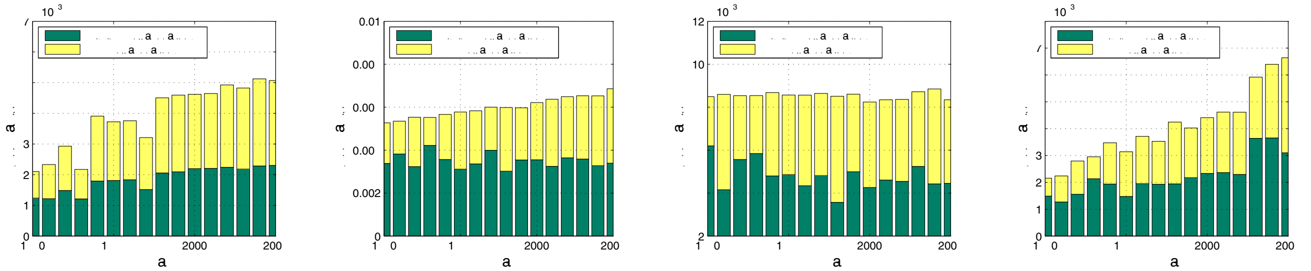
Cross-domain Collaboration Recommendation

[illegible]

ABSTRACT

[illegible] 1) **sparse connection:** [illegible]; 2) **complementary expertise:** [illegible]; 3) **topic skewness:** [illegible]



( ) DM - H ( ) DM - MI ( ) DM - I ( ) MI - DB

Figure 1: The comparison of existing collaboration and new collaboration trends over years. DM - Data Mining domain; MI - Medical Informatics domain; TH - Theory domain; VIS - Visualization domain; DB - Database domain. The trends of cross-domain collaborations in all but one case are growing (The exception between DM and VIS remain roughly constant over time). Newly formed cross-domain collaborations are significantly in all cases.

F 1.

(C L)

C L

C L

C L

C L

C L

C L

2 3 4 5 6.

2. PROBLEM DEFINITION

D 1. **Source/Target domain.** $G = (V, E, X)$, V $|V| = N$ $E \subseteq V \times V$ X $N \times d$ x_j j^{th}

S T S v_i d

G

D 2. **Domain-specific topic models.** A θ_i v_i $\{P(x_j|\theta_i)\}_j$. $P(x|\theta_i)$.

LDA 4, 15 D M

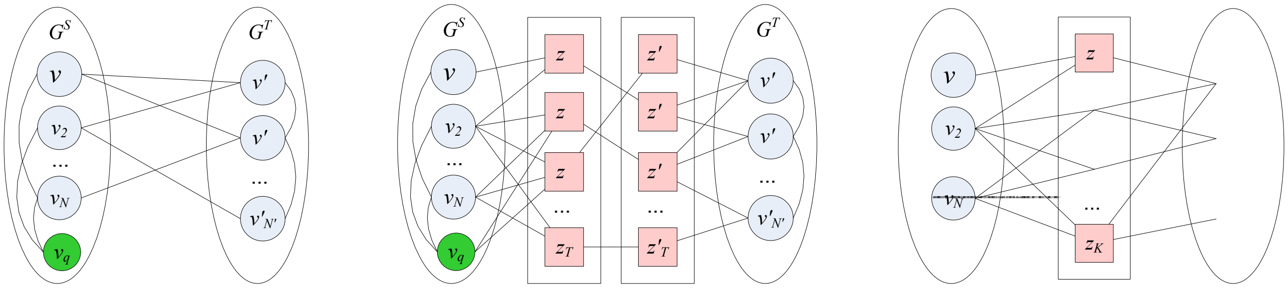
G^S G^T $V^S \cap V^T \neq \varnothing$.

P 1. **Cross-domain collaboration recommendation.** G (1) G^S G^T, (2) θ θ' v_q

θ) H

3. CROSS-DOMAIN TOPIC LEARNING

28



( ) A

( )

Input: G^S G^T

Output: $\theta, \theta', \phi, \vartheta, \lambda$

AC G^S

G^S;

AC G^T;

foreach d **do**

foreach $\in d$ **do**

$\sim$ $a(\gamma_t, \gamma)$,

$a(.)$ γ_t γ

if $= 0$ **then**

G^S G^T;

$\sim$ $l\ i(\vartheta_{vv'})$

$\vartheta_{vv'}$

end

if $= 1$ **then**

$\sim$ $l\ i(\theta_v)$

end

end

$\sim$ $l\ i(\phi_{z_{di}})$

end

Algorithm 1: C L.

