APPLICATIONS OF MINING HETEROGENEOUS INFORMATION NETWORKS

Yizhou Sun

College of Computer and Information Science

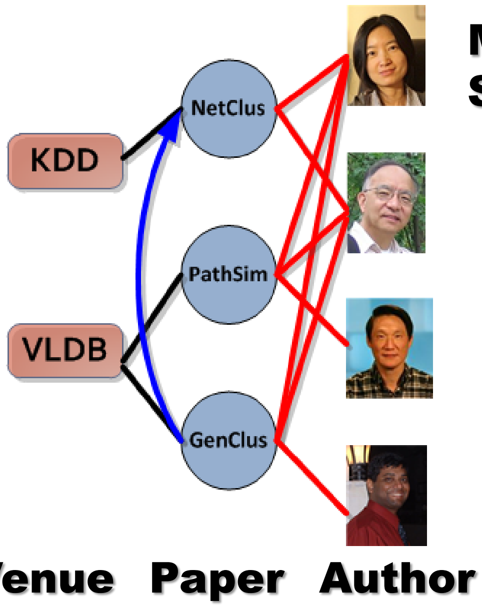
Northeastern University

yzsun@ccs.neu.edu

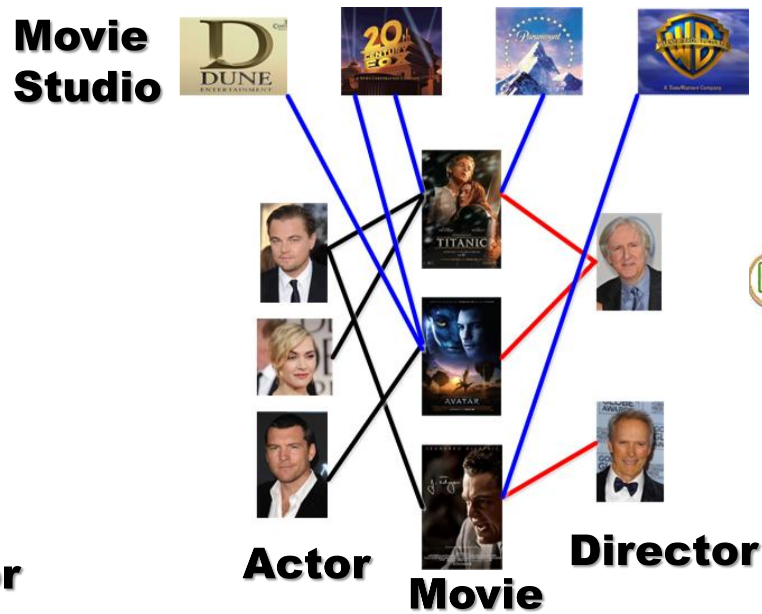
July 25, 2015

Heterogeneous Information Networks

- Multiple object types and/or multiple link types



DBLP Bibliographic Network



The IMDB Movie Network


The Facebook Network

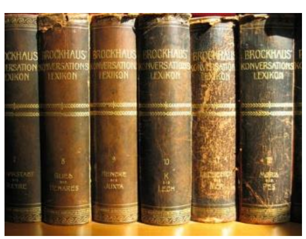
1. Homogeneous networks are ***Information loss*** projection of heterogeneous networks!
2. ***New problems*** are emerging in heterogeneous networks!

Directly Mining information richer heterogeneous networks

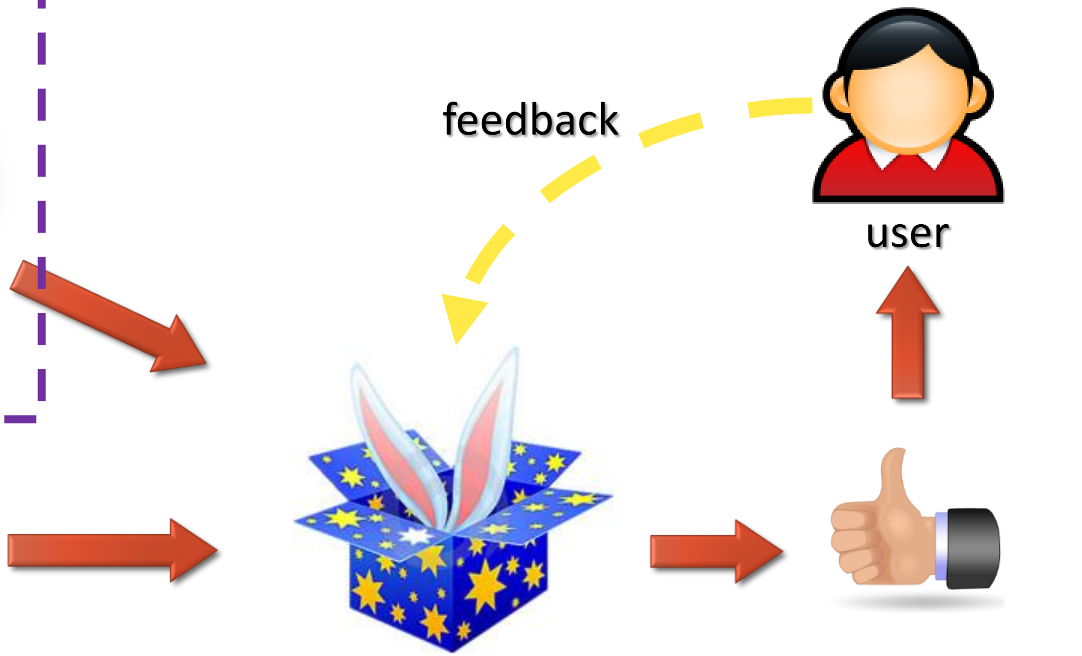
Outline

- Why Heterogeneous Information Networks?
- Entity Recommendation
- Information Diffusion
- Ideology Detection
- Summary

Recommendation Paradigm



external knowledge

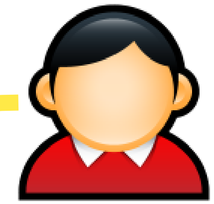


Collaborative Filtering

E.g., K-Nearest Neighbor (Sarwar WWW'01), Matrix Factorization (Hu ICDM'08, Koren IEEE CS'09), Probabilistic Model (Hofmann SIGIR'03)

Problem Definition

implicit user feedback



user

feedback

recommender system

recommendation

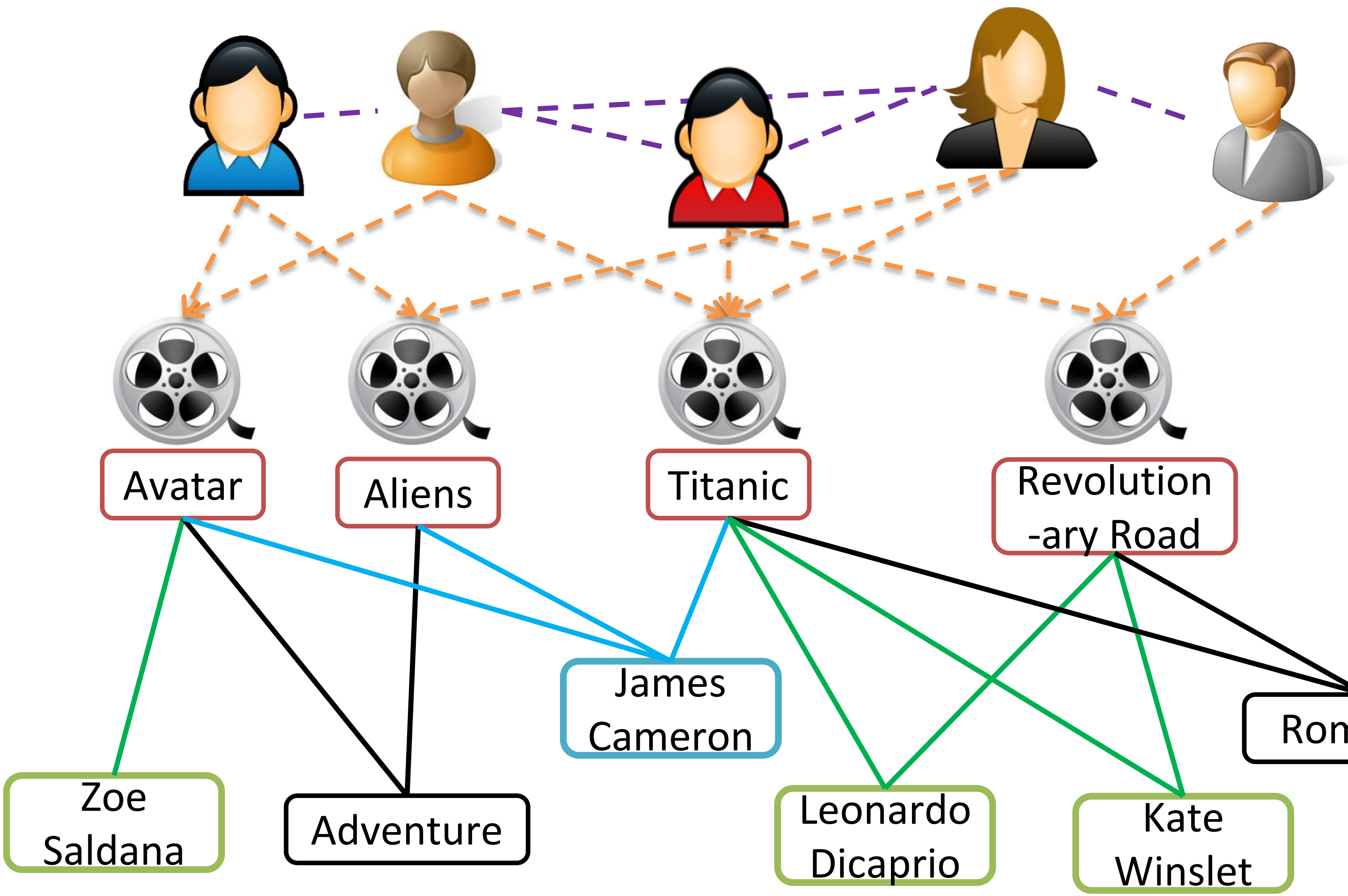
information network

hybrid collaborative filtering with information networks

Hybrid Collaborative Filtering with Networks

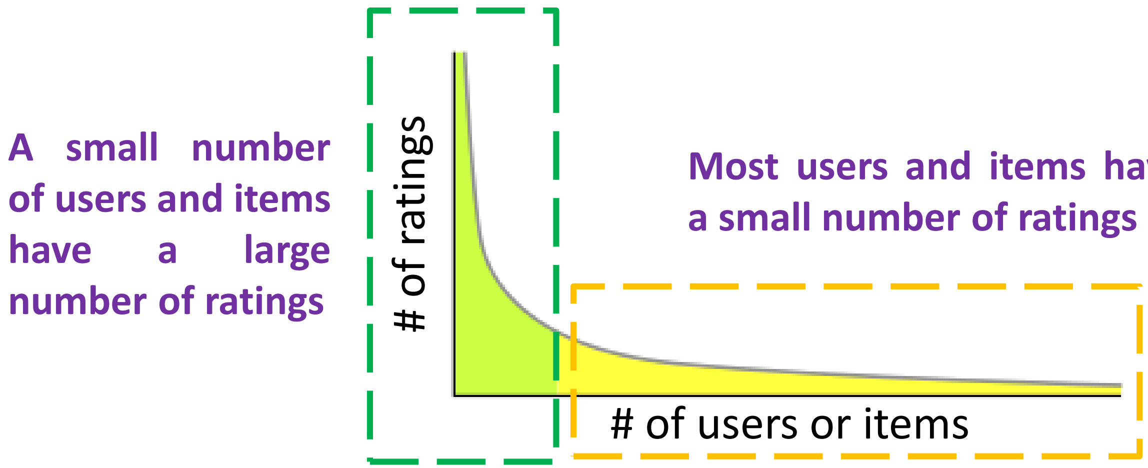
- Utilizing network relationship information can enhance the recommendation quality
- However, most of the previous studies only use single type of relationship between users or items (e.g., social network Ma,WSDM'11, trust relationship Ester, KDD'10, service membership Yuan, RecSys'11)

