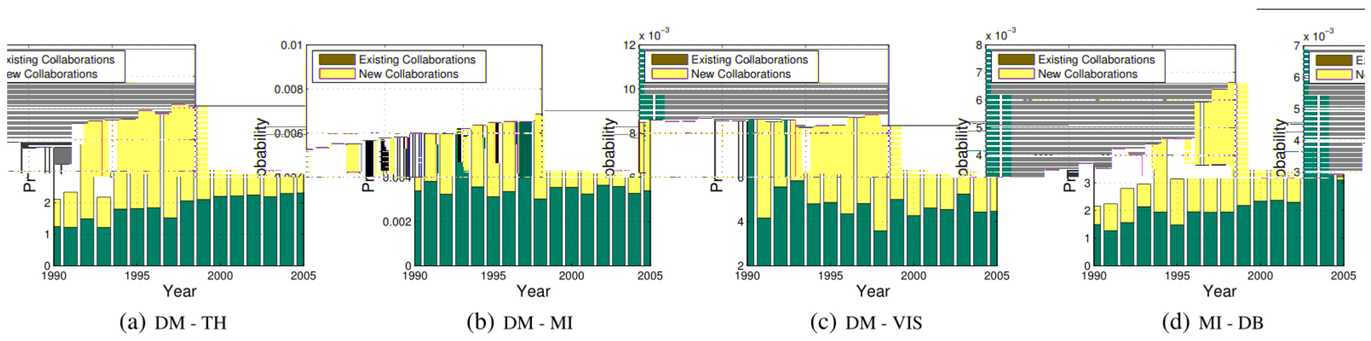
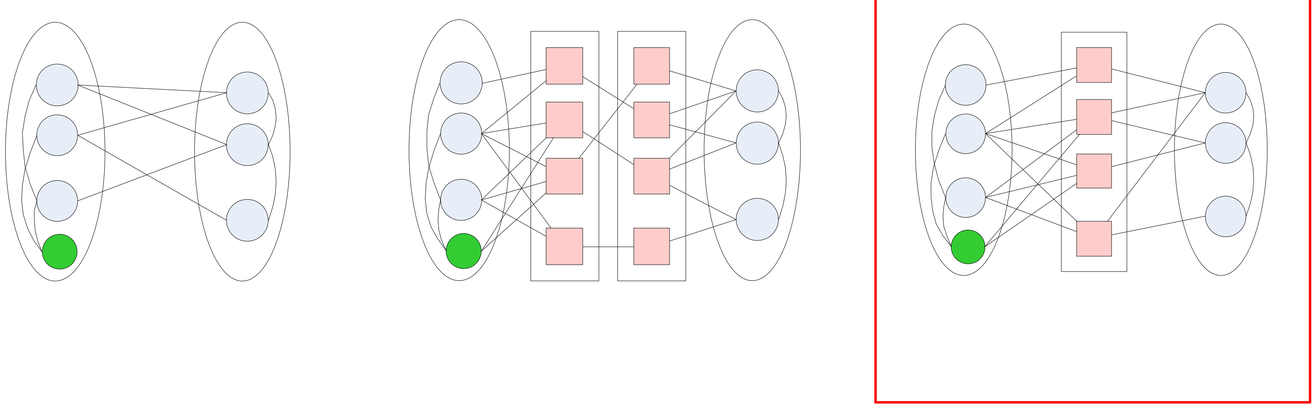
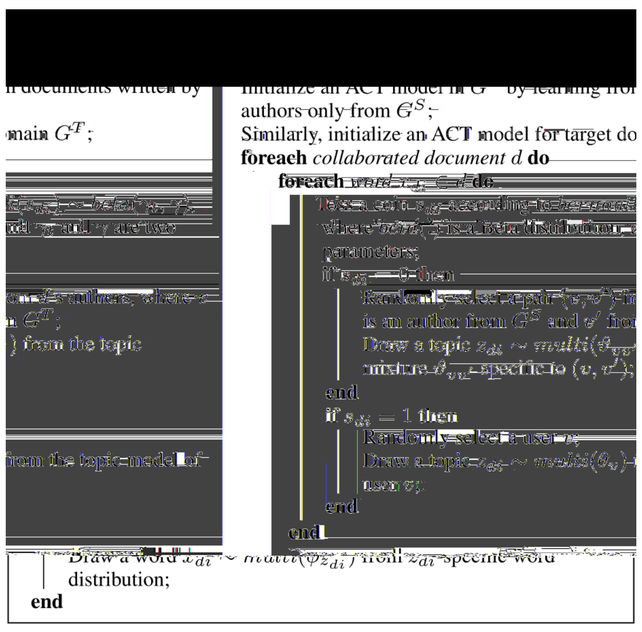
THE 18TH ACM SIGKDD CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING
KDD 2012 BEIJING
Beijing, China
August 12-16, 2012

