, Member, IEEE, , Senior Member, IEEE

Abstract— 24 .

Index Terms—

1 INTRODUCTION

A C, A' B C, A, B C A B A C t', B C, δ t' t' t $0 \leq t - t' \leq \delta$, δ A B, B C A C B, B C A' C' A, C

00 (. , .). . . , . . ( ). (. .). A C B C. . , . . . , , . , 0 , (. .,).

- 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

A

“following” link cascade model.

Organization.

2 “Following” Link Cascade Model

neighboring links

$G = (V, E, t)$, $v \in V$

$e_{uv} \in E$

u v. e A

B, A B

$t : E \rightarrow \cup \{\perp\}$

$t(e_{uv}) = n \in$ e_{uv} n, $t(e_{uv}) = \perp$

e_{uv}

$t(e_{uv})$ t_e. e' A

C, e B C.

A . 1. *Diffusion effect between links decays over time.*

discovery probability $g_{e'e}$ B

A C *diffusion probability* $h_{e'e}$

A C

B C

e' t', t'

$t' + \delta$, e

e' $g_{e'e}$,

e'

e $h_{e'e}$. e'

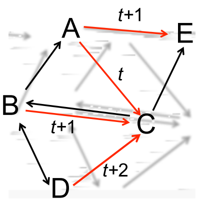
e

λ

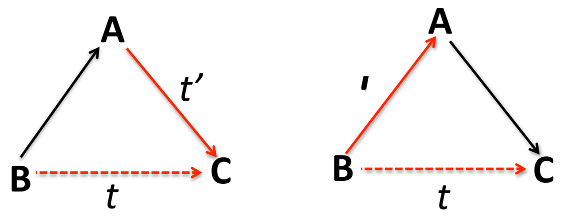
$g_{e'e}$

$t' + \lambda$ e' e.

δ



. 2.



(a) Follower diffusion

(. ., $\delta = 7$),

e, e

e',

e' B (

e) B

A C,

B A C

A C.

B C

C C.

. e_{AC}

t, B e_{BC}

e_{AC} $t+1$ e_{AC} e_{BC} $t+1$.

, e_{BC} e_{DC} $t+2$.

A e_{AE} e_{AC}

$t+1$ e_{AC} e_{AE} $t+1$.

( )

, . .,

Follower diffusion. A C

t', B C

t $t' \leq t \leq t' + \delta$, B A'

$t' - 1$,

e_{AC} e_{BC}.

C

follower diffusion.

. (A, B, C) e_{BA}

$t' - 1$.

, . ., $t' - 1$, B

A, C B,

Followee diffusion. B A

t', B C

t $t' \leq t \leq t' + \delta$, C A'

($t' - 1$),

e_{BA} e_{BC}.

B

followee diffusion.

, . ., e_{AC} e_{CB}.

3 Data and Observations

3.1 Data Collection

0,000

0/ / 0 0 / / 0 0.

,0

0

0/ / 0 0 / / 0 0.

( )

3.2 Observations

1 12

13 24

( )

( )

t' t',

( ) t

$0 \leq t - t' \leq \delta$ (δ

).

$C_\triangle$

$\triangle$, $C^+_\triangle$

B C $[t', t' + \delta]$. $|C_\triangle|$

$\triangle$. $r_\triangle$

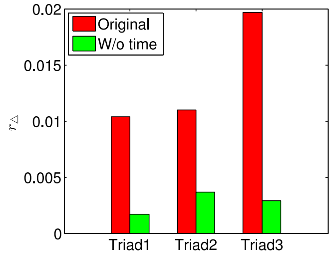
B C

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

-

-

Pattern significance.



(

≤ 0.05,

$r_\triangle$ δ

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

. ( ). $r_\triangle$.

δ

0,000

0, $r_\triangle$ ( )

0.0 ,),

B C C'

B B

A C. 0,

e_{AC}.

A C C

C

e_{AC}

Diffusion decay.

δ $1, 2, 3, 5, 7$ 10

δ, $r_\triangle$

$r_\triangle$

δ $r_\triangle$

$r_\triangle$

(.),

B A C,

C'

(.), B A C

B C'

C A

A' C' B

C

δ

Other observations.

A B.

(0-).

e_{BA},

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

A B

B C

A B (+).

A C.

A C

B C

C A (+ − 0).

A'

B A B

A' C A'

C' A' C

B C.

Summary.

(+) - -
ﬁ A C ﬀ -
ﬁ Aﬁ C Bﬁ C (+ − 0).

