The 26th ACM International Conference on Information and Knowledge Management
Singapore, 2017

Joint Topic-Semantic-aware Social Recommendation for Online Voting

[1]Hongwei Wang, [2]Jia Wang, [2]Miao Zhao, [2]Jiannong Cao, [1]Minyi Guo

[1] Shanghai Jiao Tong University
[2] The Hong Kong Polytechnic University

November 7, 2017

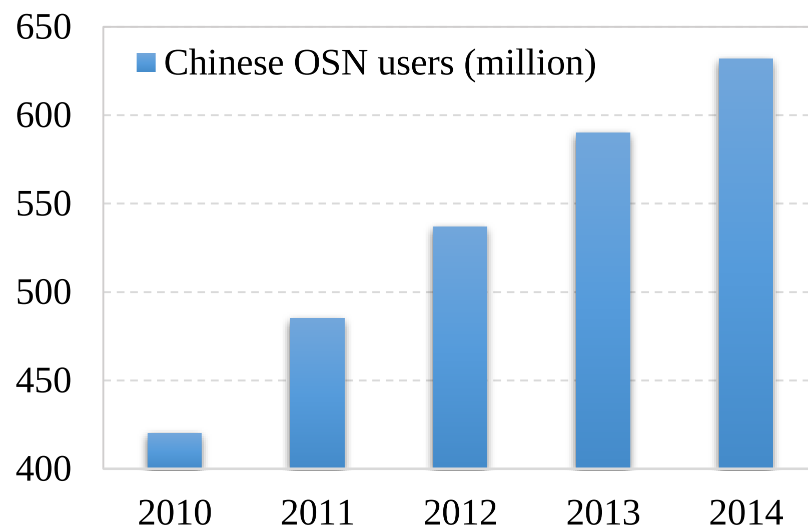
上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

Background (1/5)

Online Social Networks

- The past decade has witnessed the proliferation of online social networks
  - Facebook, Twitter, LinkedIn, Pinterest, Instagram, …
  - Weibo, Qzone, Wechat moments, Zhihu, Douban, …



Background (2/5)

Online Voting

- Online voting has recently become a popular function in online social platforms
  - users can freely initiate votings and customize voting options
  - users can share their opinion towards various interested subjects

Voting initialization *Voting propagation* *Voting details*

Background (3/5)

Voting Propagation

- In online social network, a user can …
  - initiate an original voting
  - participate an existing voting (and show to his followers)
  - simply retweet an existing voting

Initiate a voting

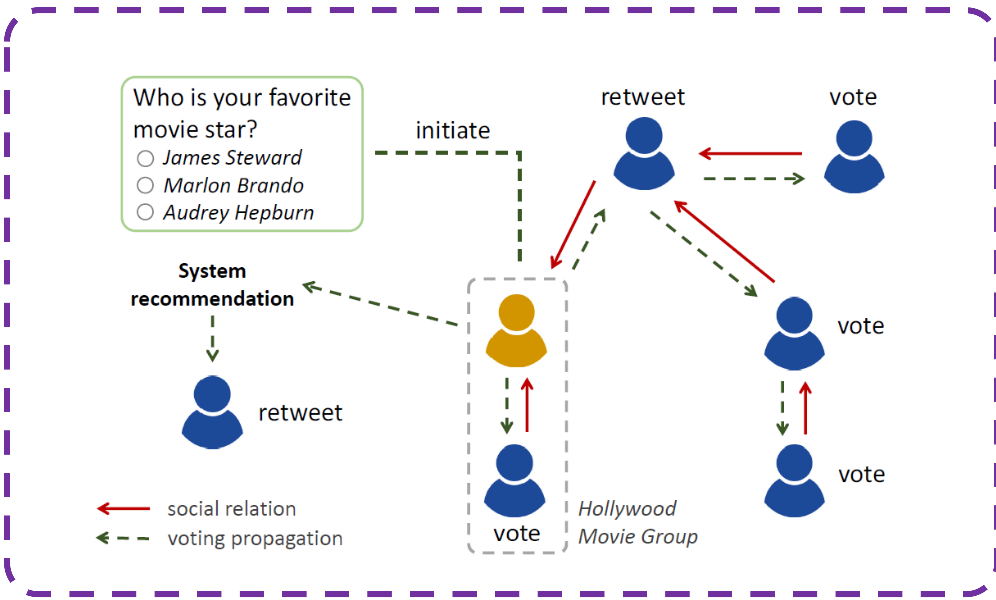
Participate a voting

Retweet a voting

Background (4/5)

