Cross-domain Collaboration Recommendation

[illegible]

ABSTRACT

In [illegible] H [illegible] ? H [illegible] ? C [illegible] ?

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

In [illegible] C [illegible] (C[illegible]L) [illegible]. F [illegible], C[illegible]L [illegible]. F [illegible], C[illegible]L [illegible]

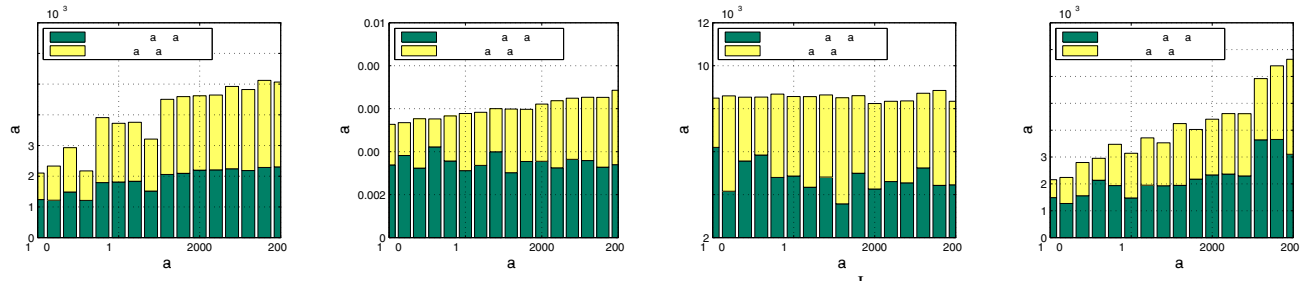


Figure 1: The comparison of existing collaboration and new collaboration trends over years. DM - Data Mining domain; MI - Medical Informatics domain; TH - Theory domain; VIS - Visualization domain; DB - Database domain. The trends of cross-domain collaborations in all but one case are growing (The exception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

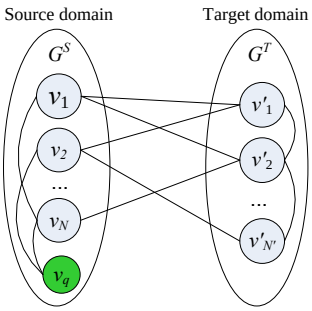
2. PROBLEM DEFINITION

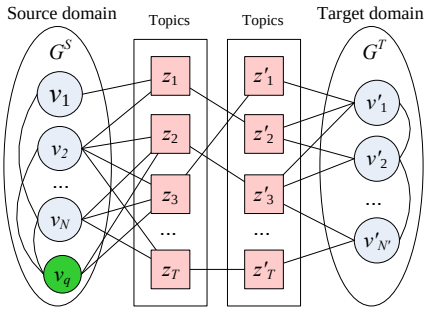
Definition 1. **Source/Target domain.**

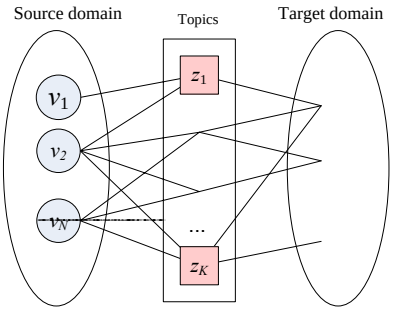
Definition 2. **Domain-specific topic models.**

Problem 1. **Cross-domain collaboration recommendation.**

3. CROSS-DOMAIN TOPIC LEARNING


Source domain
Target domain
G^S
G^T
v_1
v_2
v_N
v_q
v'_1
v'_2
$v'_{N'}$


Source domain
Topics
Topics
Target domain
G^S
G^T
v_1
v_2
v_N
v_q
z_1
z_2
z_3
z_T
z'_1
z'_2
z'_3
z'_T
v'_1
v'_2
$v'_{N'}$


Source domain
Topics
Target domain
v_1
v_2
v_N
z_1
z_K

Input: [illegible] G^S [illegible] G^T
Output: [illegible] $\theta, \theta', \phi, \vartheta$, [illegible] λ

In [illegible] AC[illegible] G^S [illegible] G^S;
[illegible] AC[illegible] G^T;
foreach [illegible] d **do**
  foreach [illegible] $\in d$ **do**
    [illegible] $\sim$ [illegible] (γ_t, γ), [illegible] γ_t [illegible] γ [illegible];
    if [illegible] $= 0$ **then**
      [illegible] G^S [illegible] G^T;
      [illegible] $\sim$ [illegible] $(\vartheta_{vv'})$ [illegible] $\vartheta_{vv'}$ [illegible];
    end
    if [illegible] $= 1$ **then**
      [illegible];
      [illegible] $\sim$ [illegible] (θ_v) [illegible];
    end
  end
  [illegible] $\sim$ [illegible] $(\phi_{z_{di}})$ [illegible];
end

Algorithm 1: [illegible] CTL.

