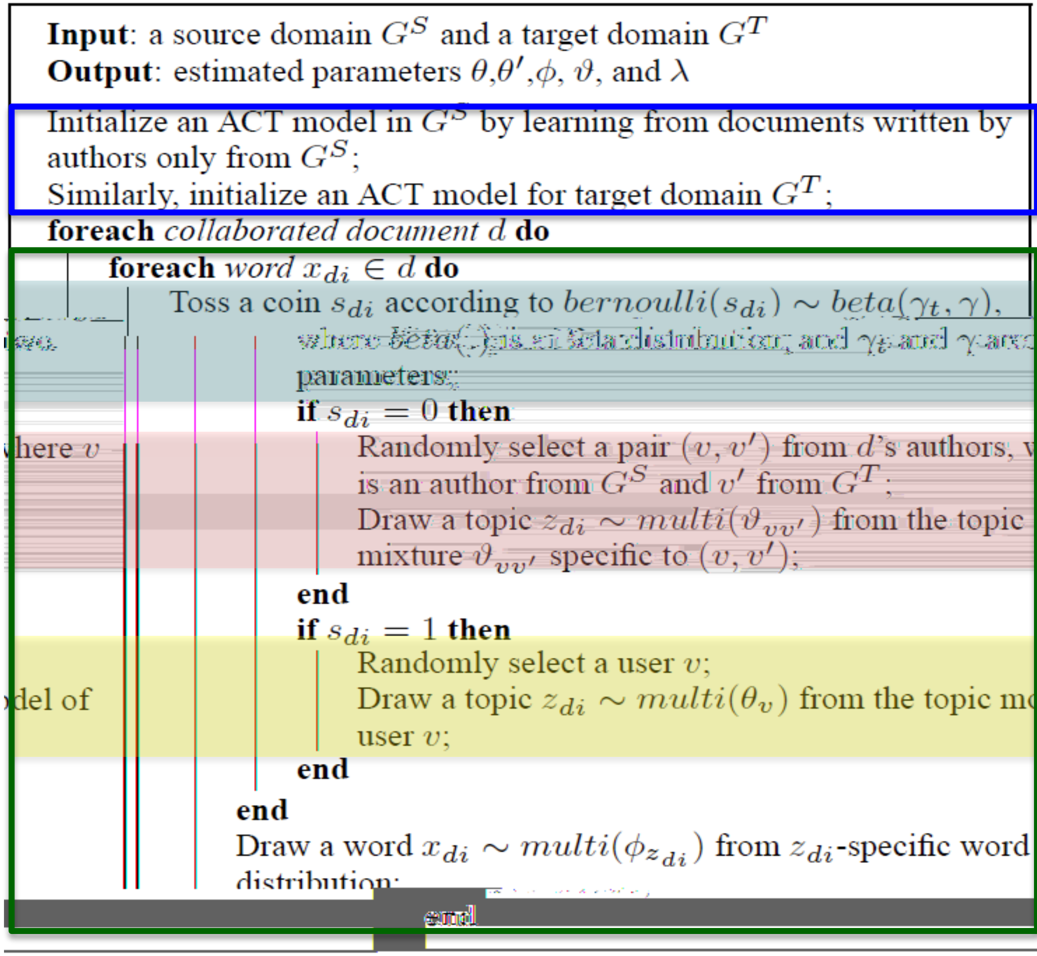


Collaboration Topics