Voting Propagation

- In online social network, a user can …
  - initiate an original voting
  - participate an existing voting (and show to his followers)
  - simply retweet an existing voting



Propagation scheme of online voting

Background (5/5)

Voting Recommendation

- Facing a large volume of diversified votings, a critical challenge is to present the "**right**" votings to the "**right**" person

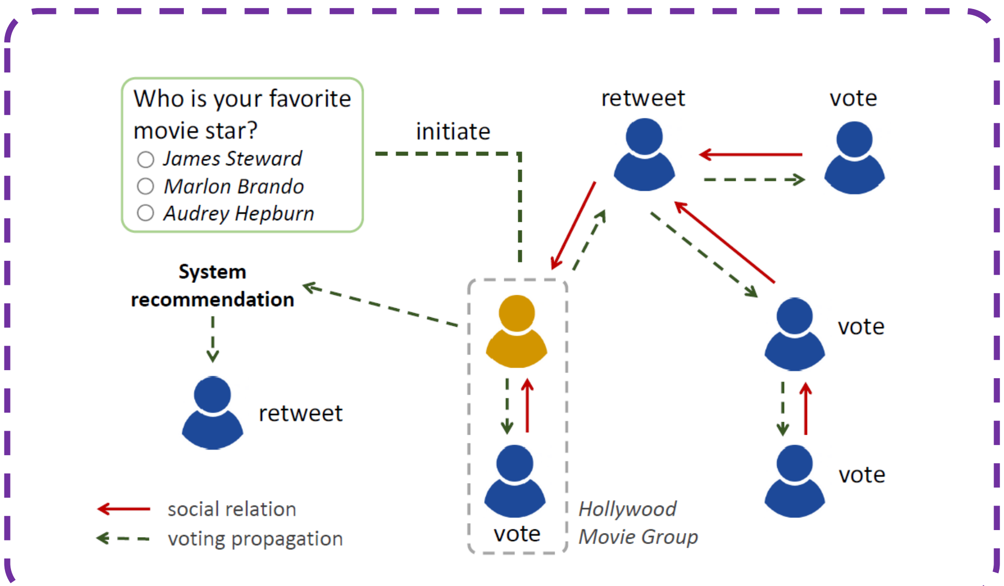
- An effective recommender system can precisely locates users' **voting favor**, thus to **improve user experience** and **maximize user engagement in votings**

- Such a recommender system can benefit a great many online services, e.g., **personalized advertising**, **market research**, **public opinion analysis**

Motivation (1/3)

Challenges for Voting Recommendation: 1

- The propagation of online votings relies heavily on social network structure
  - a user is more likely to be exposed to the votings that his friends are involved
  - a user can join different interest groups, which may potentially affects users' voting behavior



Motivation (2/3)

Challenges for Voting Recommendation: 2

- Users’ interests in votings are strongly connected with voting content in questions

How do you treat Xiong’an New Area Currently?

看 treat
雄安 Xiong’an
新区 new area
……

What are the scores of series between Warriors and Jazz?

勇士 Warriors
爵士 Jazz
系列赛 series
比分 score
……

Motivation (3/3)

Challenges for Voting Recommendation: 2

- **Topic models** (e.g., LDA)
  - discover the latent topic distribution of text
  - fail to process voting questions which are generally short and lack sufficient topic information

- **Semantic models** (e.g., word2vec)
  - learn text representations by using co-occurrence of words
  - fail to discriminate homonymy and polysemy

How to treat Xiong'an currently?

Too short to mine useful topics

What are the scores of series between Warriors and Jazz?

