[illegible]

[illegible], Member, IEEE, [illegible], Senior Member, IEEE

Abstract— [illegible]

Index Terms— [illegible]

◆

1 INTRODUCTION

[illegible]

- *J. Zhang and Z. Fang are with the Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China. E-mail: {zhangjing12, fzp13}@mails.tsinghua.edu.cn.*
- *W. Chen is with Theory Group, Microsoft Research, Beijing 100080, China. E-mail: weic@microsoft.com.*
- *J. Tang is with the Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China, and Tsinghua National Laboratory for Information Science and Technology (TNList). E-mail: jietang@tsinghua.edu.cn.*

Manuscript received 11 Jan. 2014; revised 5 Jan. 2015; accepted 9 Feb. 2015. Date of publication 25 Feb. 2015; date of current version 2 July 2015.
Recommended for acceptance by G. Das.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TKDE.2015.2407351

1041-4347 © 2015 [illegible]

2 “FOLLOWING” LINK CASCADE MODEL

neighbor-ing lin

$v \in V$

$\in E$

e A B

$n,$

$t(e_{uv}) = \perp$

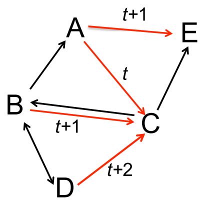
A

$C.$

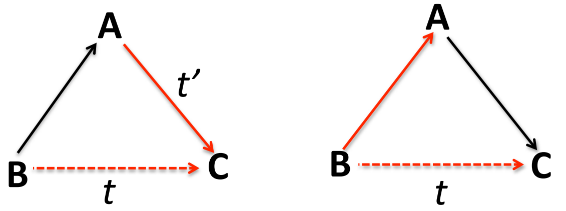
ime.

“follo

Organization.



E . 2.



(a) Follower diffusion

f
(. ., $\delta = 7$), ff .
e, e f
f .
f
f f f -
e', ff -
, -
. f e' B (f -
f e) B
A f C,
B A f C
A f C.
B C f
f C f C. ,

. ,
. , f
ff .
. f ff . e_{AC}
f t, f B f e_{BC}
e_{AC} $t+1$ e_{AC} ff e_{BC} f $t+1$.
, e_{BC} ff e_{DC} f $t+2$.
, f A f e_{AE} e_{AC}
$t+1$ e_{AC} ff e_{AE} f $t+1$.
f -
( ) , , .
ff -
, ff ,
f .
f , ,
.
f f
f f
. , f-
f , . ., f-
f .
ff

Follower diffusion. f f A f C f
t', f f B f C
t $t' \le t \le t'+\delta$, B A' f f -
$t'-1$, f
f f e_{AC} f f e_{BC}.
f f f C f ,
follower diffusion.

. f f ff
(A,B,C) e_{BA}
$t'-1$. , f f -
ff , . ., $t'-1$, B f -
f A, C f f B, .
, f ff

Followee diffusion. f f B f A f
t', f f B f C
t $t' \le t \le t'+\delta$, C A' f f -
($t'-1$),
f f f e_{BA} f f e_{BC}.
f f f B f ,
followee diffusion. .
f f ff .

ff . f ff ,
f
, ff f
. f ff ,
f , ff
f . ff -
, ff f -
. , f ff
f f
f . ff ,
, f -
.

. f ff ,
f f .
f -
f - f ,
f ff - f
. , f ff f f-
f f ,
f f .

N f
ff , . ., ff f e_{AC} e_{CB}.
, f
. f ,
f ff .

3 DATA AND OBSERVATIONS

, ff
ff ff .
f f-
f ff .

ff , -
.

3.1 Data Collection

f . -
, f .
,
, . ., ,
0,000 f f .
f f
f ’ , , , ’
, ’ , . f
f 0/ / 0 0 / / 0 0.
, ,
. ’’ -
f ,0 , ’’
, , 0
f 0/ / 0 0 / / 0 0 .
. f f (
f) f
. -
. f
.

3.2 Observations

f f ff ,
f ff .