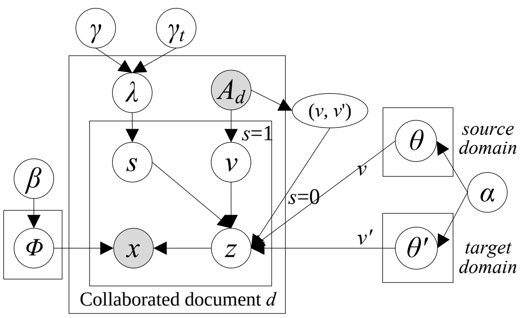


Figure 3: Graphical representation of CTL model.

[illegible] z [illegible] z' [illegible] $P(z|v)$ [illegible] AC[illegible] z' [illegible]

3.3 Cross-domain Topic Learning (CTL)

[illegible] (CTL) [illegible]

Model description. [illegible] 3 [illegible]

Table 1: Notations in the CTL model.

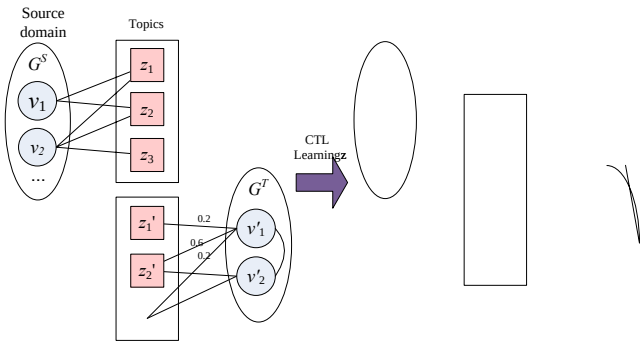
SYMBOL	DESCRIPTION
T	[illegible]
d	[illegible]
A_d	[illegible] d
[illegible]	[illegible] d
[illegible]	[illegible]
[illegible]	[illegible]
θ_v	[illegible]
$\vartheta_{vv'}$	[illegible]
ϕ_z	[illegible]
α, β	[illegible] θ, θ' [illegible] ϕ
λ	[illegible]
γ, γ_t	[illegible] λ

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

[illegible] 1: [illegible] G^S [illegible] G^T [illegible] CTL [illegible] $p(\theta_v|\alpha)$ [illegible] $p(\theta'_{v'}|\alpha)$, [illegible] α [illegible] α [illegible]

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

[illegible] 4 [illegible] CTL [illegible] CTL [illegible]


Source domain
Topics
G^S
v_1
v_2
...
z_1
z_2
z_3
CTL Learning
G^T
z_1'
z_2'
v'_1
v'_2
0.2
0.6
0.2

1,932,442 [illegible] n. T [illegible] w [illegible] n [illegible] n [illegible] n
f [illegible] 1990 2005. [illegible] n [illegible] f w n [illegible] -[illegible] n:

- **Data Mining:** [illegible] f f w n [illegible] n-
  n [illegible] n [illegible] KDD, DM, ICDM, [illegible] DM n KDD
  [illegible] n [illegible], w [illegible] n [illegible] w [illegible] w 6,282 [illegible]
  n 22,862 [illegible] n [illegible].
- **Medical Informatics:** [illegible] n [illegible] f w n [illegible] n:
  J [illegible] n f A [illegible] M [illegible] Inf [illegible] A [illegible] n,
  J [illegible] n f B [illegible] Inf [illegible], A [illegible] In [illegible] n n
  M [illegible] , IEEE T n. M [illegible]. I [illegible] n IEEE T n -
  n n Inf [illegible] n n [illegible] n B [illegible] , f [illegible]
  w w [illegible] n [illegible] w f 9,150 [illegible] n 31,851 -
  [illegible] n [illegible].
- **Theory:** [illegible] n [illegible] f w n [illegible] n [illegible], [illegible]., T C,
  F C n DA, f [illegible] w w [illegible] 5,449 [illegible] n
  27,712 [illegible] n [illegible].
- **Visualization:** [illegible] n [illegible] f w n [illegible] n [illegible] n
  [illegible] (IEEE)2 [illegible] n L