The Heterogeneous Information Network View of Recommender System



Relationship Heterogeneity Alleviates Data Sparsity

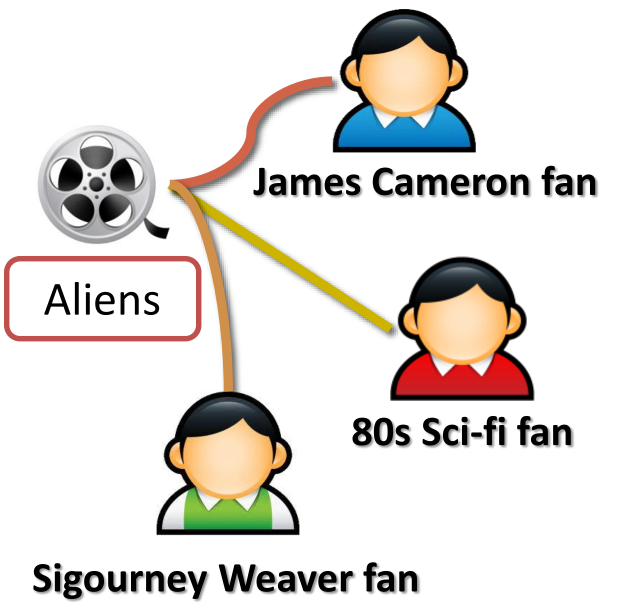
Collaborative filtering methods suffer from data sparsity issue



- Heterogeneous relationships complement each other
- Users and items with limited feedback can be connected to the network by different types of paths
  - Connect new users or items (cold start) in the information network

Relationship Heterogeneity Based Personalized Recommendation Models

Different users may have different behaviors or preferences



Different users may be interested in the same movie for different reasons

Two levels of personalization

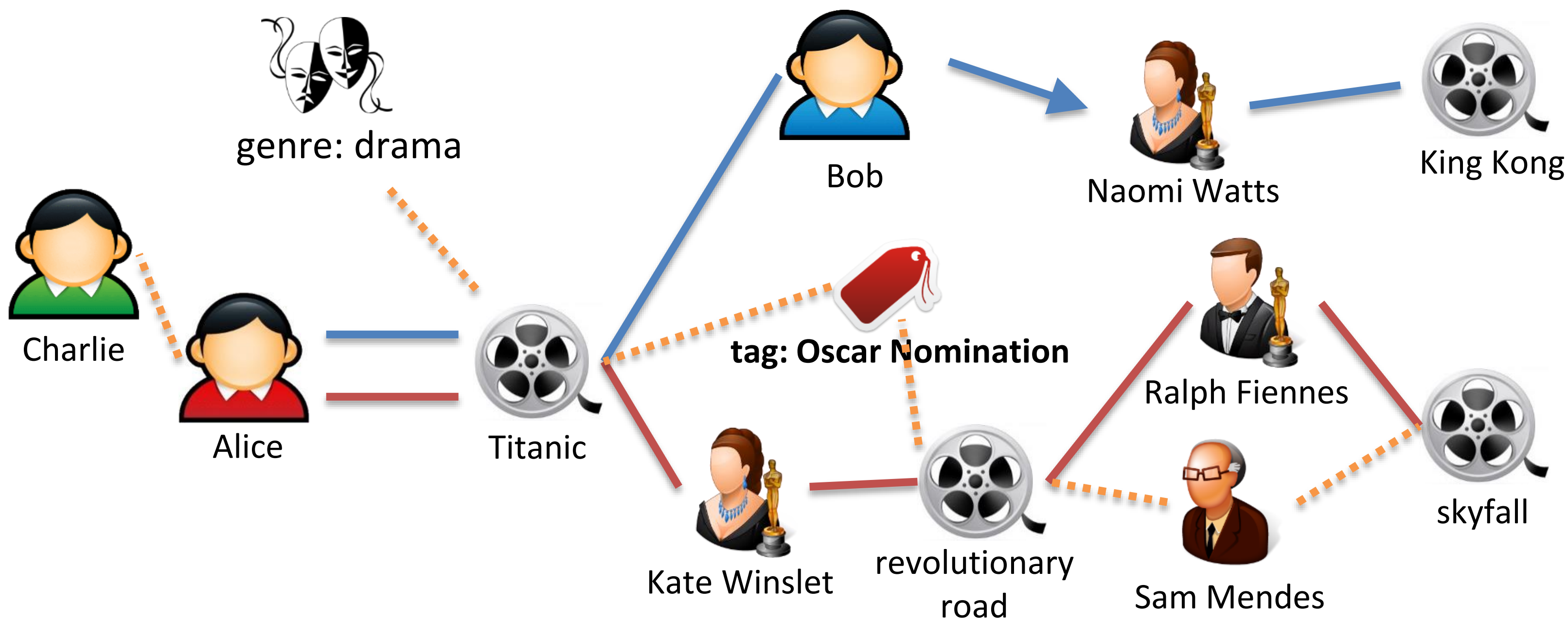
Data level

- Most recommendation methods use one model for all users and rely on personal feedback to achieve personalization

Model level

- With different entity relationships, we can learn personalized models for different users to further distinguish their differences

Preference Propagation-Based Latent Features



Generate *L* different **meta-path** (path types) connecting users and items

Propagate user implicit feedback along each meta-path

Calculate latent-features for users and items for each meta-path with **NMF** related method

Recommendation Models

Observation 1: Different meta-paths may have different importance

Global Recommendation Model

ranking score

features for user *i* and item *j*

$$\boxed{\hat{r}(u_i, e_j)} = \sum_{q=1}^{L} \theta_q \cdot \boxed{\hat{U}_i^{(q)} \hat{V}_j^{(q)T}} \qquad (1)$$

the q-th meta-path

Observation 2: Different users may require different models

Personalized Recommendation Model

user-cluster similarity

$$\hat{r}_p(u_i, e_j) = \sum_{k=1}^{c} \boxed{sim(C_k, u_i)} \sum_{q=1}^{L} \theta_q^{\{k\}} \cdot \hat{U}_i^{(q)} \hat{V}_j^{(q)T} \qquad (2)$$

c total soft user clusters

Parameter Estimation

- Bayesian personalized ranking (Rendle UAI'09)
- Objective function

sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$.

$$\min_{\Theta} \; -\sum_{u_i \in \mathcal{U}} \boxed{\sum_{(e_a > e_b) \in R_i}} \ln \sigma(\hat{r}(u_i, e_a) - \hat{r}(u_i, e_b)) + \frac{\lambda}{2}\|\Theta\|_2^2 \qquad (3)$$

for each correctly ranked item pair
i.e., u_i gave feedback to e_a but not e_b

Soft cluster users with NMF + k-means → **For each user cluster, learn one model with Eq. (3)** → **Generate personalized model for each user on the fly with Eq. (2)**

Learning Personalized Recommendation Model

Experiment Setup

- Datasets

Name	#Items	#Users	#Ratings	#Entities	#Links
IM100K	943	1360	89,626	60,905	146,013
Yelp	11,537	43,873	229,907	285,317	570,634

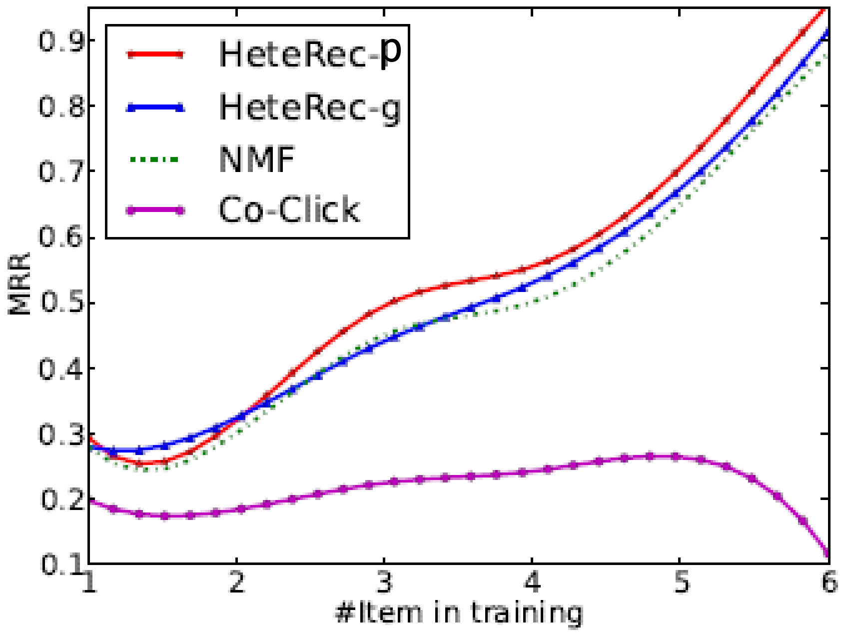
- Comparison methods:
  - Popularity: recommend the most popular items to users
  - Co-click: conditional probabilities between items
  - NMF: non-negative matrix factorization on user feedback
  - Hybrid-SVM: use Rank-SVM with plain features (utilize both user feedback and information network)

Performance Comparison

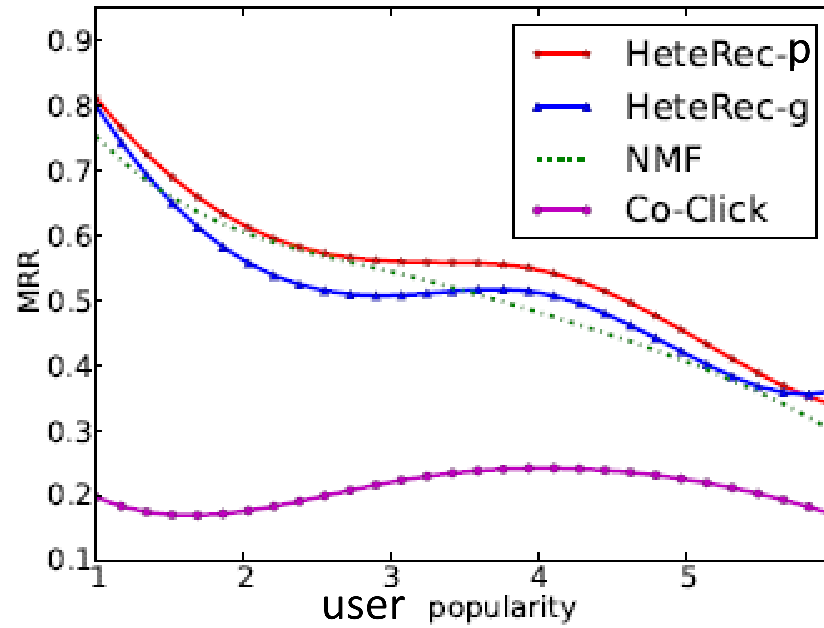
Method	IM100K				Yelp			
	Prec1	Prec5	Prec10	MRR	Prec1	Prec5	Prec10	MRR
Popularity	0.0731	0.0513	0.0489	0.1923	0.00747	0.00825	0.00780	0.0228
Co-Click	0.0668	0.0558	0.0538	0.2041	0.0147	0.0126	0.01132	0.0371
NMF	0.2064	0.1661	0.1491	0.4938	0.0162	0.0131	0.0110	0.0382
Hybrid-SVM	0.2087	0.1441	0.1241	0.4493	0.0122	0.0121	0.0110	0.0337
HeteRec-g	0.2094	0.1791	0.1614	0.5249	0.0165	0.0144	0.0129	0.0422
HeteRec-l	**0.2121**	**0.1932**	**0.1681**	**0.5530**	**0.0213**	**0.0171**	**0.0150**	**0.0513**

HeteRec personalized recommendation (HeteRec-p) provides the best recommendation results

Performance under Different Scenarios



(a) Performance Change with User Feedback Number



(b) Performance Change with User Feedback Popularity

HeteRec–p consistently outperform other methods in different scenarios

better recommendation results if users provide more feedback

better recommendation for users who like less popular items

Contributions

Entity Recommendation in Information Networks with Implicit User Feedback

(RecSys'13, WSDM'14a)