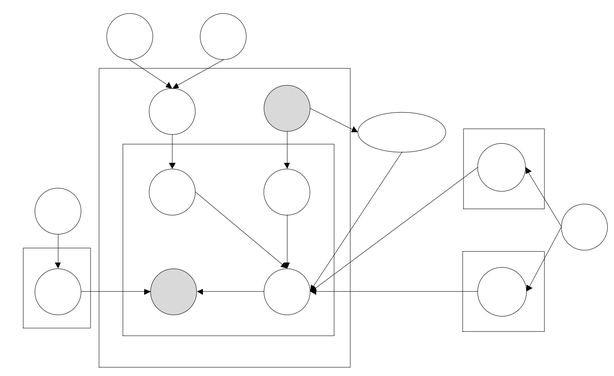


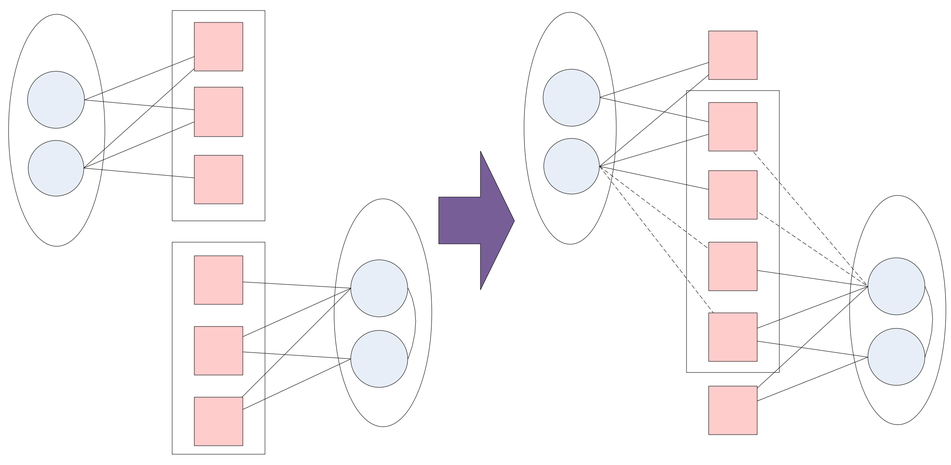
Cross-domain Topic Learning (CTL)



SYMBOL	DESCRIPTIO
T	number of topi
d	a collaborated
A_d	a set of authors
x_{di}	the ith attribute (word) in do
z_{di}	the topic assigned to attribute x_{di}
s_{di}	if x_{di} is a word from a single domain or
θ_v	multinomial distribution over topics specific
$\vartheta_{vv'}$	multinomial distribution over topics specif pair (v,v')
ϕ_z	multinomial distribution over words specific to topic z
α, β	Dirichlet priors to multinomial distributions θ, θ' and ϕ
λ	parameter for sampling the binary variable s
γ, γ_t	Beta parameters to generate λ



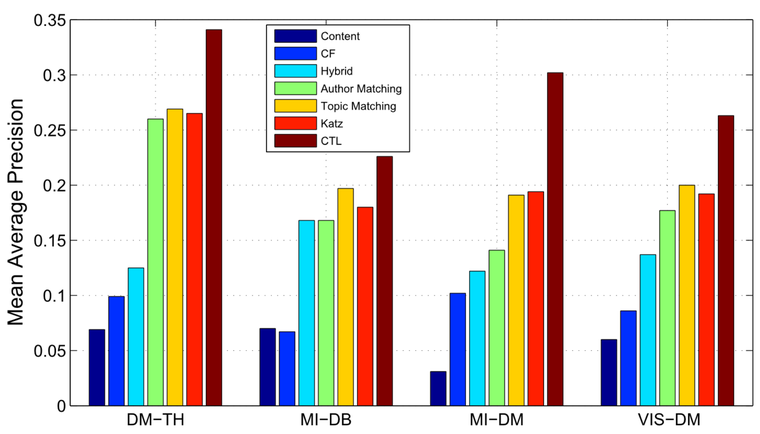




Empirical Analysis

Data Mining(DM)—
Medical Informatics(MI)—
Theory(TH)
Visualization(VIS)
Database(DB)

Baselines:



	P@10	P@20	MAP	R@100	-10	-20
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**
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	28.1	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	25.8	32.4	48.1	10.2	4.8
30.0	24.0	**35.6**	49.6	**12.2**	**6.0**	
9.6	11.8	13.2	18.9	3.1	1.8	
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**