BEHAVIORAL MODELING IN SOCIAL NETWORKS
FROM MICRO TO MACRO

Meng Jiang, *University of Illinois at Urbana-Champaign*
Peng Cui, *Tsinghua University*

ICDM 2015 TUTORIAL
Atlantic City, NJ

清華大學
Tsinghua University
1867 ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Thanks to

N S F C
National Natural Science Foundation of China

mda
Media Development Authority Singapore

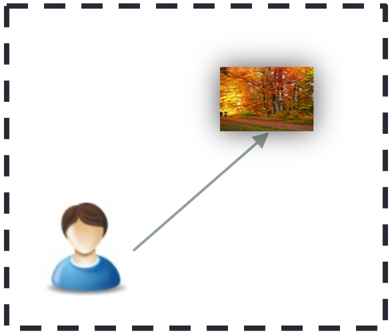
TSINGHUA UNIVERSITY- TENCENT JOINT LABORATORY

National Natural Science Foundation of China,
No. 61370022, No. 61210008, No. 61303075
International Science & Technology Cooperation Program of China,
No. 2013DFG12870
National Program on Key Basic Research Project,
No. 2011CB302206, No. 2015CB352300
NExT Research Center funded by MDA, Singapore,
WBS:R-252-300-001-490
Tsinghua-Tencent Joint Laboratory

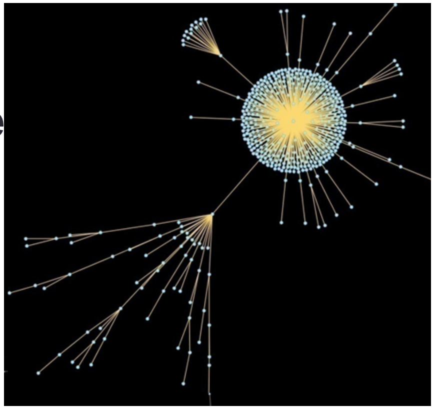
Outline

- **Prediction for natural behavior**
  - Modeling individual behavior (MICRO)
  - **Modeling information cascade (MACRO)**
- Detection for unnatural behavior
  - Suspicious behavior detection

From Micro to Macro



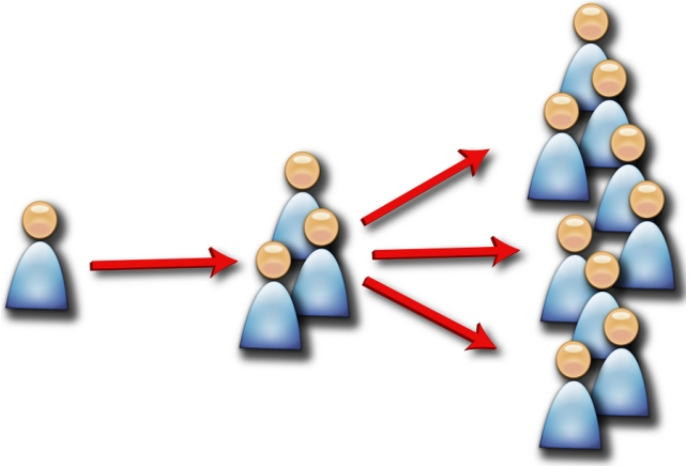
Dynamically Aggregate



Information spreading is a macro phenomenon which is driven by individual user behaviors in microscopic level.