Extraction



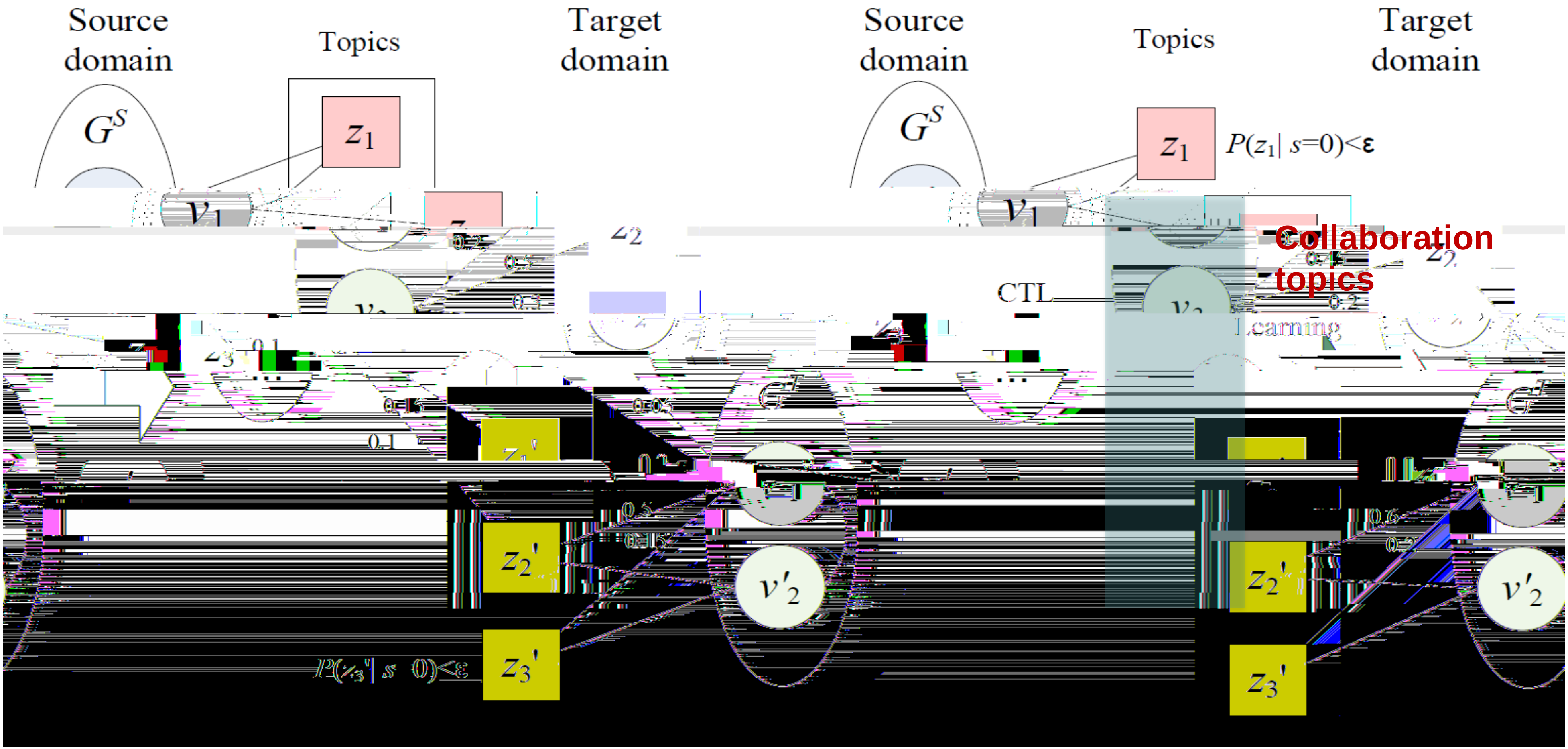
Step 1:

Step 2:



```
Input: a source domain $G^S$ and a target domain $G^T$
Output: estimated parameters $\theta$,$\theta'$,$\phi$, $\vartheta$, and $\lambda$
Initialize an ACT model in $G^S$ by learning from documents written by
authors only from $G^S$;
Similarly, initialize an ACT model for target domain $G^T$;
foreach collaborated document $d$ do
    foreach word $x_{di} \in d$ do
        Toss a coin $s_{di}$ according to $bernoulli(s_{di}) \sim beta(\gamma_t, \gamma)$,
        where $beta(\cdot)$ is a Beta distribution, and $\gamma_t$ and $\gamma$ are
        parameters;
        if $s_{di} = 0$ then
            Randomly select a pair $(v, v')$ from $d$'s authors,
            is an author from $G^S$ and $v'$ from $G^T$;
            Draw a topic $z_{di} \sim multi(\vartheta_{vv'})$ from the topic
            mixture $\vartheta_{vv'}$ specific to $(v, v')$;
        end
        if $s_{di} = 1$ then
            Randomly select a user $v$;
            Draw a topic $z_{di} \sim multi(\theta_v)$ from the topic mo
            user $v$;
        end
    end
    Draw a word $x_{di} \sim multi(\phi_{z_{di}})$ from $z_{di}$-specific word
    distribution;
end
```

Intuitive explanation of Step 2 in CTL



Collaboration topics

Experiments

Data Set and Baselines

- Arnetminer (available at http://arnetminer.org/collaboration)

Domain	Authors	Relationships	Source
Data Mining	6,282	22,862	KDD, SDM, ICDM, WSDM, PKDD
Medical Informatics	9,150	31,851	JAMIA, JBI, AIM, TMI, TITB

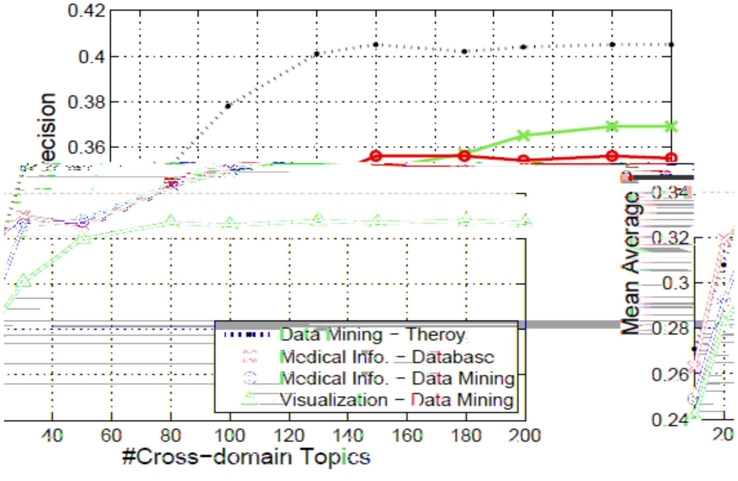
- Baselines
  - Content Similarity(Content)
  - Collaborative Filtering(CF)
  - Hybrid
  - Katz
  - Author Matching(Author), Topic Matching(Topic)

Performance Analysis

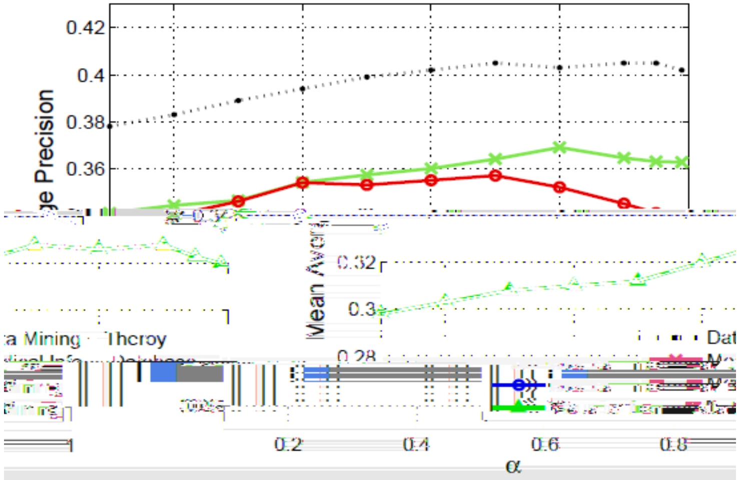
Training: collaboration before 2001   Validation: 2001-2005