Polysemy of "Warriors" and "Jazz"

In this paper…

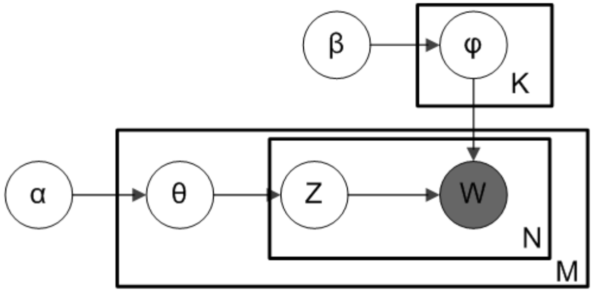
- We propose a **Joint Topic-Semantic-aware social Matrix Factorization model (JTS-MF)** for voting recommendation
- JTS-MF considers **social network structure** and **representation of voting content** in a comprehensive manner
- JTS-MF preserves the **topic-semantic-social similarity** among users and votings during matrix factorization

Joint-Topic-Semantic Embedding Learning (1/4)

Topic Distillation: LDA

$$p(\Theta, \Phi, z, \boldsymbol{w} | \boldsymbol{\alpha}, \boldsymbol{\beta})$$
$$= \prod_{z} p(\Phi_z | \boldsymbol{\beta}) \cdot \prod_{d} \left(p(\Theta_d | \boldsymbol{\alpha}) \prod_{l} \left(p(z_{d,l} | \Theta_d) p(w_{d,l} | \Phi_{z_{d,l}}) \right) \right)$$

where d traverses all group-aggregated documents

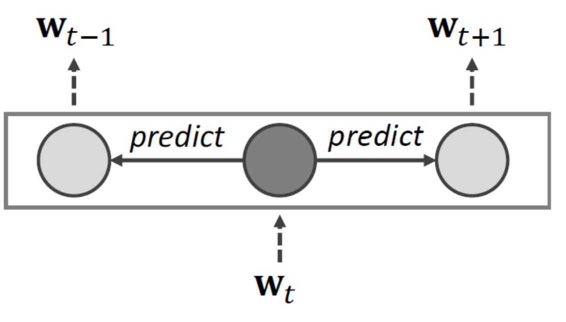


Joint-Topic-Semantic Embedding Learning (2/4)

Semantic Distillation: Skip-Gram

$$\mathcal{L}(D) = \frac{1}{T}\sum_{t=1}^{T}\sum_{-k\le c\le k,\, c\neq 0}\log p(w_{t+c}|w_t),$$

where $$p(w_i|w_t) = \frac{\exp(\mathbf{w}_i^\top \mathbf{w}_t)}{\sum_{w\in V}\exp(\mathbf{w}^\top \mathbf{w}_t)}$$



Joint-Topic-Semantic Embedding Learning (3/4)

Topical-Enhanced Word Embedding

- TEWE jointly utilizes the target word: w, the topic of the target word: z_w, and the topic of the document that the word belongs to: z_w^d
- TEWE regards each word-topics triplet $\langle w, z_w, z_w^d \rangle$ as a pseudo word and learns a unique vector $\mathbf{w}^{z,z^d}$
- TEWE objective function:

$$\mathcal{L}(D) = \frac{1}{T} \sum_{t=1}^{T} \sum_{-k \leq c \leq k, c \neq 0} \log p(\langle w_{t+c}, z_{t+c}, z_{t+c}^d \rangle | \langle w_t, z_t, z_t^d \rangle)$$