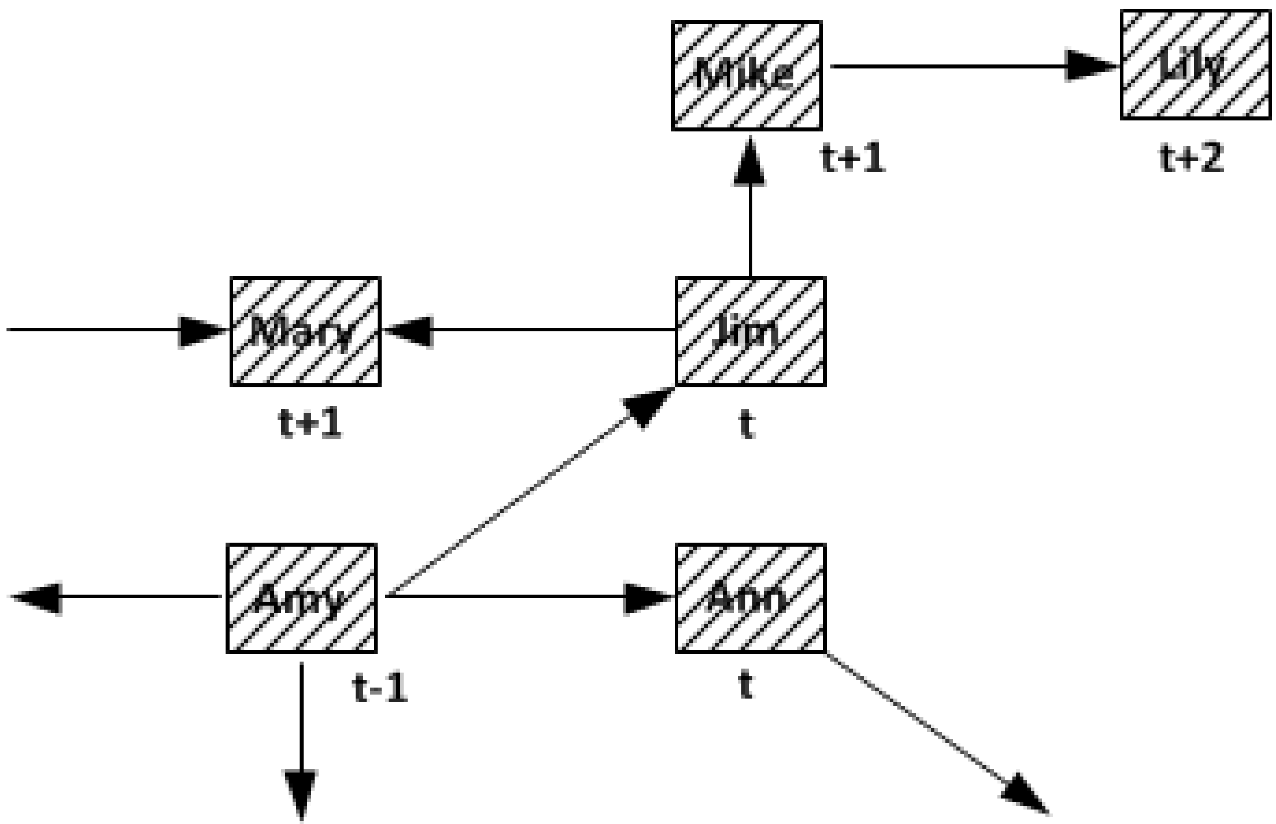
- Propose latent representations for users and items by propagating user preferences along different meta-paths
- Employ Bayesian ranking optimization technique to correctly evaluate recommendation models
- Further improve recommendation quality by considering user differences at model level and define personalized recommendation models
  - Two levels of personalization

Outline

- Why Heterogeneous Information Networks?
- Entity Recommendation
- Information Diffusion
- Ideology Detection
- Summary

Information Diffusion in Networks

- Action of a node is triggered by the actions of their neighbors



Linear Threshold Model

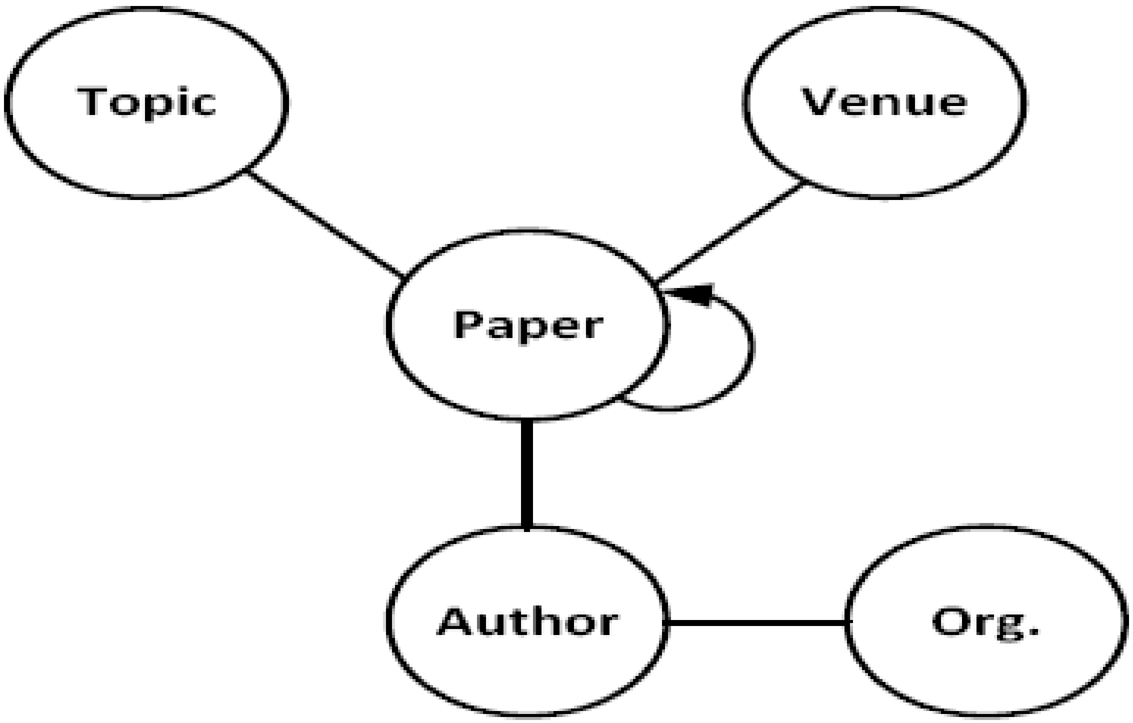
- [Granovetter, 1978]
  - If the weighted activation number of its neighbors is bigger than a pre-specified threshold θ_u, the node u is going to be activated

$$\sum_{v \in \Gamma(u)} w_{v,u}\delta(u,t) > \theta_u$$

- In other words
  - $p_u(t+1) = E[1(\sum_{v \in \Gamma(u)} w_{v,u}\delta(u,t) > \theta_u)]$

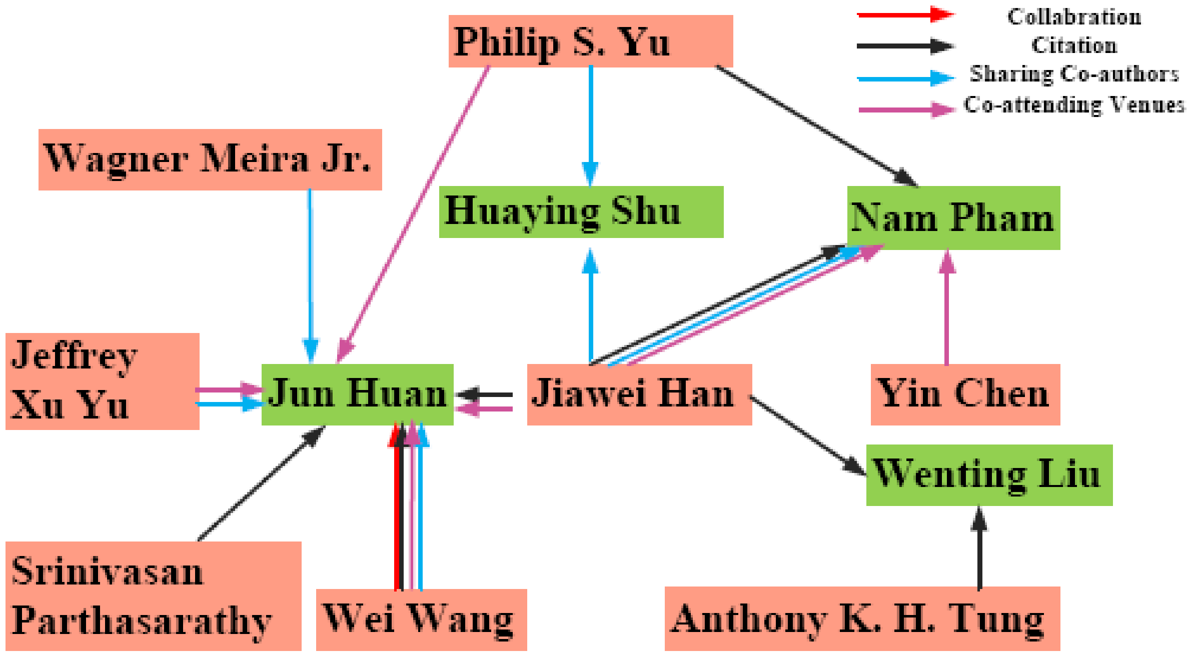
Heterogeneous Bibliographic Network

- Multiple types of objects
- Multiple types of links



Derived Multi-Relational Bibliographic Network

- Collaboration: Author-Paper-Author
- Citation: Author-Paper->Paper-Author
- Sharing Co-authors: Author-Paper-Author-Paper-Author
- Co-attending venues: Author-Paper-Venue-Paper-Author



How to generate these meta-paths ?
PathSim: Sun et.al, VLDB'11

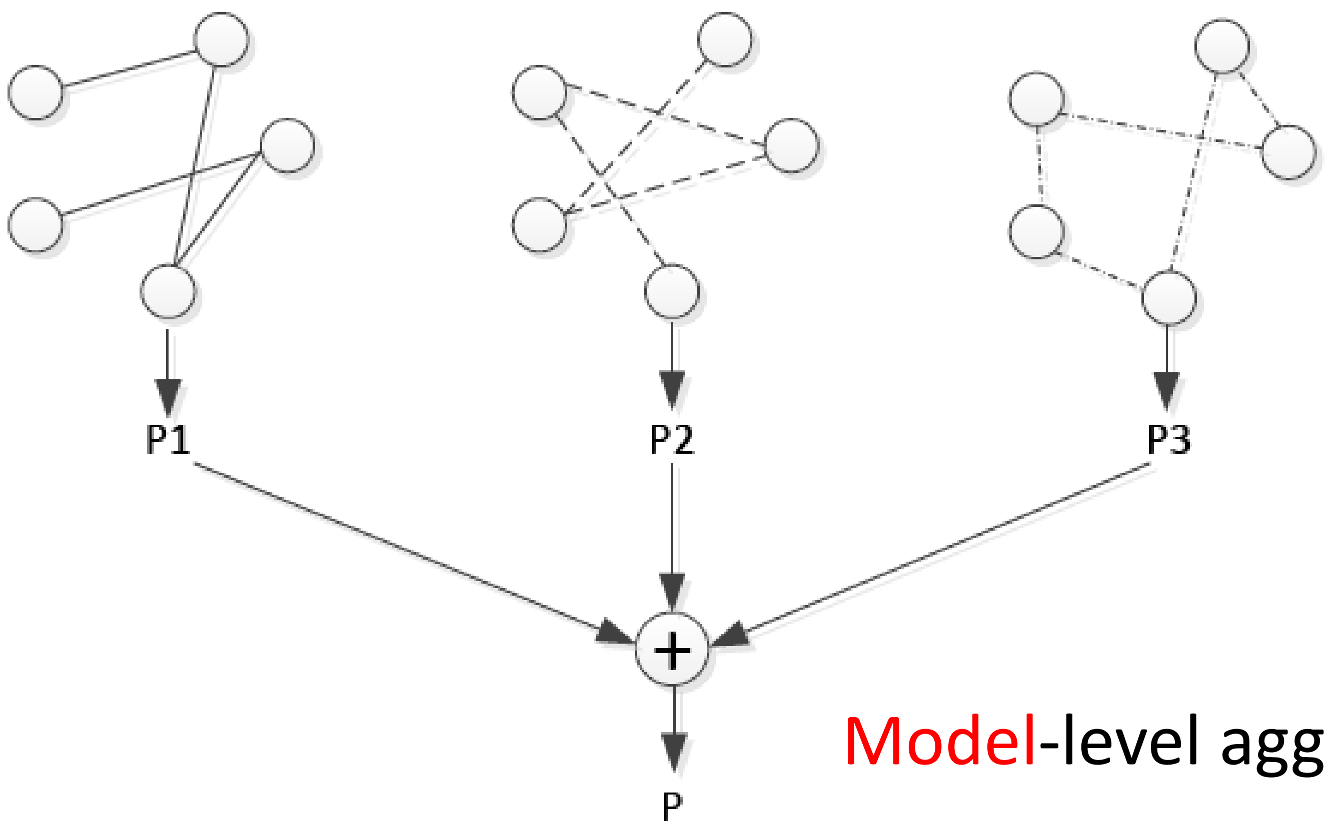
How Topics Are Propagated among Authors?