Cross Domain	ALG	P@10	P@20	MAP	R@100	ARHR -10	ARHR -20
Data Mining(S) to Theory(T)	Content	10.3	10.2	10.9	31.4	4.9	2.1
	CF	15.6	13.3	23.1	26.2	4.9	2.8
	Hybrid	17.4	19.1	20.0	29.5	5.0	2.4
	Author	27.2	22.3	25.7	32.4	10.1	6.4
	Topic	28.0	26.0	32.4	33.5	13.4	7.1
	Katz	30.4	29.8	21.6	27.4	11.2	5.9
	CTL	37.7	36.4	40.6	35.6	14.3	7.5

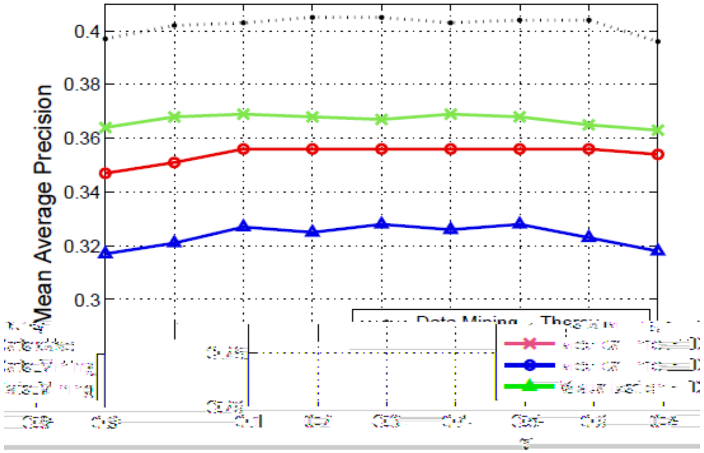
Parameter Analysis



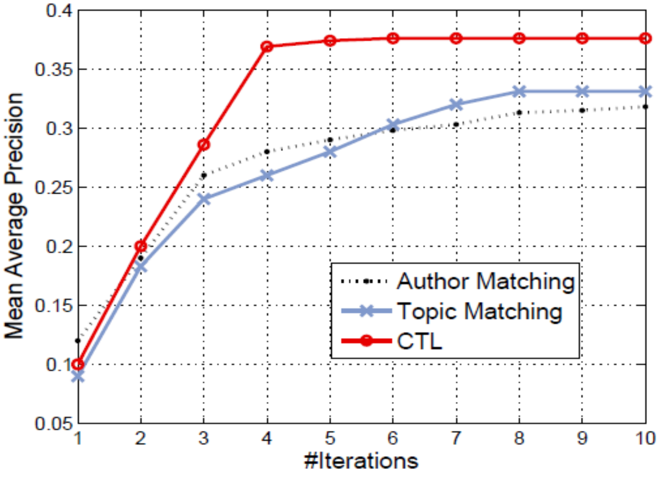
(a) number of topics T



(b) Hyperparameter α



(c) RWR parameter τ



(d) Convergence analysis