where

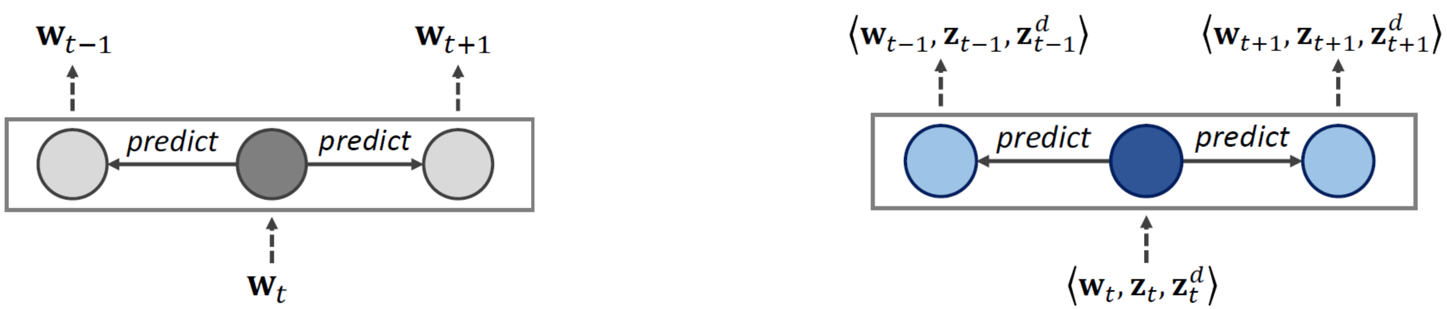
$$p(\langle w_i, z_i, z_i^d \rangle | \langle w_t, z_t, z_t^d \rangle) = \frac{\exp\left(\mathbf{w}_i^{z_i, z_i^d \top} \mathbf{w}_t^{z_t, z_t^d}\right)}{\sum_{\langle w, z, z^d \rangle \in \langle V, Z, Z \rangle} \exp\left(\mathbf{w}^{z, z^d \top} \mathbf{w}_t^{z_t, z_t^d}\right)}$$

Joint-Topic-Semantic Embedding Learning (4/4)

Topical-Enhanced Word Embedding

- TEWE representations for document d can be calculated as

$$\mathbf{e}_d = \sum_{w \in d} \text{TF-IDF}(w, d) \cdot \mathbf{w}^{z, z^d}$$



(a) Skip-Gram (b) TEWE

Recommendation Model (1/3)

Similarity Coefficients

- Normalized social-level similarity coefficient of users

$$S_{i,k} = I_{u_i,u_k} \sqrt{\frac{d_k^- + d}{d_i^+ + d_k^- + d}} \mathbf{e}_{u_i}^\top \mathbf{e}_{u_k} \qquad \widehat{S}_{i,k} = \frac{S_{i,k}}{\sum_{k \in \mathcal{F}_i^+} S_{i,k}}$$

- Normalized group-level similarity coefficient of users

$$G_{i,k} = \sum\nolimits_{G_c \in \mathcal{G}} I_{u_i,G_c} I_{u_k,G_c} \mathbf{e}_{u_i}^\top \mathbf{e}_{G_c} \qquad \widehat{G}_{i,k} = \frac{G_{i,k}}{\sum_{k \in \mathcal{G}_i} G_{i,k}}$$

- Normalized similarity coefficient of votings

$$T_{j,t} = \mathbf{e}_{v_j}^\top \mathbf{e}_{v_t} \qquad \widehat{T}_{j,t} = \frac{T_{j,t}}{\sum_{t \in \mathcal{V}_j} T_{j,t}}$$

Recommendation Model (2/3)

Loss Function

$$
\begin{aligned}
L = & \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{M} I'_{i,j} \left(R_{i,j} - Q_i P_j^{\top} \right)^2 \\
& + \frac{\alpha}{2} \sum_{i=1}^{N} \left\| Q_i - \sum_{k \in \mathcal{F}_i^+} \widehat{S}_{i,k} Q_k \right\|_2^2 \\
& + \frac{\beta}{2} \sum_{i=1}^{N} \left\| Q_i - \sum_{k \in \mathcal{G}_i} \widehat{G}_{i,k} Q_k \right\|_2^2 \\
& + \frac{\gamma}{2} \sum_{j=1}^{M} \left\| P_j - \sum_{t \in \mathcal{V}_j} \widehat{T}_{j,t} P_t \right\|_2^2 \\
& + \frac{\lambda}{2} \left(\|Q\|_F^2 + \|P\|_F^2 \right).
\end{aligned}
$$

$\boldsymbol{Q}_i$ is the latent feature of user u_i, $\boldsymbol{P}_j$ is the latent feature of voting v_j, $I'_{i,j}$ is the training weights:

$$
I'_{i,j} = \begin{cases} 1, & if\ u_i\ participates\ v_j \\ I_m, & otherwise \end{cases}
$$

Recommendation Model (3/3)