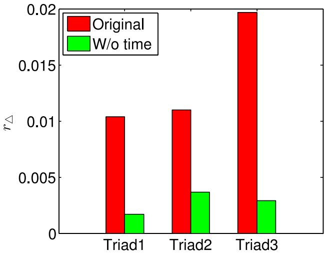
. 1 12 f ff -
13 24 f ff .
ff -
( ) -
( )
t' t',
f ff
( ) t
ff f ff ,
. -
f $0 \leq t - t' \leq \delta$ (δ
f).
, $C_\triangle$ -
$\triangle$, $C^+_\triangle$
B f C $[t', t'+\delta]$. $|C_\triangle|$ f
$\triangle$. $r_\triangle$
f B f C , -

$$r_\triangle = \frac{|C^+_\triangle|}{|C_\triangle|}. \tag{1}$$

ff ff ff -
f f

- -
- ff ff ff

Pattern significance.
f
. -f ,
f ,
. -
. , -
f ,



(

f

. f - $\leq 0.05,$

. ,

f f

. U , f-

f f f ,

. ( ).

, δ -

f 0,000 . f ff -

, - 0,

f ff . f

ff , - f , , ,

, 0.0 ,

. - . ,

, , f B f C C

f B f B f ,

f A f C. , , 0,

, -

, ff

ff f e_{AC}. , , ,

A f C f , C

. f , ff ff

f C f .

,

- . , , , e_{AC} f

f f

. ff ff , , f

. f , , , , ,

f .

Diffusion decay. ff f ff

, f δ $1, 2, 3, 5, 7$ 10 .

f δ, $r_\triangle$ f ff -

f ff ,

. . ,

f $r_\triangle$.

δ , $r_\triangle$,

ff ff f

. f f-

f , $r_\triangle$ f ff

. ,

f ff (.),

f B A f $C,$ A'

f C' . f ff

(.), B A f C

f f , B

C f f A

A' f C' . , f f B f -

C f ff f

ff . , f δ

.

f f-

f f ff . ,

$r_\triangle$ δ . ,

(. ., $t = \perp$), f

f , $r_\triangle$. -

f

. . , -

, $r_\triangle$ (f

)

(f

f), ff

ff . -

, . ff . ,

ff f ff -

f ff .

Other observations. f f ff ,

ff

- A B.

f (- , -

0-). , -

e_{BA}, - -

e_{AB}, e_{AB} e_{BA}.

-

A B ff

f B f C -

A B (+).

-

,

f , f

ff f f .

f f ff ,

ff ff -

f A C.

f (- , - ,

- , -) f f

ff . f ,

- f A C

ff f f B f C

- f C A (+ − 0).

f B f A f , B -

A' f C A'

f C' , A' C

B C.

Summary. , f f

,

ff ff ,

-

(+) - -
f A C ff -
f A f C B f C (+ $-$ 0).

4 MODEL LEARNING

, ff ff
. f
.

Likelihood function. ,
f
f .
$\theta = \{h_{e'e}, g_{e'e}\}$ -
f . , f

(e', e). ff $h_{e'e}$
$g_{e'e}$ f (e', e) f

(f e , ( ) . ( )- . ( ) . ( ) 0.0 0 / . 0 f (00 .)-00 . 0.

$y_{e'e}$ f e'
e $[t_{e'}, t_e]$, f $t_{e'}$
t_e, e' e f . , $y_{e'e}$
e' e f t_e

$$\begin{aligned} y_{e'e} &= 1 - h_\triangle g_\triangle \sum_{t=t_{e'}}^{t_e} (1-g_\triangle)^{t-t_{e'}} \\ &= h_\triangle (1-g_\triangle)^{t_e - t_{e'}+1} + (1-h_\triangle). \end{aligned} \tag{6}$$

, f ff
f . $e \in \mathcal{E}$
f
δ . f -
e' $y_{ee'}$ f $e' \in R_e$, R_e
f f e f $t_e + \delta$. -
$y_{ee'}$. ( ), $t_{e'}$
t_e t_e $t_e + \delta$.
- f

$$\log \mathcal{L} = \sum_{e\in\mathcal{E}} \left\{ \log \sum_{\vec{\alpha}_{S_e}} \prod_{e' \,\in\, S_e} x_{e'e}^{\alpha_{e'}} y_{e'e}^{1-\alpha_{e'}} + \sum_{e'\in R_e} \log y_{ee'} \right\}.$$