(a) varying the number of topics T

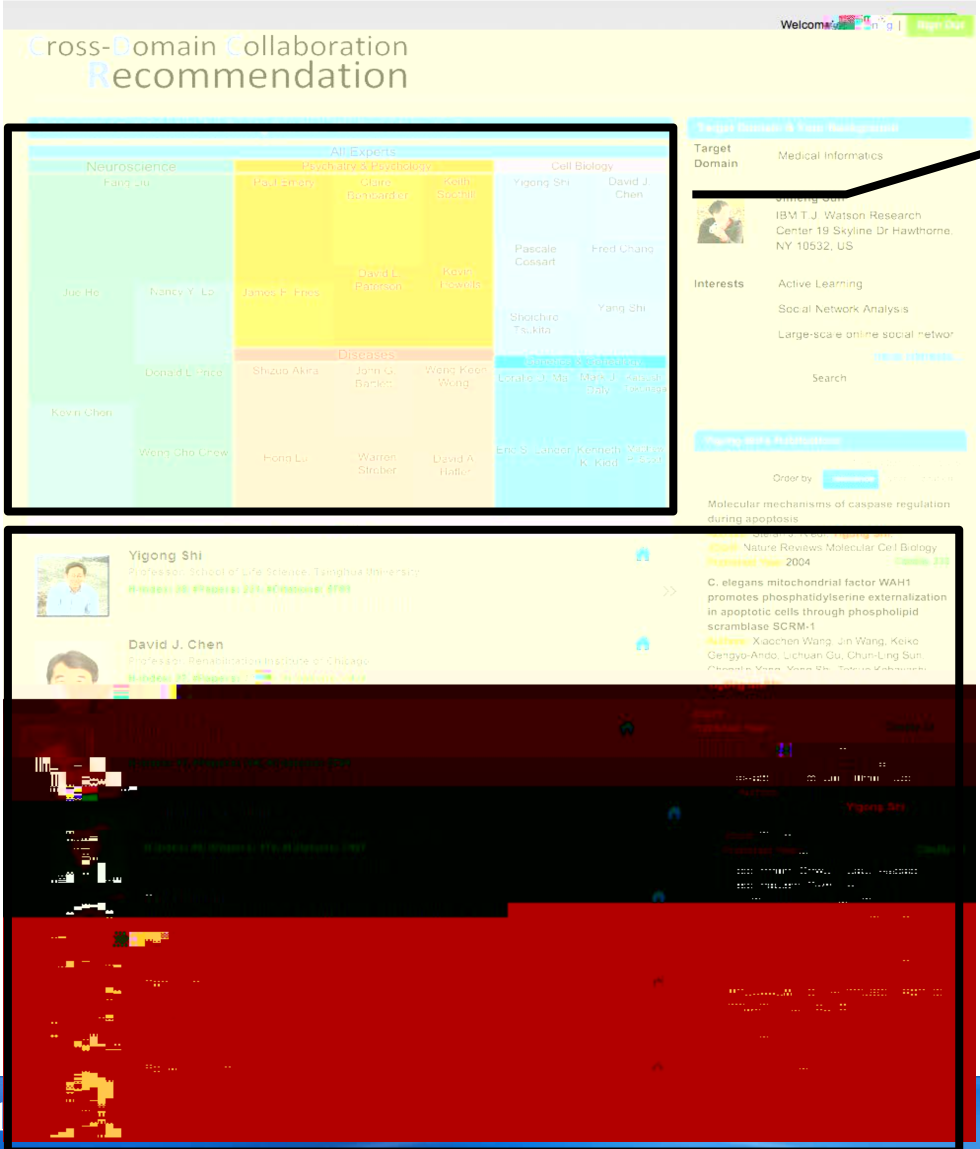
(b) varying α parameter

(c) varying the restart parameter τ in the random walk

(d) Convergence analysis

Prototype System

http://arnetminer.org/collaborator



Welcome ... | Sign Out

Cross-Domain Collaboration Recommendation

Target Domain & Your Background

Target Domain: Medical Informatics

Jimeng Sun
IBM T.J. Watson Research Center 19 Skyline Dr Hawthorne, NY 10532, US

Interests: Active Learning
Social Network Analysis
Large-scale online social networ

more interests...