Learning Algorithm

- Gradients

$$\frac{\partial L}{\partial Q_i} = \sum_{j=1}^{M} -I'_{i,j}\left(R_{i,j} - Q_i P_j^{\top}\right) P_j$$
$$+ \alpha\Big((Q_i - \sum_{k\in\mathcal{F}_i^+} \widehat{S}_{i,k} Q_k) + \sum_{t\in\mathcal{F}_i^-} -\widehat{S}_{t,i}(Q_t - \sum_{k\in\mathcal{F}_t^+} \widehat{S}_{t,k} Q_k)\Big)$$
$$+ \beta\Big((Q_i - \sum_{k\in\mathcal{G}_i} \widehat{G}_{i,k} Q_k) + \sum_{t\in\mathcal{U}} -\widehat{G}_{t,i}(Q_t - \sum_{k\in\mathcal{G}_i} \widehat{G}_{t,k} Q_k)\Big)$$
$$+ \lambda Q_i,$$

$$\frac{\partial L}{\partial P_j} = \sum_{i=1}^{N} -I'_{i,j}\left(R_{i,j} - Q_i P_j^{\top}\right) Q_i$$
$$+ \gamma\Big((P_j - \sum_{t\in\mathcal{V}_j} \widehat{T}_{j,t} P_t) + \sum_{k\in\mathcal{V}_j} -\widehat{T}_{k,j}(P_k - \sum_{t\in\mathcal{V}_k} \widehat{T}_{k,t} P_t)\Big)$$
$$+ \lambda P_j.$$

- Pseudo code

(1) Randomly initialize Q and P;
(2) At each iteration of the algorithm, do:
a) update each Q_i: $Q_i \leftarrow Q_i - \delta \frac{\partial L}{\partial Q_i}$;
b) update each P_j: $P_j \leftarrow P_j - \delta \frac{\partial L}{\partial P_j}$;
untill convergence, where δ is an adjustable learning rate.

Experiments (1/5)

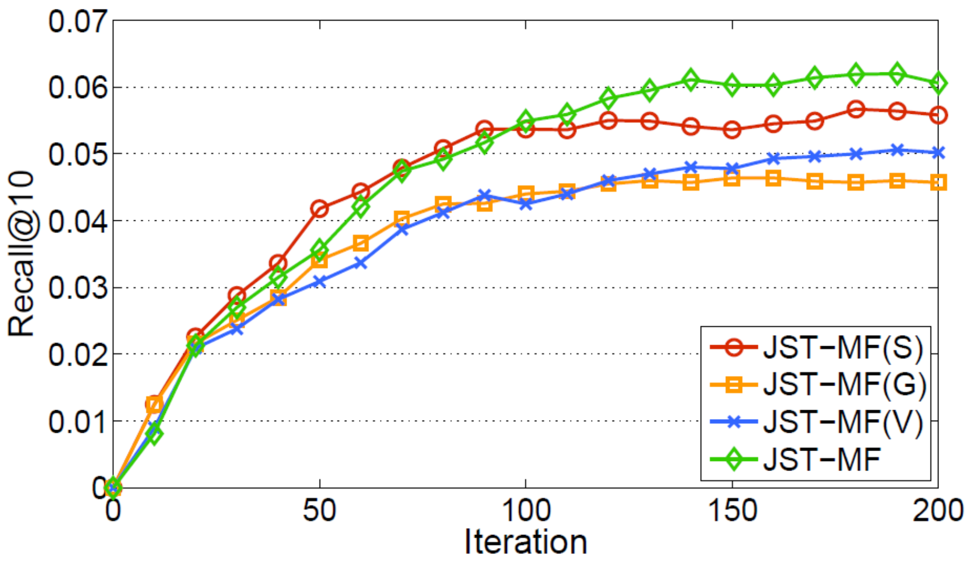
Dataset Statistics

# users	1,011,389	# groups	299,077
# users with votings	525,589	# user-voting	3,908,024
# users with groups	723,913	# user-user	83,636,677
# votings	185,387	# user-group	5,643,534

Experiments (2/5)