- To Apply Existing approaches
  - Select one relation between authors (say, A-P-A)
  - Use all the relations, but ignore the relation types
- Do different relation types play different roles?
  - Need new models!

Information loss!

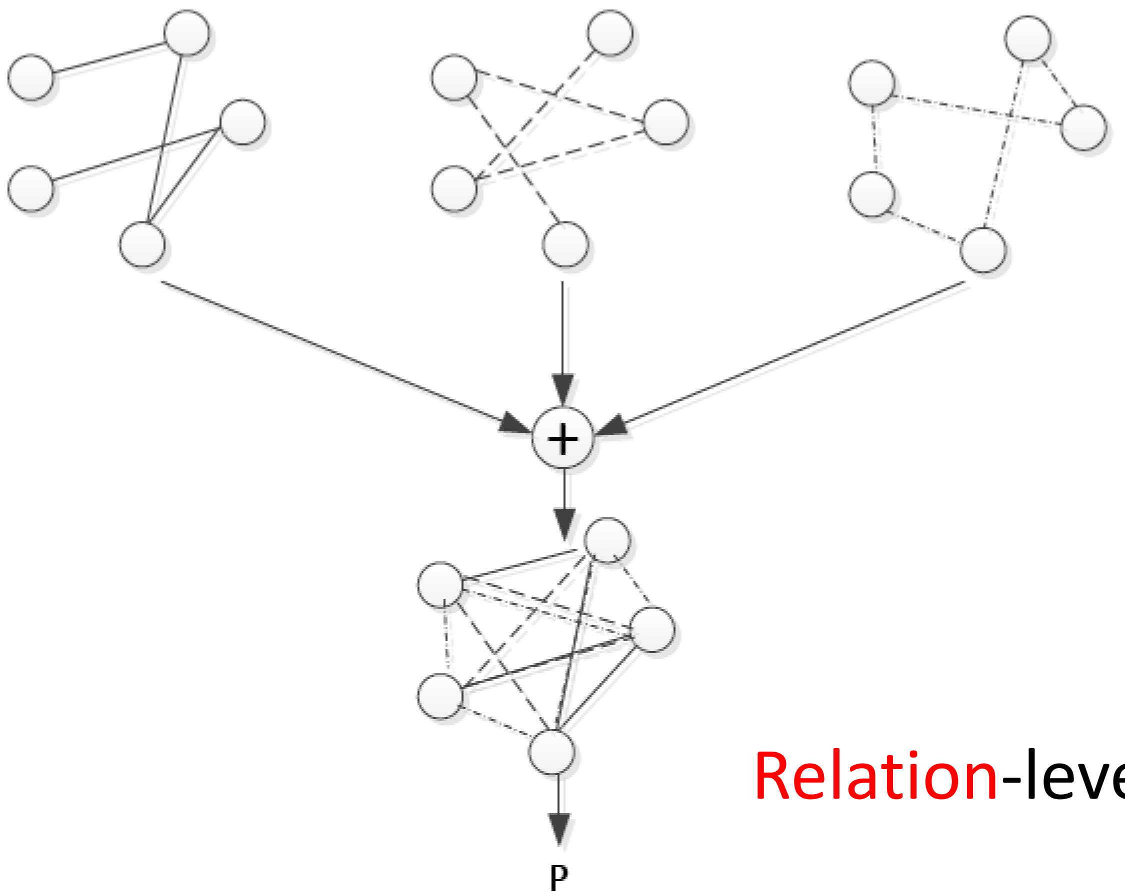
Two Assumptions for Topic Diffusion in Multi-Relational Networks

- Assumption 1: Relation independent diffusion



Model-level aggregation

- Assumption 2: Relation interdependent diffusion



Relation-level aggregation

Two Models under the Two Assumptions

- Two multi-relational linear threshold models
  - Model 1: MLTM-M
    - Model-level aggregation
  - Model 2: MLTM-R
    - Relation-level aggregation

MLTM-M

- For each relation type k
  - The activation probability for object i at time t+1:

- The collective model
  - The final activation probability for object i is an aggregation over all relation types

$$p_i^{(k)}(t+1) = \frac{\sum_{j \in \Gamma(i,k)} w_{ij}^{(k)} \delta(j,t)}{\sum_{j \in \Gamma(i,k)} w_{ij}^{(k)}}$$

$$p_i(t+1) = \frac{e^{\sum_k \beta_k p_i^{(k)}(t+1) + \beta_0}}{1 + e^{\sum_k \beta_k p_i^{(k)}(t+1) + \beta_0}}$$

Properties of MLTM-M

PROPERTY 1. *Let $p_i^{(-k)}(t+1)$ be the activation probability of i at timestamp $t+1$ without relation type k, under MLTM-M, we have*

- *if $\beta_k > 0$, then $p_i(t+1) > p_i^{(-k)}(t+1)$, that is, a relation type with positive diffusion power will increase the activation probability of i;*
- *if $\beta_k = 0$, then $p_i(t+1) = p_i^{(-k)}(t+1)$, that is, a relation type with no diffusion power (such as noise) will not change the activation probability of i; and*
- *if $\beta_k < 0$, then $p_i(t+1) < p_i^{(-k)}(t+1)$, that is, a relation type with negative diffusion power will decrease the activation probability of i.*

MLTM-R

- Aggregate multi-relational network with different weights
  - Treat the activation as in a single-relational network
  -

$$p_i(t+1) = \frac{\sum_k \beta_k \sum_{j \in \Gamma(i,k)} w_{ij}^{(k)} \delta(j,t) + \beta_0 \sum_j \delta(j,t)}{\sum_k \beta_k \sum_{j \in \Gamma(i,k)} w_{ij}^{(k)} + \beta_0 N}$$

To make sure the activation probability non-negative, weights $\beta' s$ are required non-negative

Properties of MLTM-R

PROPERTY 2. *Let* $p_i^{(-k)}(t+1)$ *be the activation probability of* i *at timestamp* $t+1$ *without relation type* k*, under MLTM-R, we have*

- $\min(p_i^{(k)}(t+1), p_i^{(-k)}(t+1)) \leq p_i(t+1) \leq \max(p_i^{(k)}(t+1), p_i^{(-k)}(t+1))$.
- *when* $\beta_k \to 0, p_i(t+1) \to p_i^{(-k)}(t+1)$*; and*
- *when* $\beta_k \to \infty, p_i(t+1) \to p_i^{(k)}(t+1)$.

How to Evaluate the Two Models?

- Test on the real action log on multiple topics!
  - $Action\ log: \{< u_i, t_i >\}$
- Diffusion model learning from action log
  - MLE estimation over $\beta' s$

Two Real Datasets

- DBLP
  - Computer Science
  - Relation types
    - APA, AP->PA, APAPA, APVPA

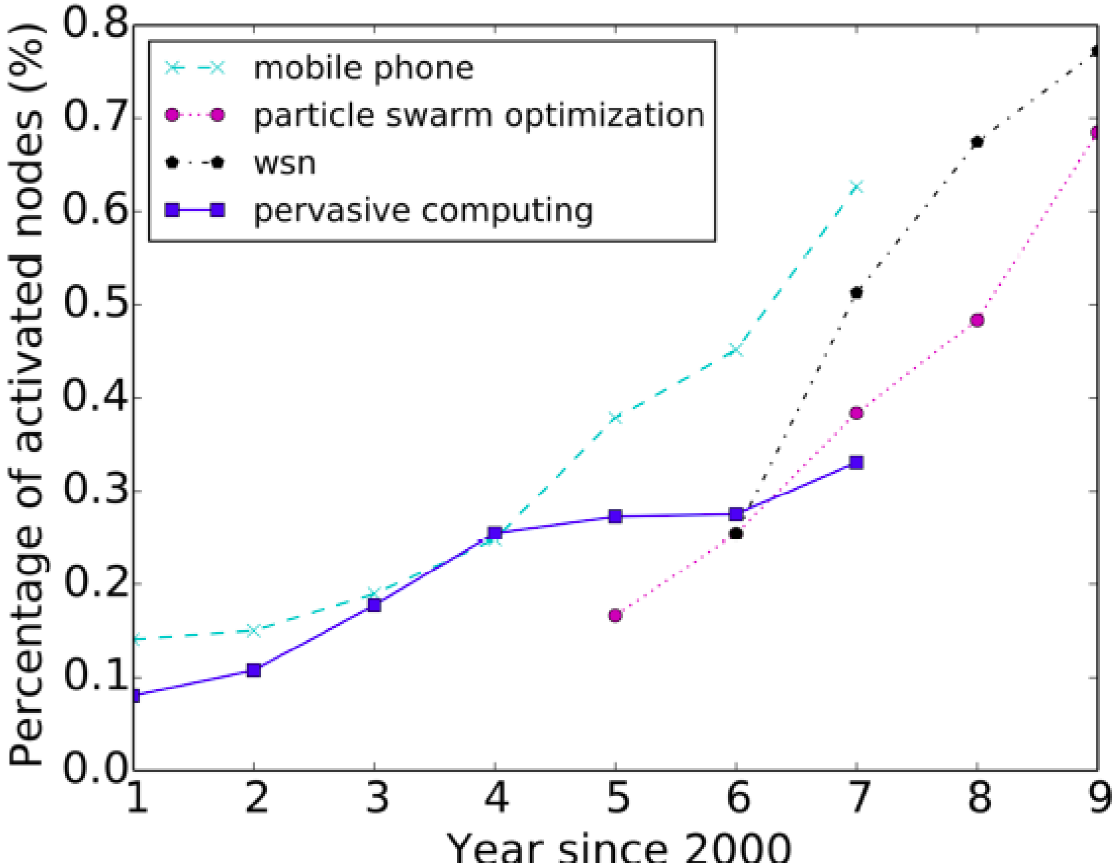
Statistics	Count
Authors	916,988
Papers	1,572,278
Venues	6,713
Author-Paper	4,135,188
Citations	2,083,947

- APS
  - Physics
  - Relation types
    - APA, AP->PA, APAPA, APOPA

Statistics	Count
Authors	323,675
Papers	463,348
Organizations [3]	41,411
Author-Paper	2,471,474
Citations	4,710,547

Topics Selected

- Select topics with increasing trends



Evaluation Methods

- Global Prediction
  - How many authors are activated at t+1
  - Error rate = ½(predicted#/true# + true#/predicted#)-1

- Local Prediction
  - Which author is likely to be activated at t+1
  - AUPR (Area under Precision-Recall Curve)

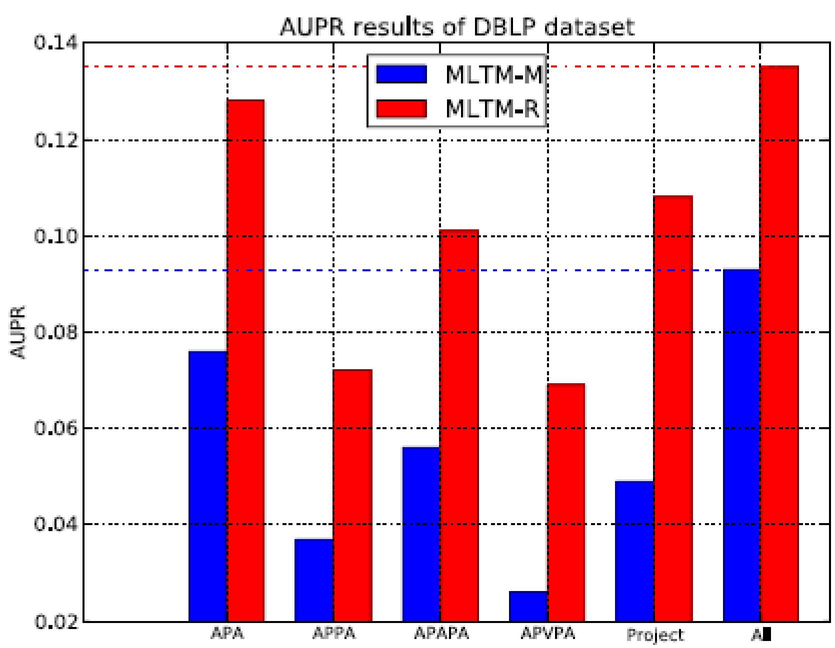
Global Prediction

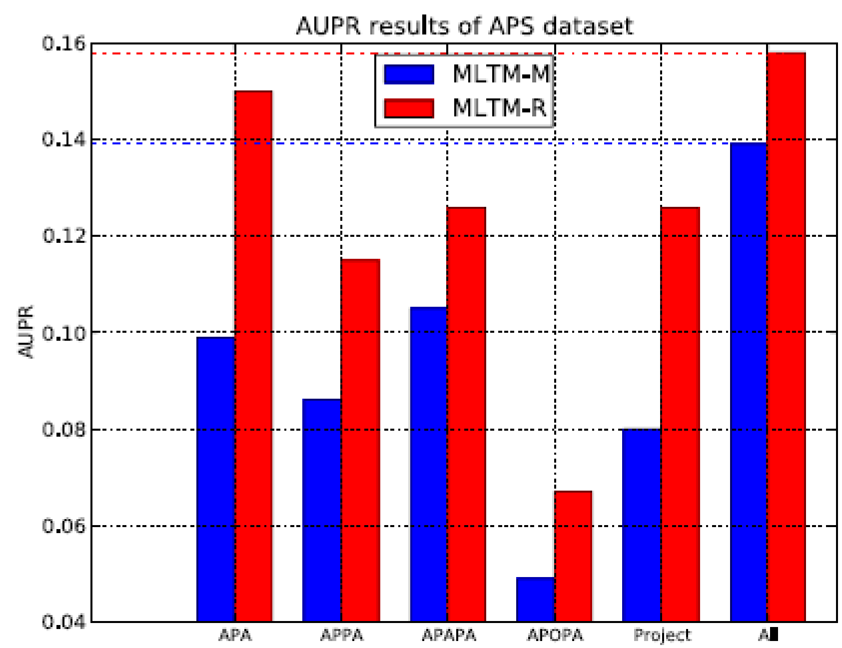
Dataset	Model	APA	APPA	APAPA	APVAP	Full
DBLP	Homo-	0.654	0.287	1.005	1.269	N/A
	MH-	**0.033**	**0.04**	**0.034**	**0.041**	**0.031**
	RH-	0.072	0.07	0.092	0.128	0.125