EM algorithm.
.
$q(e|\vec{\alpha}_{S_e}) = \frac{p(e|\vec{\alpha}_{S_e})}{p(e|S_e)}$
,
f
- f

$$\begin{aligned} \log \mathcal{L} &= \sum_{e\in\mathcal{E}} \left\{ \log \sum_{\vec{\alpha}_{S_e}} \hat{q}(e|\vec{\alpha}_{S_e}) \frac{p(e|\vec{\alpha}_{S_e})}{\hat{q}(e|\vec{\alpha}_{S_e})} + \sum_{e'\in R_e} \log y_{ee'} \right\} \\ &\geq \sum_{e\in\mathcal{E}} \left\{ \sum_{\vec{\alpha}_{S_e}} \hat{q}(e|\vec{\alpha}_{S_e}) \log \frac{p(e|\vec{\alpha}_{S_e})}{\hat{q}(e|\vec{\alpha}_{S_e})} + \sum_{e'\in R_e} \log y_{ee'} \right\}, \end{aligned}$$

ˆ f -
. $\hat{q}(e|\vec{\alpha}_{S_e}) \log \hat{q}(e|\vec{\alpha}_{S_e})$
f ,
.
$Q(\theta, \hat{\theta})$

$$Q(\theta, \hat{\theta}) = \sum_{e\in\mathcal{E}} \left\{ \sum_{\vec{\alpha}_S} \right.$$

$$h_{\triangle} = \frac{\sum_{(e',e)\in C^{+}_{\triangle}} \hat{D}_{e'e} + \sum_{(e',e)\in C^{-}_{\triangle}} \hat{B}_{e'e}}{|C_{\triangle}|}, \quad (12)$$

$$g_{\triangle} = \frac{\sum_{(e',e)\in C^{+}_{\triangle}} \hat{A}_{e'e}}{\sum_{(e',e)\in C^{-}_{\triangle}} \hat{B}_{ee'}(\delta+1) + \sum_{(e',e)\in C^{+}_{\triangle}} \hat{D}_{e'e}(t_e - t_{e'}+1)}. \quad (13)$$

. ( ) ( ), $C^{+}_{\triangle}$ $C^{-}_{\triangle}$. $A_{e'e}$, $B_{e'e}$ $D_{e'e}$ f , $A_{e'e}$ $B_{e'e}$. ( ) ( ), $D_{e'e}$ f

$$D_{e'e} = B_{e'e} + A_{e'e} - A_{e'e}B_{e'e}. \quad (14)$$

.

Algorithm 1. .

```
: $G = (V, E, t)$
: $\theta = \{h_{\triangle}, g_{\triangle}\}$
    $h_{\triangle}$ $g_{\triangle}$ (0, )
    E-  $e \in \mathcal{E}$
        $e' \in S_e$
            $x_{e'e}$ . ( )
            $y_{e'e}$ . ( )
        $e' \in S_e$
            $A_{e'e}$ . ( )
            $B_{e'e}$ . ( )
0           $D_{e'e}$ . ( )
        $e' \in R_e$
            $B_{ee'}$ . ( )
    -  $\triangle = 1$ 24
        $h_{\triangle}$ . ( )
        $g_{\triangle}$ . ( )
    Convergence
```

5 APPLICATIONS

, ff - , f f .

.

f f .

Follower maximization. v, f f k f S f v f f .

. i f f v. $u \notin S$, f f - f $(S \cup \{u\})$ (-), (S) f f S. ff (f. f), ff f f S v f ff . , f e' e . ( ). f S, f f

S (S) R . , f v .0. - , ff f f ff f f .

Followee maximization. f f - . ff f f f - - , f - f f f f f f - ff . , f f . , v, k f v f f v . f ff f .

Algorithm 2. / .

```
: $G = (V, E)$, $v$, $k$
: f /f $S$
    $S = \emptyset$ $R$ 0,000
    $i = 1$ to $k$
        $u \in V \backslash S$
            $s_u = 0$
            $r = 1$ to $R$
                $s_u += |FCM(S \cup \{u\})|$
            $s_u = s_u / R$
        $S = S \cup \{argmax_{u \in V \backslash S} s_u\}$
```

6 EXPERIMENTS

, f f f .