4 Model Learning

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

(e', e). ﬀ $h_{e'e}$
$g_{e'e}$ ﬁ (e', e) ﬁ
(ﬁ e , ( ) . ( )- . ( ) . ( ) 0.0 0 (0 . (00 (.)00 0 . 0.

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

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

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

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

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

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

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

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

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

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

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

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

.

A 1. .

: $G = (V, E, t)$

: $\theta = \{h_\triangle, g_\triangle\}$

$h_\triangle$ $g_\triangle$ (0,)

E- $e \in E$

$e' \in S_e$

$x_{e'e}$. ( )

$y_{e'e}$. ( )

$e' \in S_e$

$A_{e'e}$. ( )

$B_{e'e}$. ( )

0 $D_{e'e}$. ( )

$e' \in R_e$

$B_{ee'}$. ( )

- $\triangle = 1$ 24

$h_\triangle$. ( )

$g_\triangle$. ( )

Convergence

5 Applications

, , . . .

Follower maximization. v, k S v . . i v. $u \notin S$, - $(S \cup \{u\})$ (-), (S) S. (.), S v . , e' e . ( ). S,

S (S) R . , v .0. - , - . *Followee maximization.* - . - - , - - . , . v, k v v . .

A 2. / .

: $G = (V, E)$, v, k

: / 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

, .

6.1 Experimental Setup

. . , , - , . . , 0 , , , , . , / / 0 / / 0 0, . , . t . . , . . , - . .

Evaluation metrics.

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

$p(e|S_e)$

S_e . (0)

e e

$p(e|S_e) > \tau$.

τ 0 0.

τ.

0

$p(e|S_e)$

Comparison methods.

Basic.

SVM.

LRC.

Collaborative filtering (CF):

u, u u w, u w. u v

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

$sim(w,u)$ w u,

$I(w,v)$ w v 0

$\{v\}$ u

$CF\_score(u,v)$.

SimRank.

u v

0 . $\{v\}$

u

Katz.

u v

u v.

Random-random model (RR).

u

v, u w

w v

e_{uv}

(u,w,v)

u v

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

$|F(u)|$

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

$\{v\}$ u

$RR\_score(u,v)$.

Preferential attachment with communities (PAC).

u

v, β,

u

u $1-\beta$, u

α, v

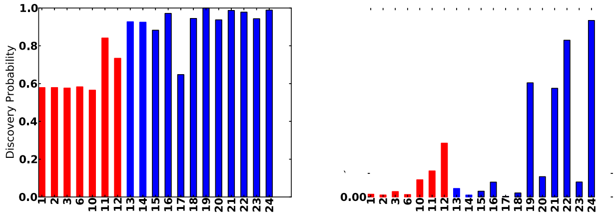
$1-\alpha$, v

u v

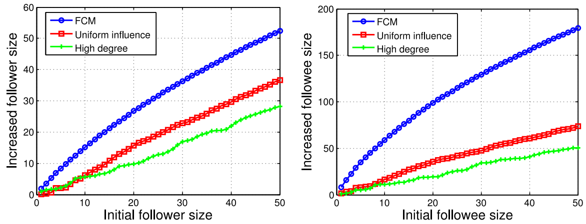
$$PAC\_score(u,v) = \beta\left(\alpha \frac{|N(v)|}{\sum_{v\in C(u)} |N(v)|} + (1-\alpha)\frac{1}{|C(u)|}\right) + (1-\beta)\left(\alpha \frac{|N(v)|}{\sum_{v\in V} |N(v)|} + (1-\alpha)\frac{1}{|V|}\right),$$

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

ʄ $u,$
$0 \,.\, V$ ʄ ʄ
ʄ $u,$ ʄ ʄ
$u.$ ʄ-
ʄ ʄ α β ʄ 0 $0.$.
-



(a) Discovery Probabilities



(a) Follower maximization (b) Followee maximization

. 11. . Y- . X-

Per-triad analysis.

Model parameter analysis.

Delay analysis. δ

δ 0

δ

δ

Convergence analysis.

0

6.3 Application Improvement

k v

v. k

v

v.

(-).

0.0 . k

0

(+ − 0).

7 Related Work

Diffusion model and influence maximization.

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.

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

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. 0 0.

Tsinghua Sci. Technol., 0

Permutation, Parametric and Bootstrap Tests of Hypotheses, 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

0 *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

0 *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

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

Computer Intensive Methods for Testing Hypotheses.

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