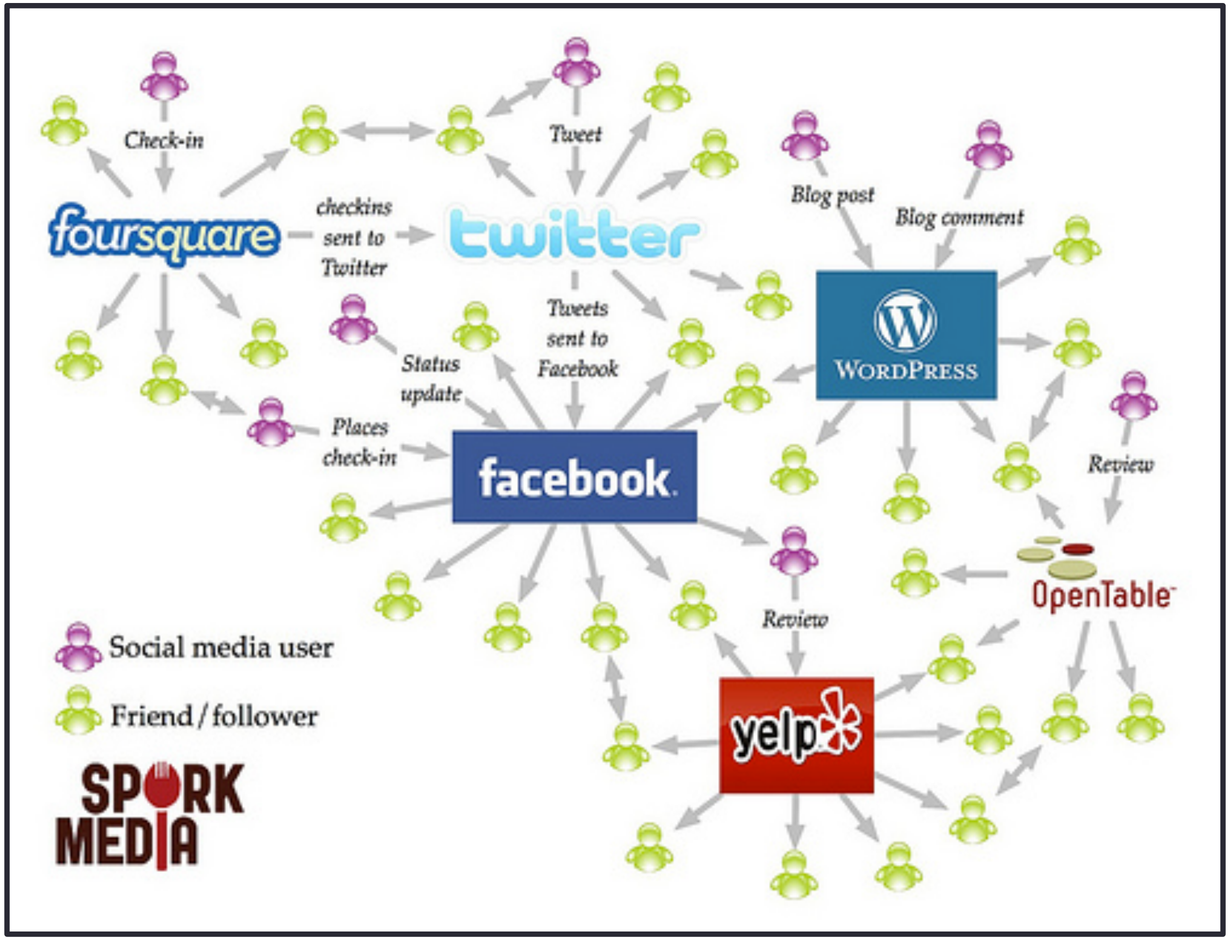
Cascades: Information Spreading

- In network environment, if decentralized nodes act on the basis of how their neighbors act at earlier time, ***cascades*** will be formed.
  - Word-of-mouth
  - Cascading
  - Diffusion
  - Propagation



Information Spreading is Ubiquitous

Social Media



Information spreading is the major way of communication in social media.