Search

Yigong Shi's Publications

Order by: relevance year citation

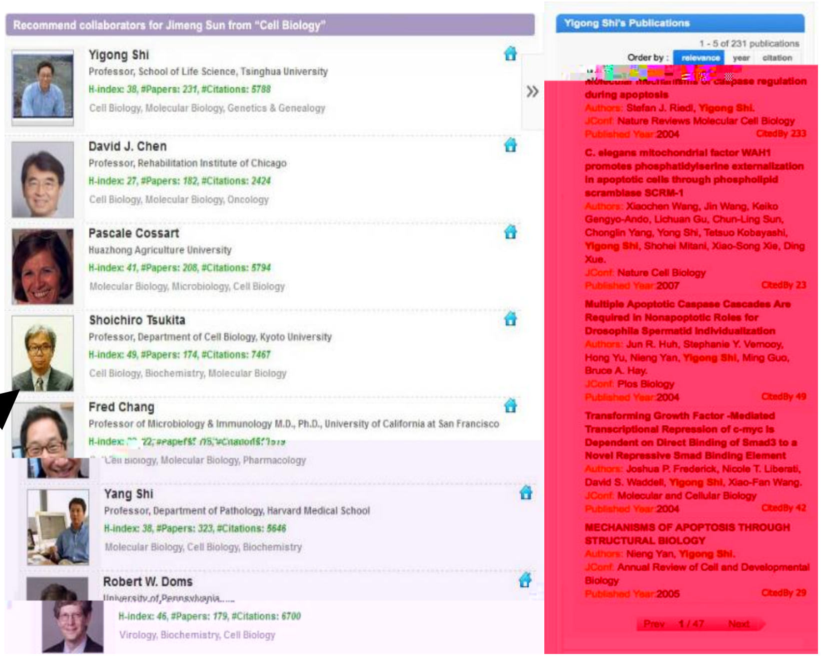
Molecular mechanisms of caspase regulation during apoptosis
Authors: Stefan J. Riedl, Yigong Shi.
JConf: Nature Reviews Molecular Cell Biology
Published Year: 2004 CitedBy 233

C. elegans mitochondrial factor WAH1 promotes phosphatidylserine externalization in apoptotic cells through phospholipid scramblase SCRM-1
Authors: Xiaochen Wang, Jin Wang, Keiko Gengyo-Ando, Lichuan Gu, Chun-Ling Sun, ...

Yigong Shi
Professor, School of Life Science, Tsinghua University
H-index: 38, #Papers: 231, #Citations: 5788

David J. Chen
Professor, Rehabilitation Institute of Chicago
H-index: 27, #Papers: ...

Treemap: representing subtopic in the target domain



Recommend collaborators for Jimeng Sun from "Cell Biology"

Yigong Shi
Professor, School of Life Science, Tsinghua University
H-index: 38, #Papers: 231, #Citations: 5788
Cell Biology, Molecular Biology, Genetics & Genealogy

David J. Chen
Professor, Rehabilitation Institute of Chicago
H-index: 27, #Papers: 182, #Citations: 2424
Cell Biology, Molecular Biology, Oncology

Pascale Cossart
Huazhong Agriculture University
H-index: 41, #Papers: 208, #Citations: 5794
Molecular Biology, Microbiology, Cell Biology

Shoichiro Tsukita
Professor, Department of Cell Biology, Kyoto University
H-index: 49, #Papers: 174, #Citations: 7467
Cell Biology, Biochemistry, Molecular Biology

Fred Chang
Professor of Microbiology & Immunology M.D., Ph.D., University of California at San Francisco
H-index: 22, #Papers: 175, #Citations: 1519
Cell Biology, Molecular Biology, Pharmacology

Yang Shi
Professor, Department of Pathology, Harvard Medical School
H-index: 38, #Papers: 323, #Citations: 5646
Molecular Biology, Cell Biology, Biochemistry

Robert W. Doms
University of Pennsylvania
H-index: 46, #Papers: 179, #Citations: 6700
Virology, Biochemistry, Cell Biology

Yigong Shi's Publications

1 - 5 of 231 publications

Order by : relevance year citation

Molecular mechanisms of caspase regulation during apoptosis
Authors: Stefan J. Riedl, Yigong Shi.
JConf: Nature Reviews Molecular Cell Biology
Published Year: 2004 CitedBy 233

C. elegans mitochondrial factor WAH1 promotes phosphatidylserine externalization in apoptotic cells through phospholipid scramblase SCRM-1
Authors: Xiaochen Wang, Jin Wang, Keiko Gengyo-Ando, Lichuan Gu, Chun-Ling Sun, Chonglin Yang, Yong Shi, Tetsuo Kobayashi, Yigong Shi, Shohei Mitani, Xiao-Song Xie, Ding Xue.
JConf: Nature Cell Biology
Published Year: 2007 CitedBy 23