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]	[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
	[illegible]	17.4	19.1	20.0	29.5	5.0	2.4
	[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
	[illegible]	30.4	29.8	31.6	27.4	11.2	5.9
	CTL	**37.7**	**36.4**	**40.6**	**35.6**	**14.3**	**7.5**
[illegible]	[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
	[illegible]	25.0	26.5	28.4	59.1	6.4	4.2
	[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
	[illegible]	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**
[illegible]	[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
	[illegible]	18.0	19.0	19.8	36.7	3.4	1.3
	[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
	[illegible]	21.2	23.8	32.4	48.1	10.2	4.8
	CTL	**30.0**	24.0	**35.6**	49.6	**12.2**	**6.0**
[illegible]	[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
	[illegible]	16.0	20.0	27.6	30.1	6.3	4.4
	[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
	[illegible]	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**

[illegible] (< 80) [illegible] 150. [illegible] CTL [illegible]

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

Restart parameter analysis. [illegible] τ [illegible] 5( ) [illegible] CTL [illegible] τ. In [illegible] τ. B [illegible] τ [illegible] τ [illegible]

Convergence analysis. [illegible] 5( ) [illegible] 10 [illegible] CTL [illegible]

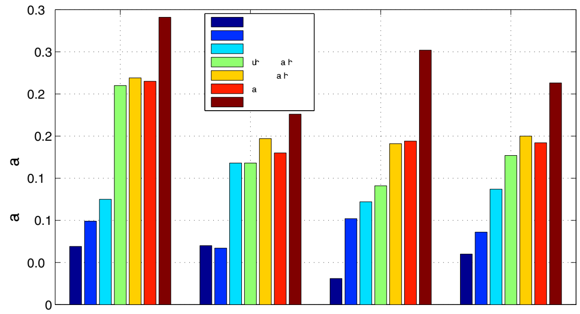


Figure 6: Performance on new collaboration prediction of all algorithms.

(w [illegible] 5 [illegible]). [illegible] CTL [illegible]

New Collaboration Prediction [illegible] 2001 [illegible] 2001-2005 [illegible] 6 [illegible] 2. In [illegible] CTL [illegible] 0.3 [illegible] MA [illegible]

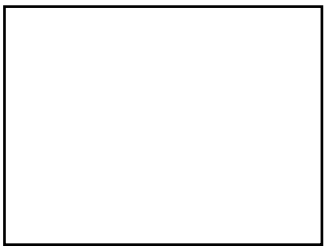
4.3 Prototype System

[illegible] CTL [illegible] [5] [illegible] CTL [illegible] 1,932,442 [illegible] A [illegible] 1 [illegible] (Cf. [illegible] 3.3) [illegible] CTL [illegible]

5. RELATED WORK

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

5 [illegible]



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

23 D. M ... n ... n A. M C ... . E ... n f ... n ... w ... w ... . In *KDD'07*, ... 500 509, 2007.

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

25 D. ... C ... n ... n ... n. In *KDD'10*, ... 979 988, 2010.

26 ... , D. ... , A. G ... n ... n n. A ... n ... f ... n n ... n n ... w ... . In *SIGIR'11*, ... 1337 1338, 2011.

27 M. ... , ... n ... G. f ... f ... nf ... n ... . In *KDD'04*, ... 306 315, 2004.

28 J. ... n, H. ... , D. C ... n C. E ... f ... n n ... n ... n n ... . In *ICDM'05*, ... 418 425, 2005.

29 J. ... n, L. ... n D. C ... n. M ... - ... - ... n ... n. In *SDM'09*, ... 1147 1158, 2009.

30 J. ... n, J. ... n, ... J. n, ... n, K. C ... , L. ... n ... n ... w ... . *Mac ... L a ... J ... a*, 82(2):211 237, 2011.

31 J. ... n, J. ... n, L. ... , J. L, L. ... n ... n ... . A ... E ... n n ... n n ... f ... w ... . In *KDD'08*, ... 990 998, 2008.

32 L. ... n ... n H. L ... n ... n ... n ... n ... n. In *KDD'09*, ... 817 826, 2009.

33 ... n, J. ... n, ... , C. ... n, B. G ... n ... L. ... n ... n f ... n w ... n ... n. *N ... c ...*, 76(1):71 83, 2012.

34 C. ... n ... n D. M. B ... . C ... n f ... n n ... n ... . In *KDD'11*, ... 448 456, 2011.

35 ... n, L. C ... n ... n ... . E ... n ... n: f ... n ... n ... n ... -n ... n ... n. In *R cS ... '11*, ... 245 252, 2011.

36 J. ... n, J. ... n ... n J. L. E ... n n ... n ... w ... . In *DASFAA'07*, ... 1066 1069, 2007.

8. APPENDIX

A ... n ... , w ... n ... n ... (B ... n ...) ... n $\theta, \theta', \vartheta, \lambda, \varphi$, ...