Learning Curves

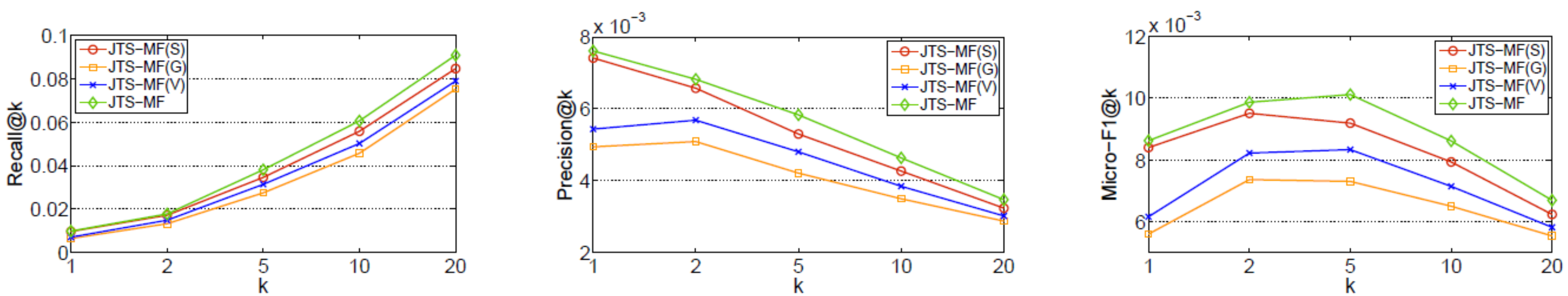
- $\alpha = 10$ for JTS-MF(S)
- $\beta = 140$ for JTS-MF(G)
- $\gamma = 30$ for JTS-MF(V)
- $\alpha = 10, \beta = 140, \gamma = 30$ for JTS-MF



Experiments (3/5)

Study of JTS-MF

- JTS-MF > JTS-MF(S) > JTS-MF(V) > JTS-MF(G)
- Social-level friends are more helpful than group-level friends when determining users' voting interest
- The three types of similarities can be well combined by JTS-MF to achieve even better result



(a) (b) (c)

Experiments (4/5)

Comparison of Models

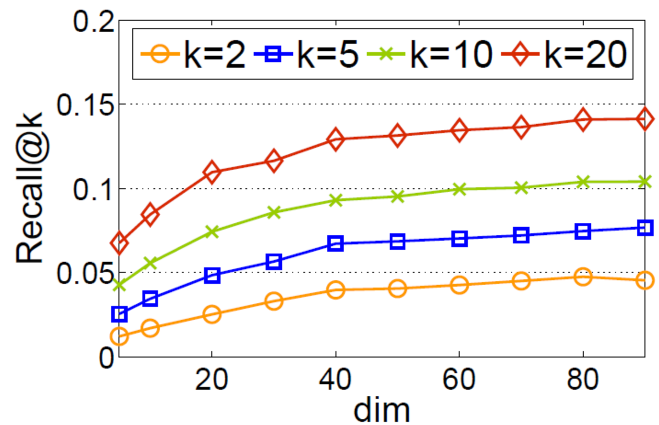
- MostPop performs worst
- JTS-MF outperforms Topic-MF and Semantic-MF
- JTS-MF performs best when k is small, but is slightly inferior to JTS-MF(S) when k gets larger