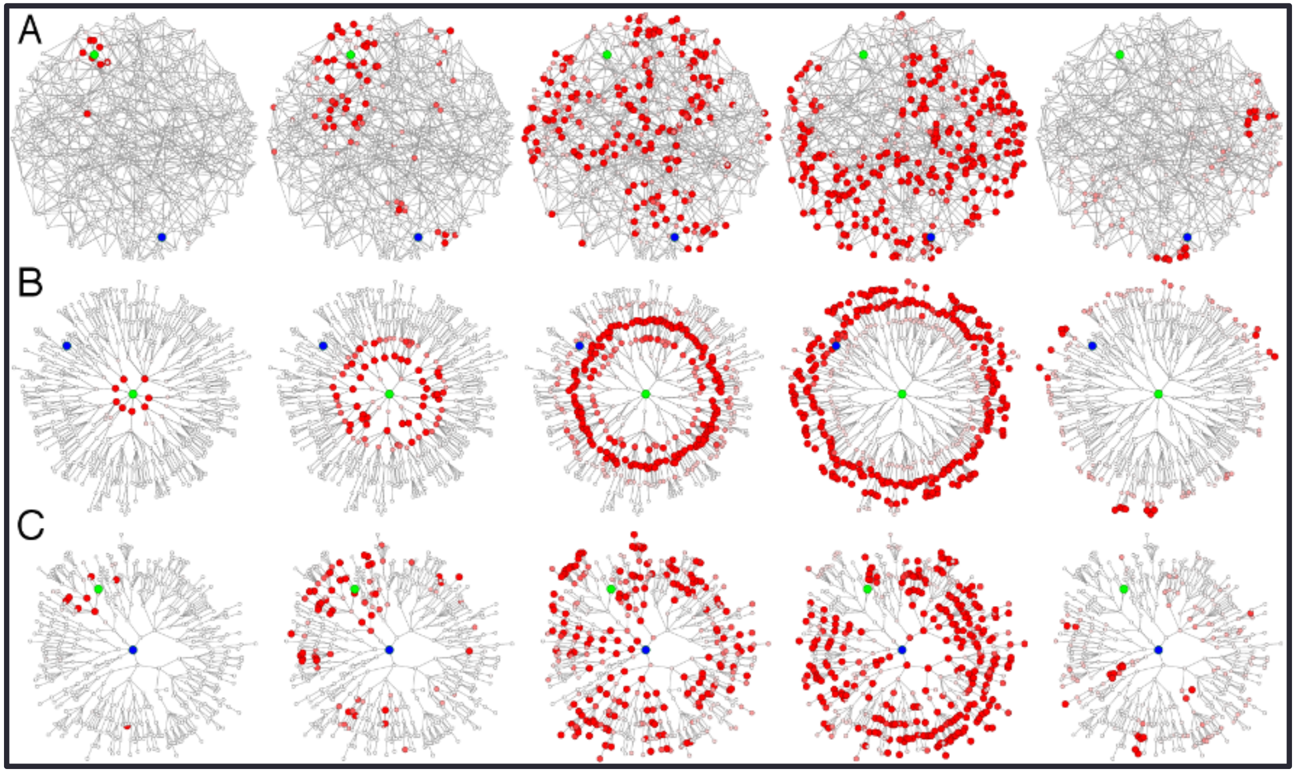
Information Spreading is Ubiquitous

Word-of-Mouth (Marketing)

How to fully exploit the power of word-of-mouth in marketing?

Information Spreading is Ubiquitous

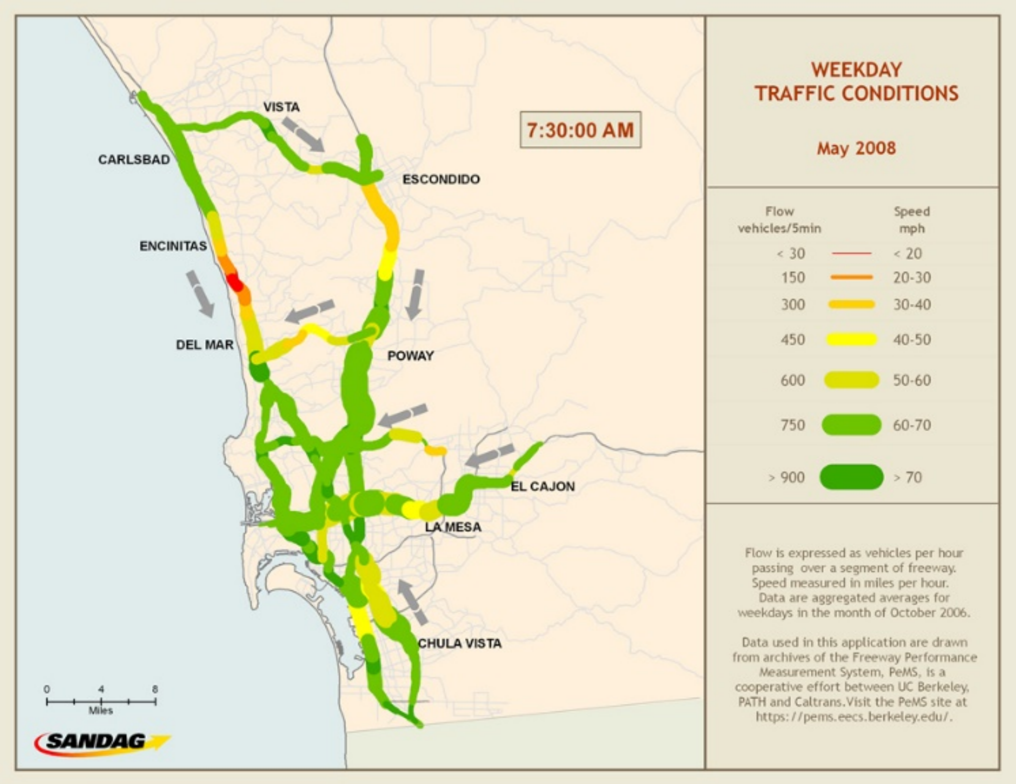
Epidemics



Share similar dynamic process as information spreading.

Information Spreading is Ubiquitous

Traffic



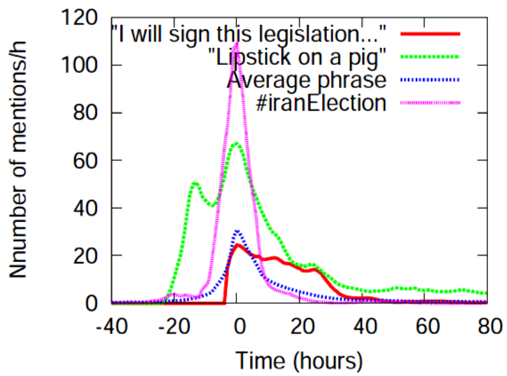
Traffic jams spread through road network. How to model, predict and intervene?