Dataset	Model	APA	APPA	APAPA	APOAP	Full
APS	Homo-	0.249	0.398	0.107	0.144	N/A
	MH-	**0.045**	**0.052**	**0.039**	**0.068**	**0.052**
	RH-	0.073	0.082	0.076	0.079	0.11

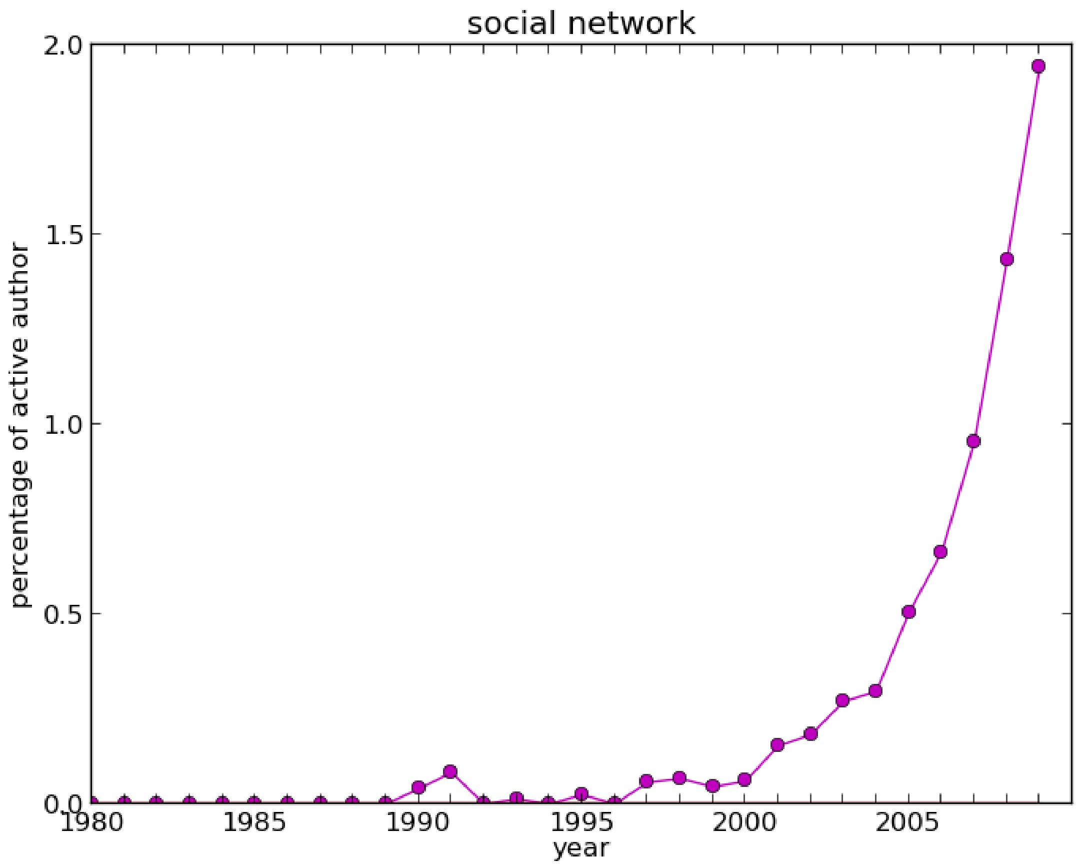
Local Prediction - AUPR

- 1: Different Relation Play Different Roles in Diffusion Process
- 2: Relation-Level Aggregation is better than Model-Level Aggregation





Case Study


social network
percentage of active author
year
2.0
1.5
1.0
0.5
0.0
1980
1985
1990
1995
2000
2005

Prediction Results on "social network" Diffusion

AUPR					
Year	MLTM-M	APA	APPA	APAPA	APVPA
2006	**0.0708**	0.0538	0.0288	0.0382	0.012
2007	**0.0696**	0.0596	0.0293	0.0442	0.0187
2008	**0.0861**	0.0767	0.0394	0.0549	0.0231
held-out logL					
Year	MLTM-M	APA	APPA	APAPA	APVPA
2006	**-3436.8302**	-3529.7381	-3750.592	-3605.3715	-3899.8318
2007	**-5723.0076**	-5881.2261	-6248.871	-5973.8478	-6368.6163
2008	**-7663.0198**	-7933.8971	-8506.7911	-8088.0607	-8990.3446
global					
Year	MLTM-M	APA	APPA	APAPA	APVPA
2006	**0.0366**	0.0413	0.0477	0.0471	0.0547
2007	**0.0337**	0.0574	0.0942	0.0587	0.091
2008	0.0022	0.0169	0.0295	0.0116	**0.0008**

AUPR					
Year	MLTM-R	APA	APPA	APAPA	APVPA
2006	**0.0996**	0.0865	0.0826	0.0695	0.0396
2007	**0.1157**	0.1026	0.0912	0.0802	0.0608
2008	**0.1498**	0.1433	0.1088	0.1068	0.0492
held-out logL					
Year	MLTM-R	APA	APPA	APAPA	APVPA
2006	**-3154.8047**	-3349.3748	-3367.2633	-3391.4317	-3457.8147
2007	**-5182.5405**	-5451.7943	-5527.3942	-5499.6015	-5701.3318
2008	**-7151.585**3	-7306.9863	-7597.4606	-7503.8617	-8137.539
global					
Year	MLTM-R	APA	APPA	APAPA	APVPA
2006	0.0227	0.0052	**0.0018**	0.0024	0.0196
2007	0.0128	0.0112	0.0005	**0.0001**	0.0091
2008	0.0529	**0.002**	0.0232	0.0326	0.0802

Year	MLTM-M
2006	**0.0708**
2007	**0.0696**
2008	**0.0861**

Year	MLTM-M
2006	**-3436.8302**
2007	**-5723.0076**
2008	**-7663.0198**

Year	MLTM-M
2006	**0.0366**
2007	**0.0337**
2008	0.0022

WIN!

Year	MLTM-R
2006	**0.0996**
2007	**0.1157**
2008	**0.1498**

Year	MLTM-R
2006	**-3154.8047**
2007	**-5182.5405**
2008	**-7151.5853**

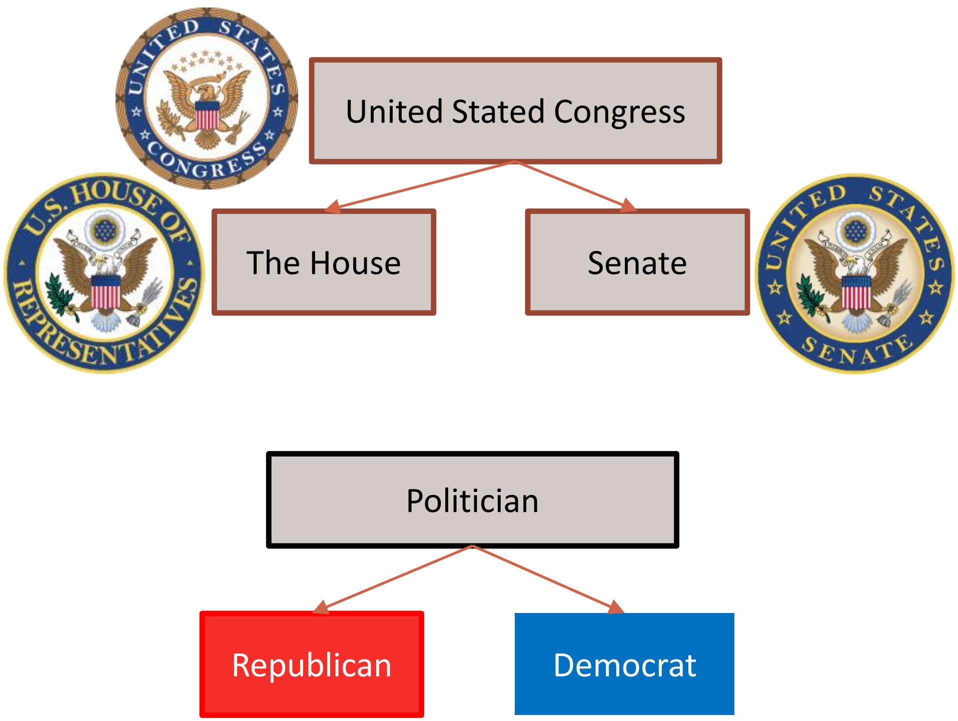
Year	MLTM-R
2006	0.0227
2007	0.0128
2008	0.0529

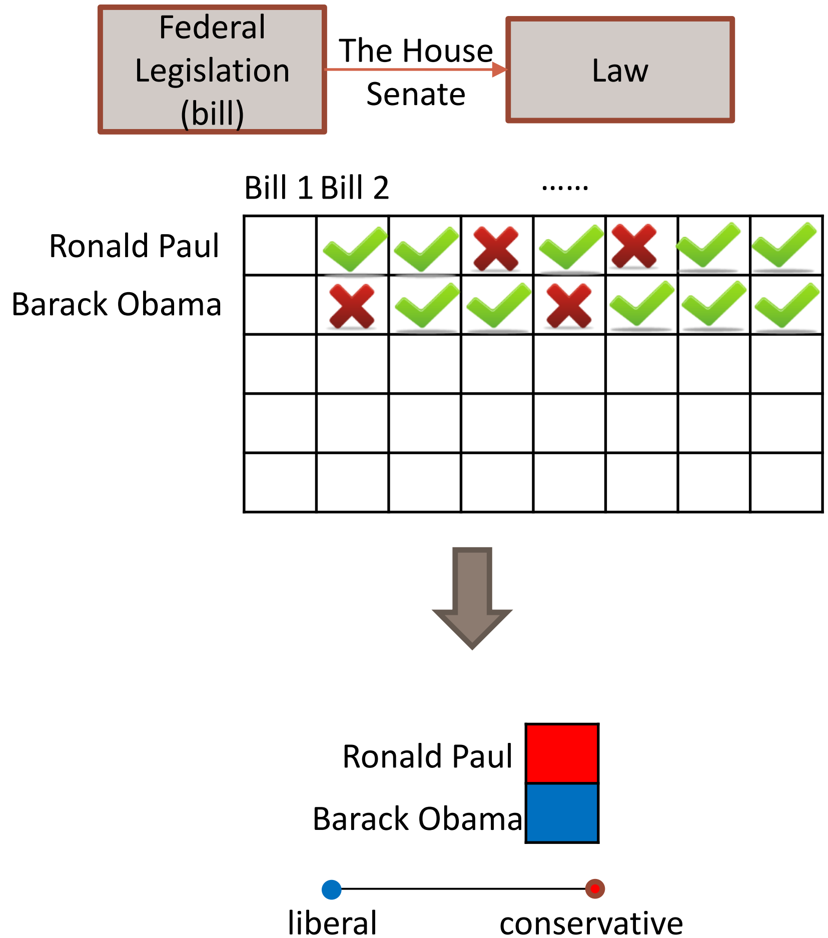
Outline

- Why Heterogeneous Information Networks?
- Entity Recommendation
- Information Diffusion
- Ideology Detection
- Summary