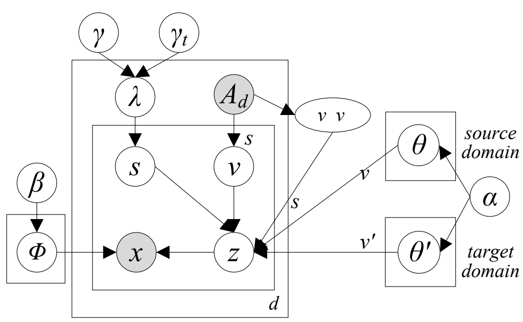


Figure 3: Graphical representation of CTL model.

z z' v $P(z|v)$ AC z' E . 1

3.3 Cross-domain Topic Learning (CTL)

A

C

(C L)

Model description.

F 3

Table 1: Notations in the CTL model.

T	
d	
A_d	d
	d
θ_v	
$\vartheta_{vv'}$	
ϕ_z	
α, β	θ, θ' ϕ
λ	
γ, γ_t	λ

C L . (F

A_d d; v (v, v') x; s $(s = 1)$ $s = 0$; θ θ' $\vartheta_{vv'}$ (v, v'); α λ; s; γ γ_t λ. 1 C L

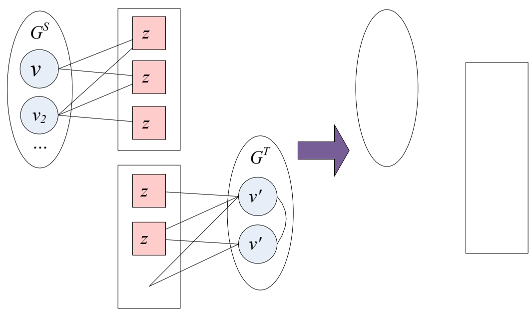
F A 1: G^S G^T

C L $p(\theta_v|\alpha)$ $p(\theta'_{v'}|\alpha)$, α α

C L

F x_{di} d, s $p(s|d) \sim beta(\gamma_t, \gamma)$, $beta(.)$ $s = 1$, v (v') x_{di} z_{di} v, θ_v ($s = 0$, (v, v') $\vartheta_{vv'}$ θ_v $\theta_{v'}$ (M . F 10 5 15 (10 5). $\tilde{\theta}_v = < \theta_v, 0, \ldots, 0 >$, 0 $\tilde{\theta}'_{v'} = < 0, \ldots, 0, \theta'_v >$. $\vartheta_{vv'}$ $\tilde{\theta}_v + \tilde{\theta}_{v'}$, θ_v $\theta_{v'}$ 5 . F x_{di} z_{di} $\vartheta_{vv'}$.

F 4 C L . C L



G^S
v
v_2
...
z
z
z
z
z
G^T
v'
v'

1,932,442 [illegible] 1990 [illegible] 2005 [illegible]:

- **Data Mining:** [illegible]: KDD, [illegible] DM, ICDM [illegible] DM [illegible] KDD [illegible], [illegible] 6,282 [illegible] 22,862 [illegible].
- **Medical Informatics:** [illegible]: J [illegible] A [illegible] M [illegible] I [illegible] A [illegible], J [illegible] B [illegible] I [illegible], A [illegible] I [illegible] M [illegible], IEEE [illegible] M [illegible] I [illegible] IEEE [illegible] I [illegible] B [illegible], [illegible] 9,150 [illegible] 31,851 [illegible].
- **Theory:** [illegible], [illegible] C, F [illegible] C [illegible] DA, [illegible] 5,449 [illegible] 27,712 [illegible].
- **Visualization:** [illegible]

Table 2: Recommendation performance by different methods on the four cross-domain test cases (%). Content— Content Similarity; CF— Collaborative Filtering; Author— Author Matching; Topic— Topic Matching.