6.1 Experimental Setup

.

. , , - , f . f . , 0 , f f f , , , . f f , f / / 0 / / 0 0, .

, f .

t . .

, .

. f , f - . .

f f

. ff , -

ff , f

. f f

, f -

f ff .

ff -

f f f

f ff f .

f f f -

, , , . f f

Evaluation metrics. -

f f

f ,

f .

f -

, . ,

f -f f

. $\{h_\triangle\}$ $\{g_\triangle\}$

, f

$p(e|S_e)$ f

S_e . (0) .

. f

e f . e

f f $p(e|S_e) > \tau$. , , -

, ,

- -

ff f τ f 0 0. ,

f τ.

.

f f f .

f

f f . (-

f), , , 0

( ) f f -

f f

f f .

f

f . f -

$p(e|S_e)$ f

f .

Comparison methods.

. f -

.

Basic. f f

(. , - f

f).

SVM. f f -

- f -

f .

LRC. f f -

f f .

f

.

Collaborative filtering (CF): -

. f

f u, u

. f u

w, u f

w. -

, f u f v -

f f

$$CF\_score(u,v) = \sum_w I(w,v) sim(w,u),$$

$sim(w,u)$ w u,

. ., f -

$I(w,v)$ f w f v 0 -

. $\{v\}$ u

$CF\_score(u,v)$.

SimRank. f -

u f v

f f -

0 . $\{v\}$ f

u .

Katz. f

u f v -

f u v. f -

, .

.

Random-random model (RR). -

- .

, f u

f v, u w

f , w v -

f . e_{uv} -

(u,w,v) .

, f u f v

$$RR\_score(u,v) = \frac{1}{|F(u)|} \sum_w I(u,w) I(w,v) \frac{1}{|F(w)|},$$

$|F(u)|$ f f

u, $I(u,w)$ f u f w 0 -

. $\{v\}$ u

$RR\_score(u,v)$.

Preferential attachment with communities (PAC).

.

f u

f v, β,

u f f -

u $1-\beta$, u f

. α, v f

f (. ., f -

)

$1-\alpha$, v f f

. ,

f u f v

$$PAC\_score(u,v) = \beta\left(\alpha \frac{|N(v)|}{\sum_{v \in C(u)} |N(v)|} + (1-\alpha)\frac{1}{|C(u)|}\right) + (1-\beta)\left(\alpha \frac{|N(v)|}{\sum_{v \in V} |N(v)|} + (1-\alpha)\frac{1}{|V|}\right),$$

$|N(v)|$ f f f v. $C(u)$

f

f u,

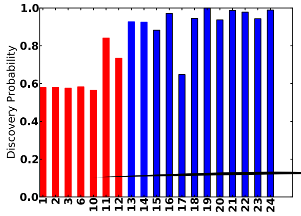
0 . V f f

f u, f f

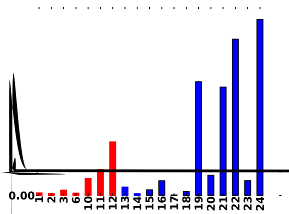
u. f-

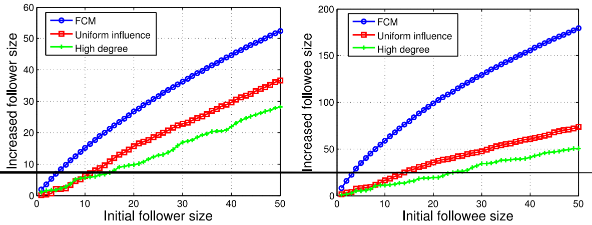
f fα β f 0 0. .

-



(a) Discovery Probabilities





(a) Follower maximization (b) Followee maximization

E . 11. [illegible]

-

,

.

f .

Per-triad analysis.

f

f ff f , , f

f ff

ff ff ff . . f -

f - . -

-

. , f -

f

. f

ff ff , - .

, f

ff ff , ff -

, , , ,

ff ff . -

. f

ff f ,

ff f

. ,

ff f f f ff f ,

f

.