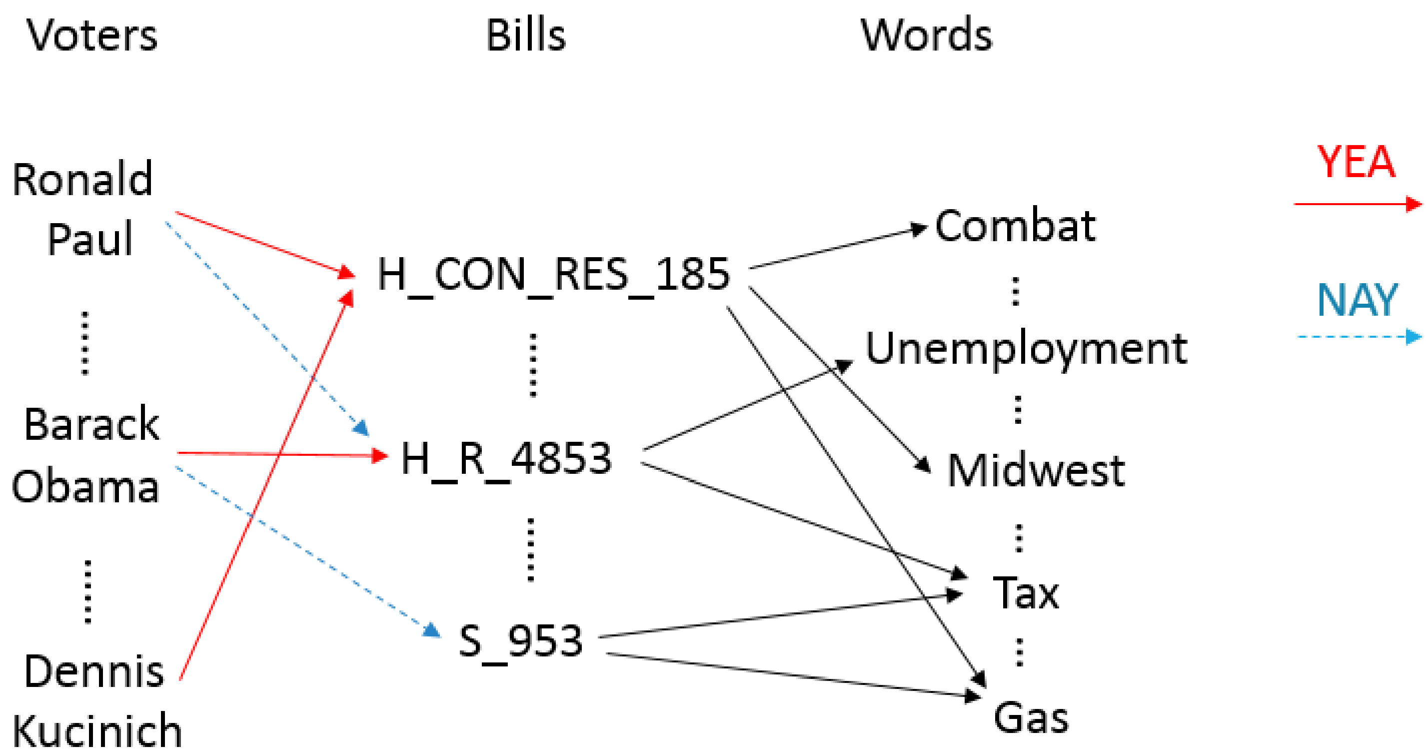
- Topic-Factorized Ideal Point Estimation Model for Legislative Voting Network (KDD'14, Gu, Sun et al.)

Background


United Stated Congress
The House
Senate
Politician
Republican
Democrat

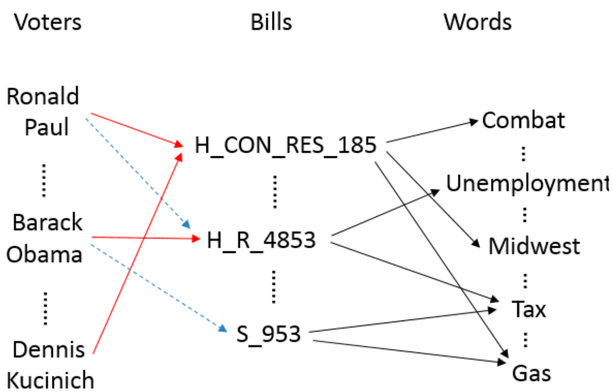

Federal Legislation (bill)
The House
Senate
Law
Bill 1 Bill 2
......
Ronald Paul
Barack Obama
Ronald Paul
Barack Obama
liberal
conservative

Legislative Voting Network



Problem Definition

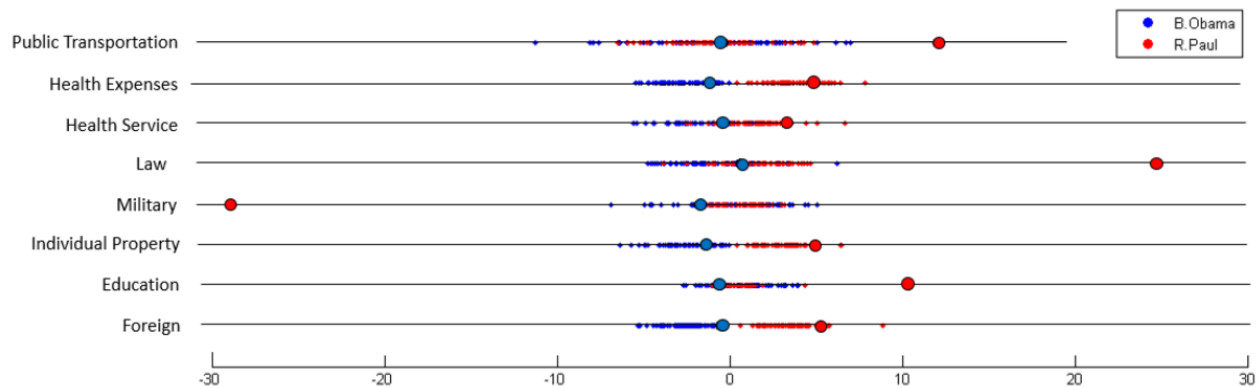
Input:
Legislative Network



Output:

$\boldsymbol{x}_u$: Ideal Points for Politician u

$\boldsymbol{a}_d$: Ideal Points for Bill d



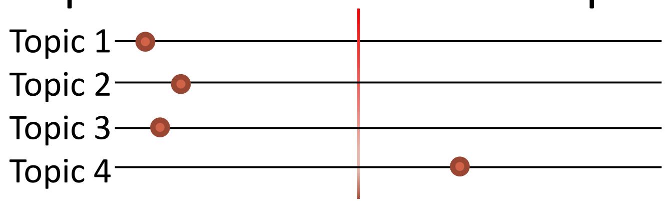
$\boldsymbol{x}_u$'s on different topics

Existing Work

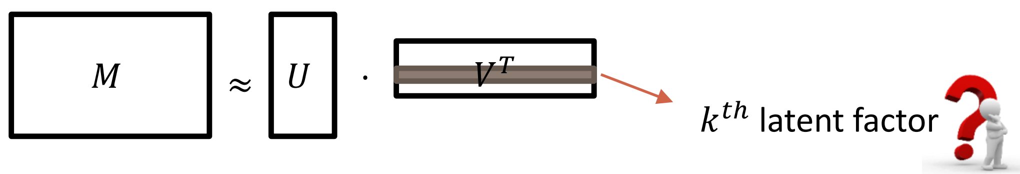
- 1-dimensional ideal point model (Poole and Rosenthal, 1985; Gerrish and Blei, 2011)
- High-dimensional ideal point model (Poole and Rosenthal, 1997)
- Issue-adjusted ideal point model (Gerrish and Blei, 2012)

Motivations

- Voters have different positions on different topics.



- Traditional matrix factorization method cannot give the meanings for each dimension.



- Topics of bills can influence politician’s voting, and the voting behavior can better interpret the topics of bills as well.

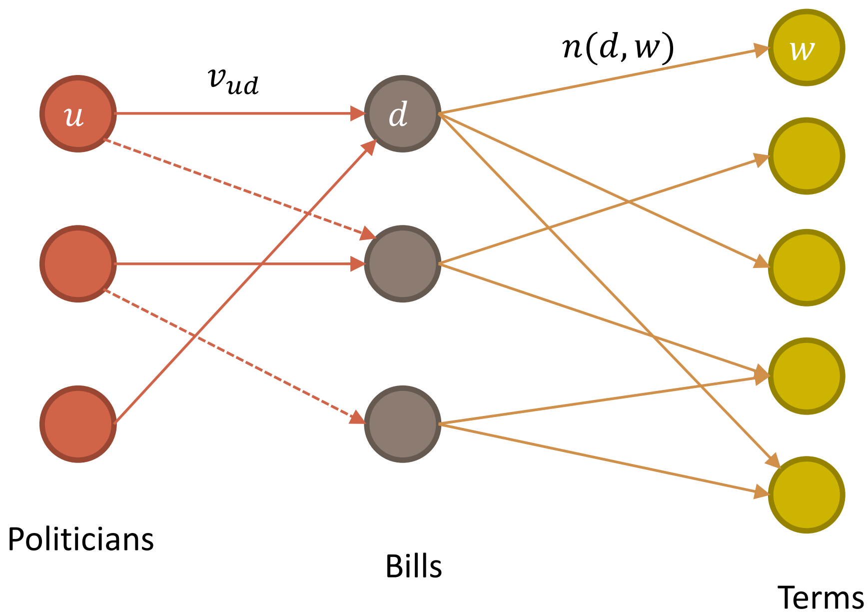
Topic Model:

- Health
- Public Transport
- …

Voting-guided Topic Model:

- Health Service
- Health Expenses
- Public Transport
- …

Topic-Factorized IPM



Heterogeneous Voting Network

Entities:

- Politicians
- Bills
- Terms

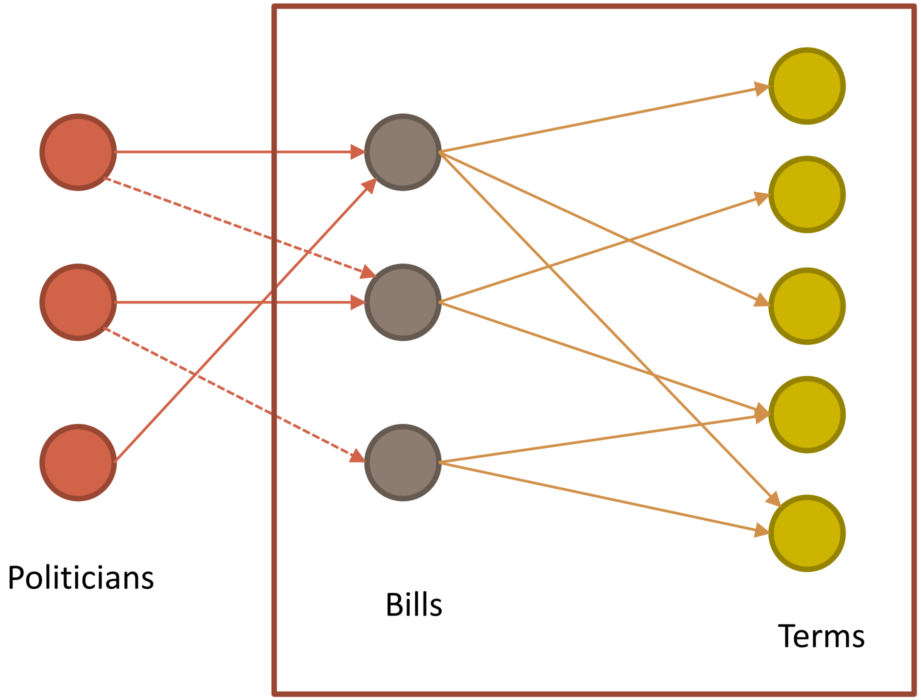
Links:

- (P, B)
- (B, T)

Parameters to maximize the likelihood of generating two types of links:

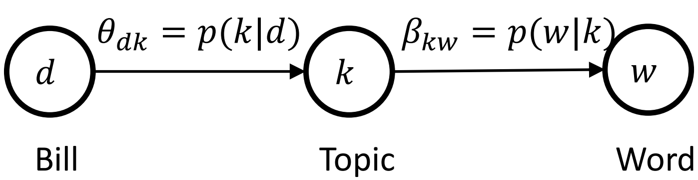
- Ideal points for politicians
- Ideal points for bills
- Topic models

Text Part



Text Part

- We model the probability of each word in each document as a mixture of categorical distributions, as in PLSA (Hofmann, 1999) and LDA (Blei et al., 2003)

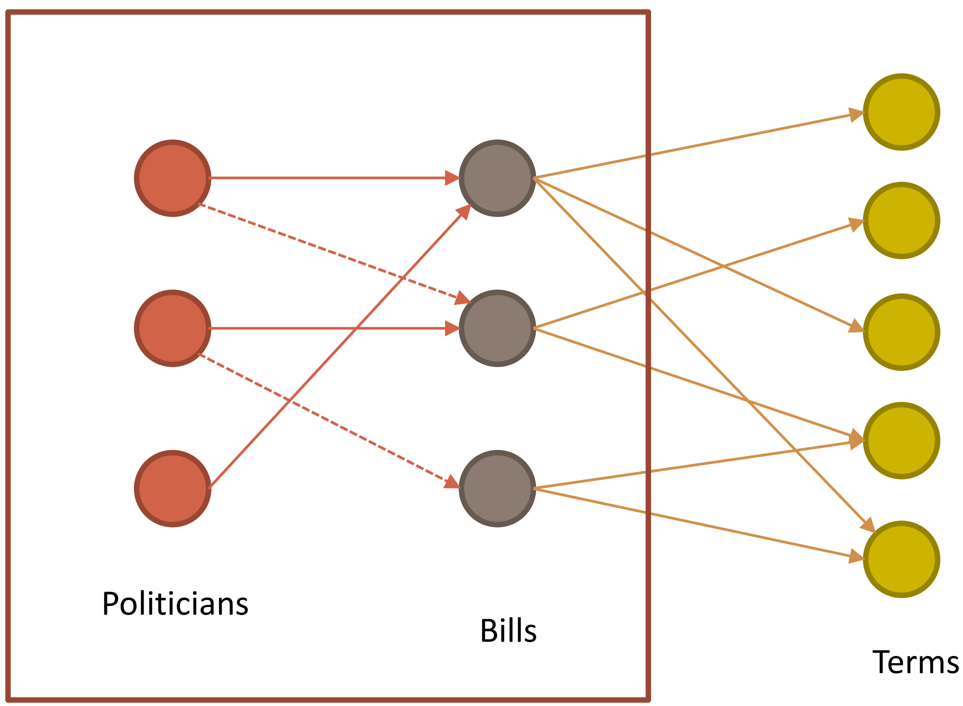


$$\boldsymbol{w}_d = \big(n(d,1), n(d,2), \ldots, n(d,N_w)\big)$$

$$p(\boldsymbol{w}_d|\boldsymbol{\theta},\boldsymbol{\beta}) \propto \prod_w \Big(\sum_k \theta_{dk}\beta_{kw}\Big)^{n(d,w)}$$

$$p(\boldsymbol{W}|\boldsymbol{\theta},\boldsymbol{\beta}) \propto \prod_d \prod_w \Big(\sum_k \theta_{dk}\beta_{kw}\Big)^{n(d,w)}$$

Voting Part



Intuitions:

- The more similar of the ideal points of *u* and *d*, the higher probability of “YEA” link
- The higher portion a bill belongs to topic *k*, the higher weight of ideal points on topic *k*

Voting Part

		Topic 1	Topic 2	......		Topic k	......	Topic K	
Voter u	$\boldsymbol{x}_u$	x_{u1}	x_{u2}			x_{uk}		x_{uK}	$x_{uk} \in \boldsymbol{R}$
Bill d	$\boldsymbol{a}_d$	a_{d1}	a_{d2}			a_{dk}		a_{dK}	$a_{dk} \in \boldsymbol{R}$

	d_1	d_2			......			d_{N_D}
u_1	0	1	-1	1	1	1	1	1
u_2	0	0	-1	1	1	1	-1	1
⋮								
u_{N_U}	1	1	1	1	-1	1	0	0

User-Bill voting matrix $\boldsymbol{V}$

$$\left.\begin{array}{ccc} \theta_{d1} & x_{u1} & a_{d1} \\ & \vdots & \\ \theta_{dk} & x_{uk} & a_{dk} \\ & \vdots & \\ \theta_{dK} & x_{uK} & a_{dK} \end{array}\right] \Rightarrow$$

$$\cancel{\hat{r}_{ud} = \sum_{k=1}^{K} x_{uk} a_{dk}}$$

$$\cancel{\hat{r}_{ud} = \sum_{k=1}^{K} \theta_{dk} x_{uk} a_{dk}}$$

$$p(v_{ud} = 1) = \sigma\Big(\sum_k \theta_{dk} x_{uk} a_{dk} + b_d\Big) \quad \xrightarrow{\text{YEA}}$$

$$p(v_{ud} = -1) = 1 - \sigma\Big(\sum_k \theta_{dk} x_{uk} a_{dk} + b_d\Big) \quad \xrightarrow{\text{NAY}}$$

$$p(\boldsymbol{V}|\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{A}, \boldsymbol{b}) = \prod_{(u,d):v_{ud}\neq 0} \left(p(v_{ud} = 1)^{\overset{I_{\{v_{ud}=1\}}}{\boxed{\frac{1+v_{ud}}{2}}}} p(v_{ud} = -1)^{\overset{I_{\{v_{ud}=-1\}}}{\boxed{\frac{1-v_{ud}}{2}}}}\right)$$

Combining Two Parts Together

- The final objective function is a linear combination of the two average log-likelihood functions over the *word links* and *voting links*.

$$J(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{X}, \boldsymbol{A}, \boldsymbol{b}) = (1-\lambda)\boxed{\frac{\sum_{d,w} n(d,w)\log(\sum_k \theta_{dk}\beta_{kw})}{N_F}} + \lambda\boxed{\frac{\sum_{(u,d):v_{ud}\neq 0}(\frac{1+v_{ud}}{2}\log p(v_{ud}=1) + \frac{1-v_{ud}}{2}\log p(v_{ud}=-1))}{N_V}}$$

s.t.

$$0 \le \theta_{dk} \le 1, \quad \sum_k \theta_{dk} = 1 \quad \text{and} \quad 0 \le \beta_{kw} \le 1, \quad \sum_w \beta_{kw} = 1$$

- We also add an l_2 regularization term to A and X to reduce over-fitting.

$$J(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{X}, \boldsymbol{A}, \boldsymbol{b}) = (1-\lambda)\frac{\sum_{d,w} n(d,w)\log(\sum_k \theta_{dk}\beta_{kw})}{N_F} + \lambda\frac{\sum_{(u,d):v_{ud}\neq 0}(\frac{1+v_{ud}}{2}\log p(v_{ud}=1) + \frac{1-v_{ud}}{2}\log p(v_{ud}=-1))}{N_V}$$
$$\boxed{-\frac{1}{2\sigma^2}(\sum_u ||\boldsymbol{x}_{\boldsymbol{u}}||_2^2 + \sum_d ||\boldsymbol{a}_d||_2^2)}$$

Learning Algorithm

- An iterative algorithm where ideal points related parameters (X, A, b) and topic model related parameters (θ, β) enhance each other.
  - Step 1: Update X, A, b given θ, β
    - Gradient descent
  - Step 2: Update θ, β given X, A, b
    - Follow the idea of expectation-maximization (EM) algorithm: maximize a lower bound of the objective function in each iteration

$$
\begin{aligned}
&\sum_{d,w} n(d,w) \log \left(\sum_k \theta_{dk} \beta_{kw} \right) \\
&= \sum_{d,w} n(d,w) \log \left(\sum_k p(k|d,w) \frac{\theta_{dk}\beta_{kw}}{p(k|d,w)} \right) \\
&\geq \sum_{d,w} n(d,w) \sum_k p(k|d,w) \log \frac{\theta_{dk}\beta_{kw}}{p(k|d,w)} \\
&= \sum_{d,w} n(d,w) \sum_k p(k|d,w) \log \theta_{dk}\beta_{kw} - c
\end{aligned}
$$

Learning Algorithm

- Update θ: A nonlinear constrained optimization problem.

  Remove the constraints by a logistic function based transformation:

$$\theta_{dk} = \begin{cases} \dfrac{e^{\mu_{dk}}}{1 + \sum_{k'=1}^{K-1} e^{\mu_{dk'}}} & \text{if } 1 \le k \le K-1 \\[2ex] \dfrac{1}{1 + \sum_{k'=1}^{K-1} e^{\mu_{dk'}}} & \text{if } k = K \end{cases}$$

  and update μ_{dk} using gradient descent.

- Update β:

  Since β only appears in the topic model part, we use the same updating rule as in PLSA:

$$\beta_{kw}^{new} = \frac{\sum_d n(d,w)p(k|d,w)}{\sum_{d,w} n(d,w)p(k|d,w)} \quad \text{where} \quad p(k|d,w) = \frac{\theta_{dk}\beta_{kw}^{old}}{\sum_{k'} \theta_{dk'}\beta_{k'w}^{old}}$$

Data Description

- Dataset:
  - U.S. House and Senate roll call data in the years between 1990 and 2013.*
    - 1,540 legislators
    - 7,162 bills
    - 2,780,453 votes (80% are “YEA”)
  - Keep the latest version of a bill if there are multiple versions.
  - Randomly select 90% of the votes as training and 10% as testing.

* Downloaded from http://thomas.loc.gov/home/rollcallvotes.html

Evaluation Measures

- *Root mean square error (RMSE)* between the predicted vote score and the ground truth

$$\text{RMSE} = \sqrt{\frac{\sum_{(u,d):v_{ud}\neq 0}\left(\frac{1+v_{ud}}{2} - p(v_{ud}=1)\right)^2}{N_V}}$$

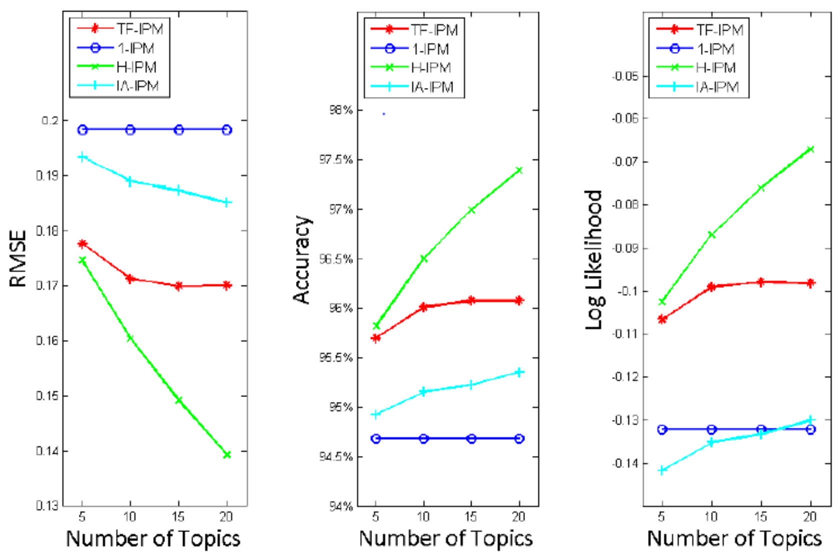
- *Accuracy* of correctly predicted votes (using 0.5 as a threshold for the predicted accuracy)

$$\text{Accuracy} = \frac{\sum_{u,d}\left(I_{\{p(v_{ud}=1)>0.5\ \&\&\ v_{ud}=1\}} + I_{\{p(v_{ud}=1)<0.5\ \&\&\ v_{ud}=-1\}}\right)}{N_V}$$