Cross domain	ALG	P@10	P@20	MAP	R@100	ARHR -10	ARHR -20
[illegible]	C [illegible]	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
	H [illegible]	17.4	19.1	20.0	29.5	5.0	2.4
	A [illegible]	27.2	22.3	25.7	32.4	10.1	6.4
	[illegible]	28.0	26.0	32.4	33.5	13.4	7.1
	K [illegible]	30.4	29.8	31.6	27.4	11.2	5.9
	C L	**37.7**	**36.4**	**40.6**	**35.6**	**14.3**	**7.5**
[illegible]	C [illegible]	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
	H [illegible]	25.0	26.5	28.4	59.1	6.4	4.2
	A [illegible]	26.2	29.6	32.2	54.8	10.5	**5.4**
	[illegible]	29.4	26.3	34.7	59.3	**11.5**	5.2
	K [illegible]	27.5	28.3	30.7	57.2	10.5	5.0
	C L	**32.5**	**30.0**	**36.9**	**59.8**	11.4	**5.4**
[illegible]	C [illegible]	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
	H [illegible]	18.0	19.0	19.8	36.7	3.4	1.3
	A [illegible]	20.1	23.8	29.3	**64.4**	5.3	2.1
	[illegible]	26.0	**25.0**	33.9	48.1	10.7	5.6
	K [illegible]	21.2	23.8	32.4	48.1	10.2	4.8
	C L	**30.0**	24.0	**35.6**	49.6	**12.2**	**6.0**
[illegible]	C [illegible]	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
	H [illegible]	16.0	20.0	27.6	30.1	6.3	4.4
	A [illegible]	22.0	25.2	27.7	31.1	11.9	6.7
	[illegible]	26.3	25.0	32.3	31.4	13.2	8.8
	K [illegible]	23.0	25.1	29.3	30.2	10.4	5.4
	C L	**28.3**	**26.0**	**32.8**	**36.3**	**14.0**	**9.1**

[illegible] (< 80), [illegible] 150. [illegible] C L [illegible]

Hyperparameter analysis. [illegible] α [illegible] C L [illegible] F[illegible] 5( ) [illegible] C L [illegible] α [illegible] $T = 120$) [illegible] α, [illegible] 0.03 [illegible] C L [illegible] α.

Restart parameter analysis. [illegible] τ [illegible] F[illegible] 5( ) [illegible] C L [illegible] τ [illegible] τ. B [illegible] τ [illegible] τ [illegible]

Convergence analysis. [illegible] F[illegible] 5( ) [illegible] -D M [illegible] 10 [illegible] C L [illegible]

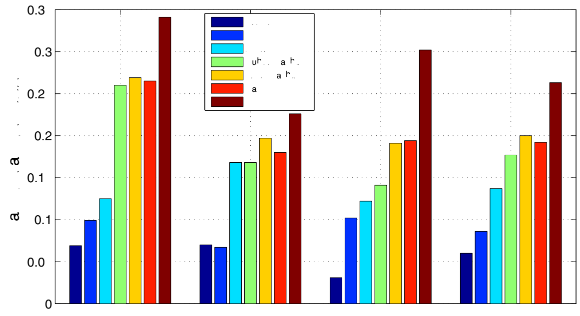


Figure 6: Performance on new collaboration prediction of all algorithms.

([illegible] 5 [illegible]). [illegible] C L [illegible]

New Collaboration Prediction [illegible] C [illegible] 2001 [illegible] 2001-2005 [illegible] F[illegible] 6 [illegible] 2. I [illegible] C L [illegible] 0.3 [illegible] MA [illegible]

4.3 Prototype System

[illegible] C L [illegible][5]. [illegible] C L [illegible] ([illegible] 1,932,442 [illegible]) A [illegible] 1 [illegible] (C [illegible] 3.3) [illegible] C L [illegible]

5. RELATED WORK

C [illegible] C [illegible] [7] [illegible] H [illegible] K [illegible] [17] [illegible]

[5] [illegible]://[illegible]

[22] L. L . : A . C b. a c ,
2(1):1?6, 1993.

[23] D. M A. M C . E
. I KDD'07, 500 509, 2007.

[24] J. . M . N w
E a d J a M d c , 354:2463 2472, J 2006.

[25] D. . C . I KDD'10,
979 988, 2010.

[26] . , D. , A. G , . A.
. I SIGIR'11, 1337 1338, 2011.

[27] M. , . , . G .
. I KDD'04, 306 315, 2004.

[28] J. , H. , D. C , C. F .
. I ICDM'05,
418 425, 2005.

[29] J. , L. , D. C . M - -
. I SDM'09, 1147 1158, 2009.

[30] J. , J. , J , , K. C , L. ,
. Mac
L a J a , 82(2):211 237, 2011.

[31] J. , J. , L. , J. L , L. , . A :
E . I KDD'08,
990 998, 2008.

[32] L. H. L .
I KDD'09, 817 826, 2009.

[33] E. , J. , . L , C. , B. G , . L .
. N c ,
76(1):71 83, 2012.

[34] C E D. M. B . C
. I KDD'11, 448 456, 2011.

[35] . , L. C , . F :
- .
I R cS '11, 245 252, 2011.

[36] J. , J. , J. L . E I
DASFAA'07, 1066 1069, 2007.

8. APPENDIX

A , (B) $\theta, \theta', \vartheta, \lambda, \varphi$,
10 4075(10 0 (,)-256( )-255 / 1 01 1 2115. 7,)-145() J/ 1 41 0 -1.1660.0001 (I) / 1 11