Model parameter analysis. ,

ff -

f f ff -

, f f -

ff . f . -

f f ff

f ff ,

f ff

f ff .

f ff

, f ff ,

ff , f

- ff , ff . .

ff

f $r_\triangle'$, ff -

ff f ff

- f ff .

Delay analysis. δ

f f .

δ ff f f -

. . 0 - f

f δ f f ff

. , f -

.

ff ff f

, -

. f , δ f

.

Convergence analysis. f -

f . . 0

. -

. f

f .

6.3 Application Improvement

f f ff f ,

ff

f f

f .

k f v

f f

v. k f -

v f -

f f v.

.

, f /f

f f /f . -

f -

. f /f

f (- -

). f

f /f -

ff -

f 0.0 . f k f /f

, ff

f .

. , 0 ,

f /f

(+ − 0).

ff . -

f ff

,

ﬁ .
ﬁ / ﬁ
ﬀ . -
ﬀ ﬀ ﬀ
ﬀ ﬁ / ﬁ
ﬀ ﬁ .

7 RELATED WORK

Diffusion model and influence maximization.
-
ﬁ
.
, ﬀ , -

REFERENCES

, *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 00 , .

Proc. 4th ACM Int. Conf. Web Search Data Mining, 0 ,

, *Science*, . , . 0 , .

, *Proc. IEEE 12th Int. Conf. Data Mining*, 0 , . 0.

, *Nature*, . , . , 0 .

Proc. 11th SIAM Int. Conf. Data Mining, 0 , . 0.

f. f. ., , , 0 .

Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 0 0, . 0 0 .

, *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 00 , . 0 .

, *Phys. Rev. E*, . 0, . , . 0 , 00 .

, *Proc. 16th Int. Conf. World Wide Web*, 00 , . 0.

, *Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 00 , . .

, *Networks, Crowds, and Markets: Reasoning about a Highly Connected World.* , U. . U . , 0 0.

, *Tsinghua Sci. Technol.*, . , . , . , 0 .

, *Permutation, Parametric and Bootstrap Tests of Hypotheses*, . . , , U , 00 .

, *Proc. 3rd ACM Int. Conf. Web Search Data Mining*, 0 0, . 0.

, *Amer. J. Sociol.*, . , . , . 0 0, .

, *Proc. 13th Int. Conf. World Wide Web*, 00 , . 0 .

, *Proc. 11th SIAM Int. Conf. Data Mining Workshop*, 00 .

, *Proc. 8th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 00 , . .

, *Psychometrika*, . , . , . , .

, *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 00 , . .

, *Proc. Roy. Soc. A*, . , . 00 , .

, *Proc. ACM Conf. Inf. Knowl. Manage.*, 0 0, . 0 .

, *Intell. Data Anal.*, . , . , 0 .

, *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 0 , . .

, *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 00 , . 0.

, *Proc. 21st Int. Conf. World Wide Web*, 0 0, . 0.

, *Proc. 11th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 00 , . .

, *Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 00 , . 0 .

, *Math. Biosci.*, . 0, . , .

, *Proc. ACM Int. Conf. Web Search Data Mining*, 0 , . .

, *J. Amer. Soc. Inf. Sci. Technol.*, . , . , . 0 0 , 00 .

, *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 0 0, . .

, *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 0 0, . .

, *Data Mining Knowl. Discovery*, . , . , . , 0 .

, *Sci. Technol.*, . , . , . , 0 .

, *ACM Trans. Knowl. Discovery Data*, . , . , . , 0 .

, *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 0 , . .

, *Phys. Rev. E*, . , . , . 0 0 , 00 .

, *Computer Intensive Methods for Testing Hypotheses.* , , U , .

, *Proc. 4th Int. AAAI Conf. Weblogs Social Media*, 0 0, . .

, *Int. J. Semantic Web Inf. Syst.*, . , . , . 0 , 0 .

, *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 0 , . 0 0 .

, *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 00 , . 0 .

, *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 0 , . .

, *Proc. 6th Int. Conf. Data Mining*, 00 , . .

, *Proc. 7th IEEE Int. Conf. Data Mining*, 00 , . .