, , , *Member, IEEE*, , *Senior Member, IEEE*

Abstract—

Index Terms—

1 Introduction

00 (. , .).

A C,).

A' B C, A, B C

(. .).

A C A C

A B B C.

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

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

(. .,

).

2 "FOLLOWING" LINK CASCADE MODEL

neighboring links

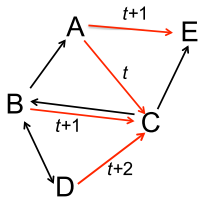
$G = (V, E, t)$, $v \in V$ $e_{uv} \in E$ u v. e A B, A B . $t : E \to \cup \{\perp\}$.

$t(e_{uv}) = n \in$ e_{uv} n, . $t(e_{uv}) = \perp$ e_{uv} , . $t(e_{uv})$ t_e. e' A C, e B C.

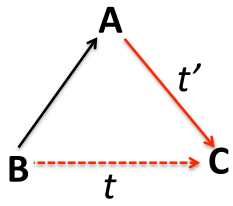
A **1.** *Diffusion effect between links decays over time.*

, *discovery probability* $g_{e'e}$ B A C , *diffusion probability* $h_{e'e}$ A C B C . , .

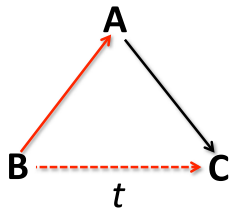
e' t', t' $t' + \delta$, e e' $g_{e'e}$, , e' e $h_{e'e}$. , e . λ $g_{e'e}$ $t' + \lambda$ e' e. δ



. 2.



(a) Follower diffusion



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

e, e

e',

B (e)

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 $t' \le t \le t' + \delta$, 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 $t' \le t \le t' + \delta$, 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

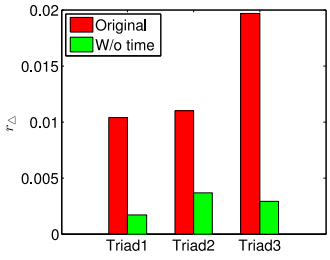
, ,
.

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

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

-
-
-

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



(

.I - ≤ 0.05,
.I ,
. U ,
. ( ).
, δ -
0,000 .I -
, - 0,
.I
, - , , ,
, 0.0 ,
. - .I ,
, , B C C
B B ,
A C. , 0,
, -
, -
e_{AC}.I , , ,
A C , C
. , -
C .
, -
.I , , , e_{AC}
.
. , ,
. , , , , ,
.
Diffusion decay. -
, δ 1, 2, 3, 5, 7 10 .
δ, $r_\triangle$ -
,
. . ,
$r_\triangle$.
δ , $r_\triangle$,
-
. -
, $r_\triangle$ -
. ,
(.),
B A C, A'
C' . -
(.), B A C
, B
C A
A' C' . , B -
C -
. , δ
.

. , -
$r_\triangle$ δ . ,
(. ., $t = \perp$), ,
$r_\triangle$. -
. . , -
, $r_\triangle$ ( )
( ),
. -
, . , ,
- -
.
Other observations. I , ,
- A B.
(- , -
0-).I ,
e_{BA}, - -
e_{AB}, e_{AB} e_{BA}.
-
A B
B C -
A B (+).
-
,
, ,
. I ,
-
A C.
(- , - ,
- , -)
. ,
- A C
B C
- C A (+ - 0).
B A , B -
A' C A'
C' , A' C
B C.
Summary. , ,
,
, ,
-

(+) -
ɬ A C ɬɬ -
ɬ A ɬ C B ɬ C (+ − 0).

4 MODEL LEARNING

I , ɬɬ ɬɬ
. ɬ
.

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

(e', e). ɬɬ $h_{e'e}$
$g_{e'e}$ ɬ (e', e) ɬ
(I ɬ e , ( ) . ( )- . ( ) . ( ) 0.0 0 ; . . .0 ; . . .

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

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

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

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

EM algorithm.

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

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

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

$Q(\theta, \hat{\theta})$

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

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

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

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

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

.

A 1. .

: $G = (V, E, t)$
: $\theta = \{h_{\triangle}, g_{\triangle}\}$
I $h_{\triangle}$ $g_{\triangle}$ (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

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

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

A 2. / .

: $G = (V, E)$, v, k
: I , / S
I $S = \emptyset$ 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

I ,
, ,
.

6.1 Experimental Setup

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

Evaluation metrics.

Comparison methods.

Basic.

SVM.

LRC.

Collaborative filtering (CF):

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

SimRank.

Katz.

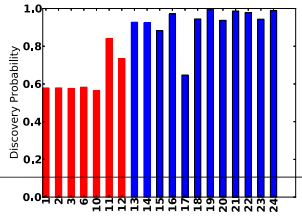
Random-random model (RR).

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

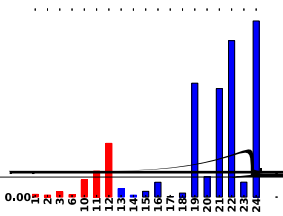
Preferential attachment with communities (PAC).

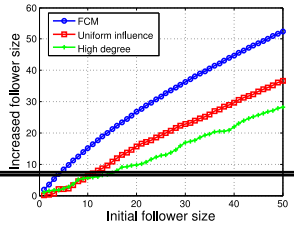
$$PAC\_score(u,v) = \beta\left(\alpha \frac{|N(v)|}{\sum_{v\in C(u)} |N(v)|} + (1-\alpha)\frac{1}{|C(u)|}\right) + (1-\beta)\left(\alpha \frac{|N(v)|}{\sum_{v\in V} |N(v)|} + (1-\alpha)\frac{1}{|V|}\right),$$

ξ u,
0 . V ξξ
ξ u, ξ ξ
u. ξ-
ξ ξ α βξ 0 0. .
-

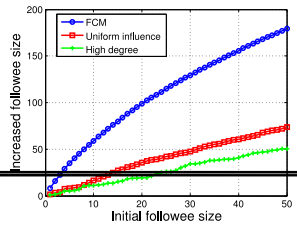


(a) Discovery Probabilities





(a) Follower maximization



(b) Followee maximization

Per-triad analysis.

Model parameter analysis.

Delay analysis.

Convergence analysis.

6.3 Application Improvement

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,

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

Nature,

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

Phys. Rev. E, 0 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

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, 0

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

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

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