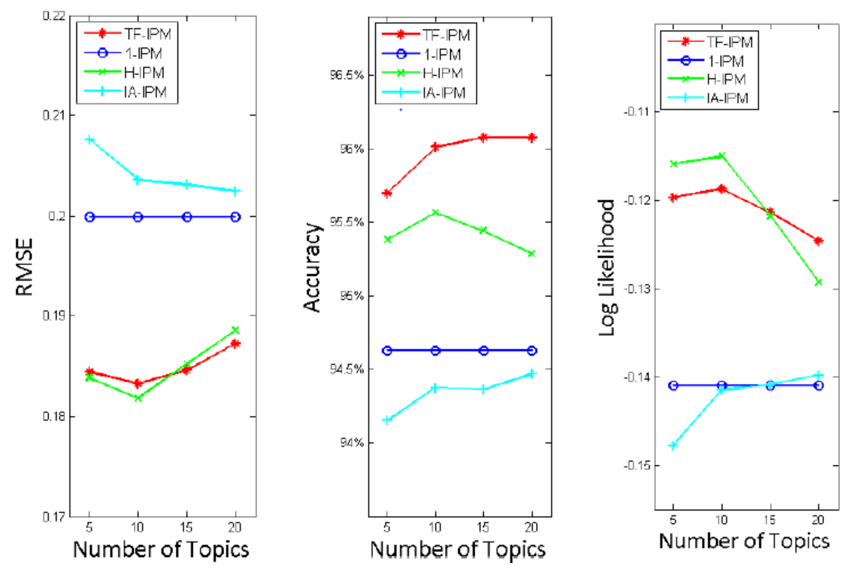
- *Average log-likelihood* of the voting link

$$\text{AvelogL} = \frac{\sum_{(u,d):v_{ud}\neq 0}\left(\frac{1+v_{ud}}{2}\log p(v_{ud}=1) + \frac{1-v_{ud}}{2}\log p(v_{ud}=-1)\right)}{N_V}$$

Experimental Results



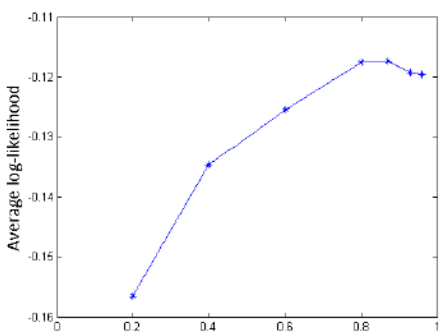
Training Data set



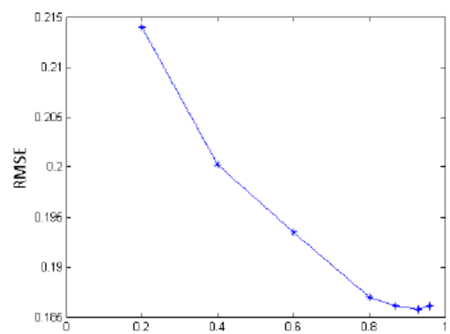
Testing Data set

Parameter Study

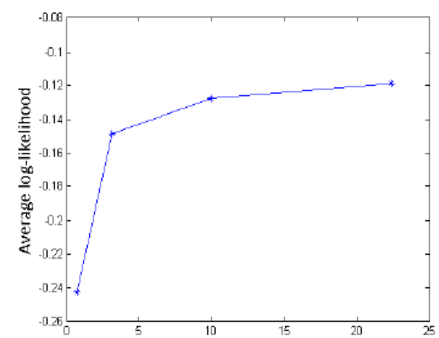
$$J(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{X}, \boldsymbol{A}, \boldsymbol{b}) = (1-\lambda) \cdot avelogL(text) + \lambda \cdot avelogL(voting) - \frac{1}{2\sigma^2}\left(\sum_u ||\boldsymbol{x_u}||_2^2 + \sum_d ||\boldsymbol{a}_d||_2^2\right)$$



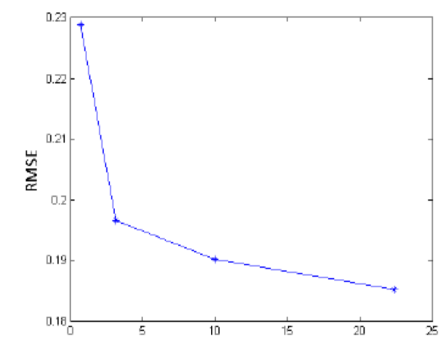
(a) Average log-likelihood



(b) RMSE



(a) Average log-likelihood



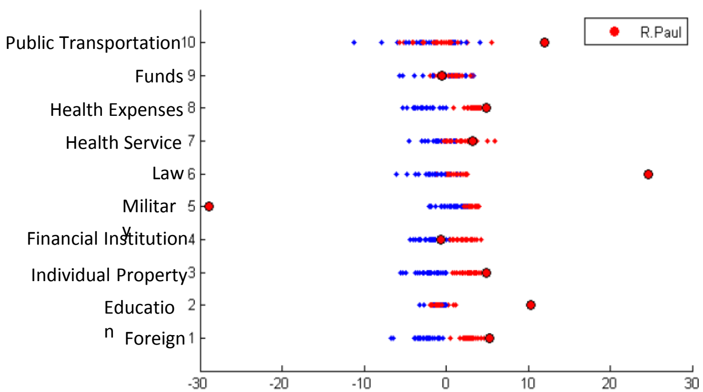
(b) RMSE

Parameter study on λ

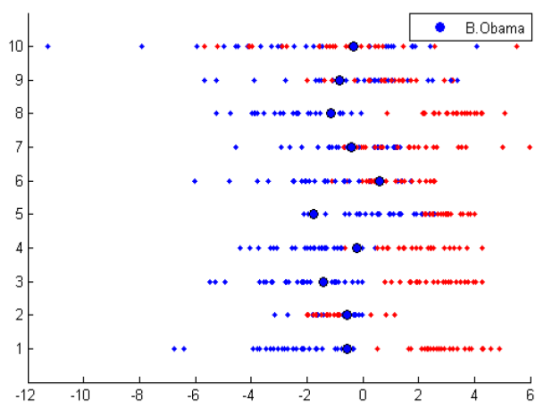
Parameter study on σ (regularization coefficient)

Case Studies

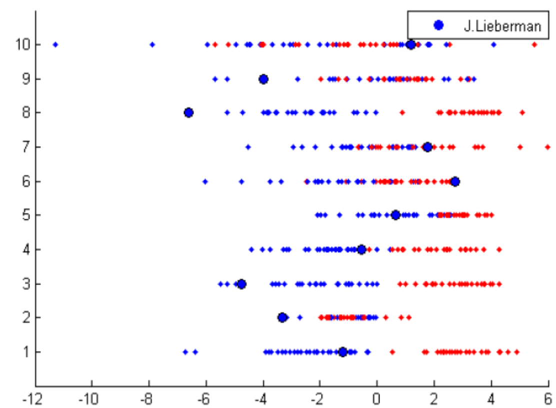
- Ideal points for three famous politicians: (Republican, Democrat)
  - Ronald Paul (R), Barack Obama (D), Joe Lieberman (D)



Ronald Paul



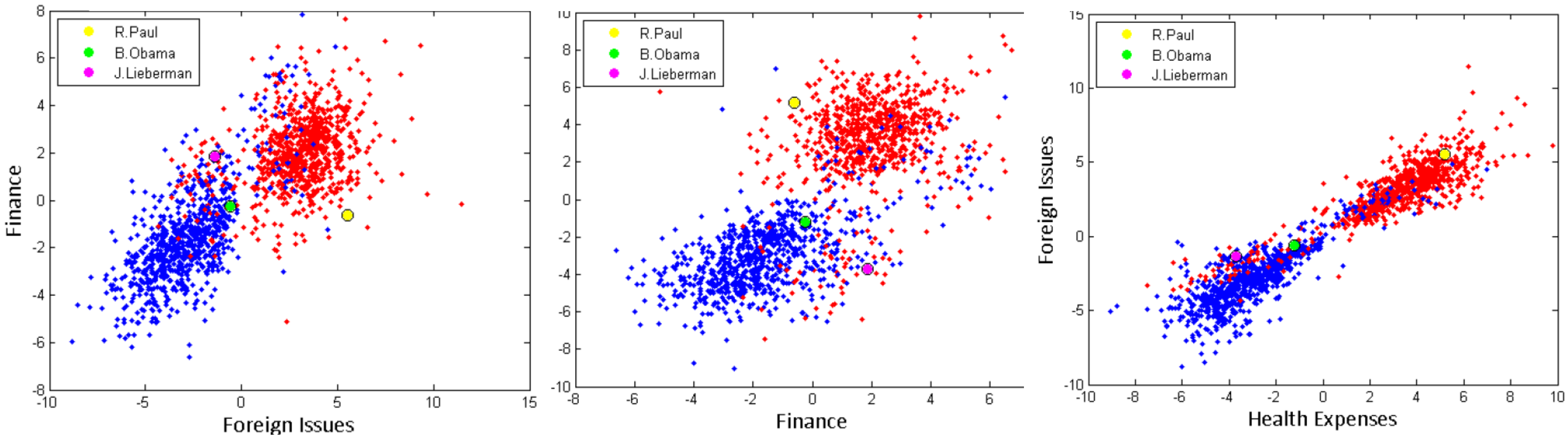
Barack Obama



Joe Lieberman

Case Studies

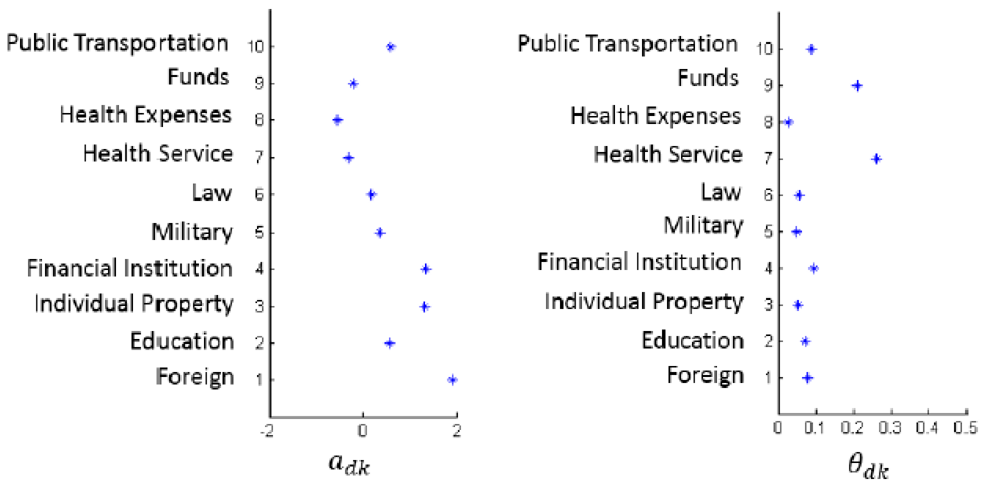
- Scatter plots over selected dimensions:
  (Republican, Democrat)



Case Studies

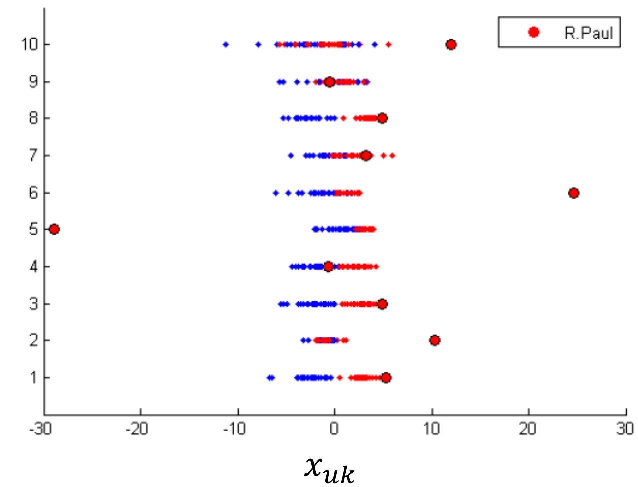
Bill: *H_RES_578* — **109th Congress (2005-2006)**

It is about supporting the government of Romania to improve the standard health care and well-being of children in Romania.



(a) Bill Positions (b) Topic Distributions

$$p(v_{ud} = 1) = \sigma(\sum_k \theta_{dk} x_{uk} a_{dk} + b_d)$$



R. Paul $\xrightarrow{\text{YEA}}$ *H_RES_578*

For Unseen Bill d:

Topic Model → $\boldsymbol{\theta}_d$

TF-IPM → $\boldsymbol{x}_u$

Experts/Algorithm → $\boldsymbol{a}_d$

$p(v_{ud} = 1)$

Outline

- Why Heterogeneous Information Networks?
- Entity Recommendation
- Information Diffusion
- Ideology Detection
- Summary

Summary

- Heterogeneous Information Networks are networks with multiple types of objects and links

- Principles in mining heterogeneous information networks
  - Meta-path-based mining
    - Systematically form new types of relations
  - Relation strength-aware mining
    - Different types of relations have different strengths
  - Relation semantic-aware mining
    - Different types of relations need different modeling

Q & A

THANK YOU!