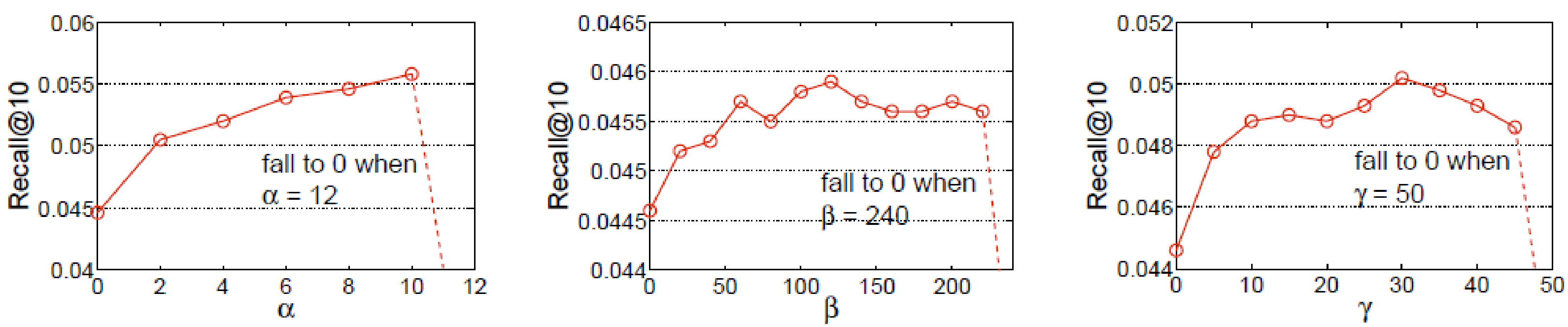
Model	Metric	k							
		1	2	5	10	20	50	100	500
JTS-MF(S)	*Recall*	0.0097	0.0172	0.0346	0.0558	0.0846	**0.1529**	**0.2229**	**0.4392**
	Precision	0.007416	0.006575	0.005300	0.004271	0.003238	**0.002341**	**0.001707**	**0.000672**
	Micro-F1	0.008401	0.009511	0.009192	0.007935	0.006238	**0.004612**	**0.003387**	**0.001343**
JTS-MF(G)	*Recall*	0.0065	0.0133	0.0275	0.0457	0.0752	0.1360	0.2051	0.4216
	Precision	0.004944	0.005092	0.004212	0.003500	0.002877	0.002082	0.001570	0.000645
	Micro-F1	0.005601	0.007365	0.007306	0.006503	0.005542	0.004102	0.003116	0.001289
JTS-MF(V)	*Recall*	0.0071	0.0149	0.0314	0.0502	0.0789	0.1387	0.2049	0.4176
	Precision	0.005439	0.005685	0.004805	0.003846	0.003021	0.002124	0.001568	0.000639
	Micro-F1	0.006161	0.008223	0.008335	0.007145	0.005819	0.004184	0.003112	0.001277
JTS-MF	*Recall*	**0.0099**	**0.0178**	**0.0381**	**0.0606**	**0.0908**	0.1520	0.2187	0.4297
	Precision	**0.007614**	**0.006823**	**0.005834**	**0.004637**	**0.003475**	0.002327	0.001674	0.000658
	Micro-F1	**0.008625**	**0.009868**	**0.010118**	**0.008615**	**0.006695**	0.004585	0.003322	0.001314
MostPop	*Recall*	0.0042	0.0085	0.0191	0.0313	0.0517	0.0974	0.1455	0.3086
	Precision	0.003221	0.003261	0.002921	0.002403	0.001972	0.001482	0.001119	0.000469
	Micro-F1	0.003637	0.004721	0.005062	0.004468	0.003804	0.002925	0.002218	0.000937
Basic-MF	*Recall*	0.0063	0.0129	0.0274	0.0446	0.0727	0.1368	0.2050	0.4198
	Precision	0.004845	0.004944	0.004192	0.003411	0.002783	0.002094	0.001569	0.000643
	Micro-F1	0.005489	0.007151	0.007271	0.006337	0.005361	0.004125	0.003114	0.001283
Topic-MF	*Recall*	0.0076	0.0147	0.0311	0.0495	0.0781	0.1395	0.2076	0.4210
	Precision	0.005834	0.005636	0.004766	0.003787	0.002991	0.002136	0.001589	0.000644
	Micro-F1	0.006609	0.008152	0.008266	0.007035	0.005761	0.004207	0.003154	0.001287
Semantic-MF	*Recall*	0.0093	0.0169	0.0333	0.0545	0.0860	0.1471	0.2142	0.4293
	Precision	0.007120	0.006476	0.005102	0.004173	0.003293	0.002252	0.001639	0.000657
	Micro-F1	0.008065	0.009368	0.008849	0.007752	0.006342	0.004437	0.003254	0.001313

Experiments (5/5)

Parameter Sensitivity

- Dimension of latent features
- Trade-off parameters α, β, and γ





(a) (b) (c)

Summary

- In this paper we study the problem of **recommending online votings to users** in social networks
- We propose **Topical-Enhanced Word Embedding (TEWE)** to jointly consider topics and semantics of words and documents
- We propose **Joint Topic-Semantic-aware social Matrix Factorization (JTS-MF)** model to learn latent features of users and votings based on social network structure and TEWE representation
- Experiment results prove the **competitiveness of JTS-MF** and demonstrate the **efficacy of TEWE representation**

Q & A

Thanks!