Multiple Apoptotic Caspase Cascades Are Required in Nonapoptotic Roles for Drosophila Spermatid Individualization
Authors: Jun R. Huh, Stephanie Y. Vernooy, Hong Yu, Nieng Yan, Yigong Shi, Ming Guo, Bruce A. Hay.
JConf: Plos Biology
Published Year: 2004 CitedBy 49

Transforming Growth Factor -Mediated Transcriptional Repression of c-myc Is Dependent on Direct Binding of Smad3 to a Novel Repressive Smad Binding Element
Authors: Joshua P. Frederick, Nicole T. Liberati, David S. Waddell, Yigong Shi, Xiao-Fan Wang.
JConf: Molecular and Cellular Biology
Published Year: 2004 CitedBy 42

MECHANISMS OF APOPTOSIS THROUGH STRUCTURAL BIOLOGY
Authors: Nieng Yan, Yigong Shi.
JConf: Annual Review of Cell and Developmental Biology
Published Year: 2005 CitedBy 29

Prev 1 / 47 Next

Recommend Collaborators & Their relevant publications

Conclusion

- Study the problem of cross-domain collaboration recommendation
- Propose the cross-domain topic model for recommending collaborators
- Experimental results in a coauthor network demonstrate the effectiveness and efficiency of the proposed approach

Future work

- Connect cross-domain collaborative relationships with social theories (e.g. social balance, social status, structural hole)

- Apply the proposed method to other networks

Thanks!

System: http://arnetminer.org/collaborator

Code&Data: http://arnetminer.org/collaboration

Challenge always be side with opportunity!

- Sparse connection:
  - cross-domain collaborations are rare;
- Complementary expertise:
  - cross-domain collaborators often have different expertise and interest;
- Topic skewness:
  - cross-domain collaboration topics are focused on a subset of topics.

Performance Analysis

Cross Domain	ALG	P@10	P@20	MAP	R@100	ARHR -10	ARHR -20
Medical Info.(S) to Database(T)	Content	10.1	10.9	12.5	45.9	3.6	2.1
	CF	18.3	20.2	21.4	47.6	5.3	3.9
	Hybrid	25.0	26.5	28.4	59.1	6.4	4.2
	Author	26.2	29.6	32.2	54.8	10.5	5.4
	Topic	29.4	26.3	34.7	59.3	11.5	5.2
	Katz	27.5	28.3	30.7	57.2	10.5	5.0
	CTL	32.5	30.0	36.9	59.8	11.4	5.4

Performance Analysis

Cross Domain	ALG	P@10	P@20	MAP	R@100	ARHR -10	ARHR -20
Medical Info.(S) to Data Mining(T)	Content	5.8	5.7	9.5	19.8	1.9	0.9
	CF	13.7	17.8	18.9	34.3	2.7	1.3
	Hybrid	18.0	19.0	19.8	36.7	3.4	1.3
	Author	20.1	23.8	29.3	64.4	5.3	2.1
	Topic	26.0	25.0	33.9	48.1	10.7	5.6
	Katz	21.2	23.8	32.4	48.1	10.2	4.8
	CTL	30.0					

Performance Analysis

Cross Domain	ALG	P@10	P@20	MAP	R@100	ARHR -10	ARHR -20
Visual.(S) to Data Mining(T)	Content	9.6	11.8	13.2	18.9	3.1	1.8
	CF	14.0	20.8	26.4	29.4	6.9	4.3
	Hybrid	16.0	20.0	27.6	30.1	6.3	4.4
	Author	22.0	25.2	27.7	31.1	11.9	6.7
	Topic	26.3	25.0	32.3	31.4	13.2	8.8
	Katz	23.0	25.1	29.3	30.2	10.4	5.4
	CTL	28.3	26.0	32.8	36.3	14.0	9.1