How to understand information spreading mechanism, and furthermore, predict the information spreading process?

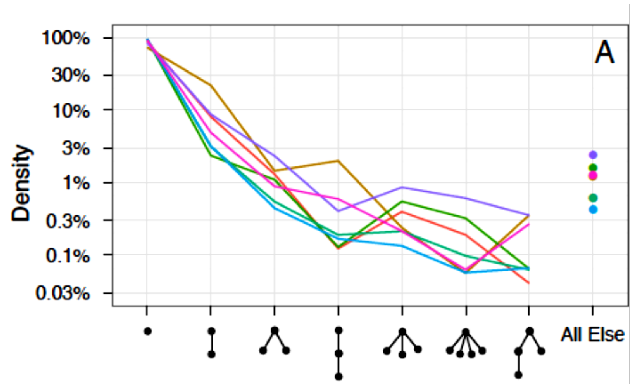
Related Research

Understand

Rise and fall patterns in cascading curve

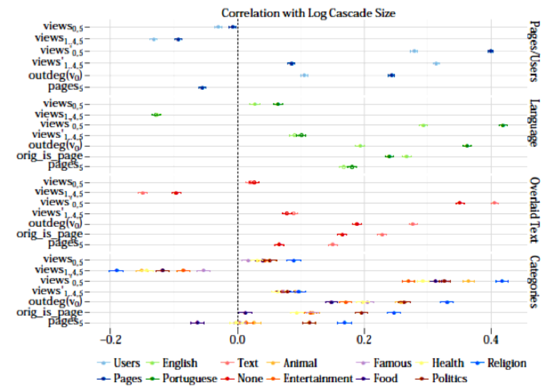


Structural units of information cascades



Prediction

Popularity prediction



Regard cascades as a whole and extract cascade-level features for understanding and prediction.

Macro Phenomenon v.s. Micro Mechanism

Information spreading is driven by a cascade of user adoption behaviors.

Micro

Behavioral Dynamics

Macro

Information Spreading

Behavior-Driven Information Spreading Modeling

Ultimate Goal: Bridge the gap between macro phenomena of information spreading and micro behavioral mechanism.

One-Hop Cascade Prediction

Predict the collective response of a user's followers

SIGIR'11, AAAI'11

Cascading Outbreak Prediction

Predict whether the information will break out in future

KDD'13

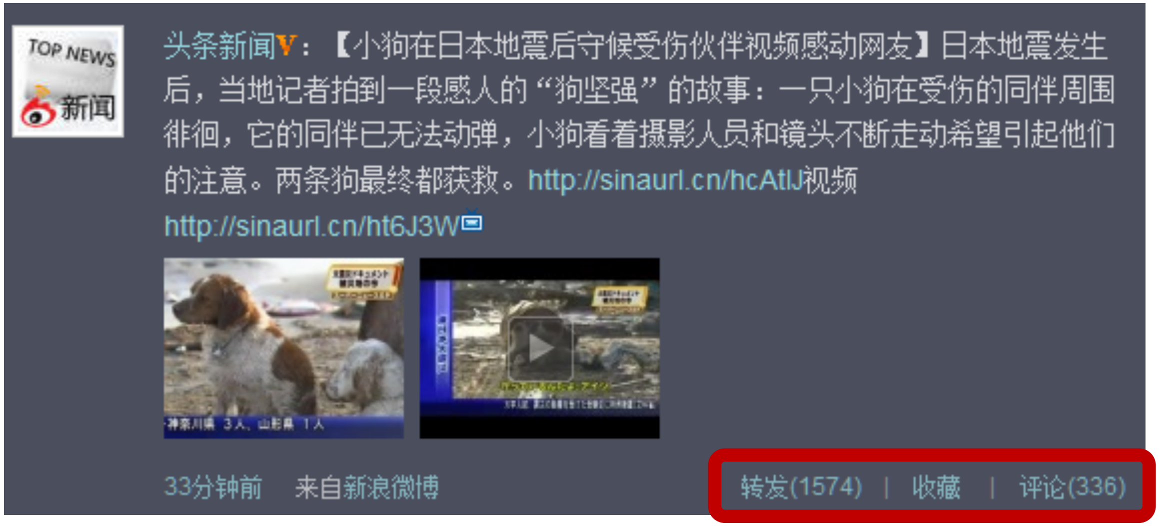
Dynamic Process Prediction

Predict the dynamic cascading process of a piece of information

ICDM'15

The problem:

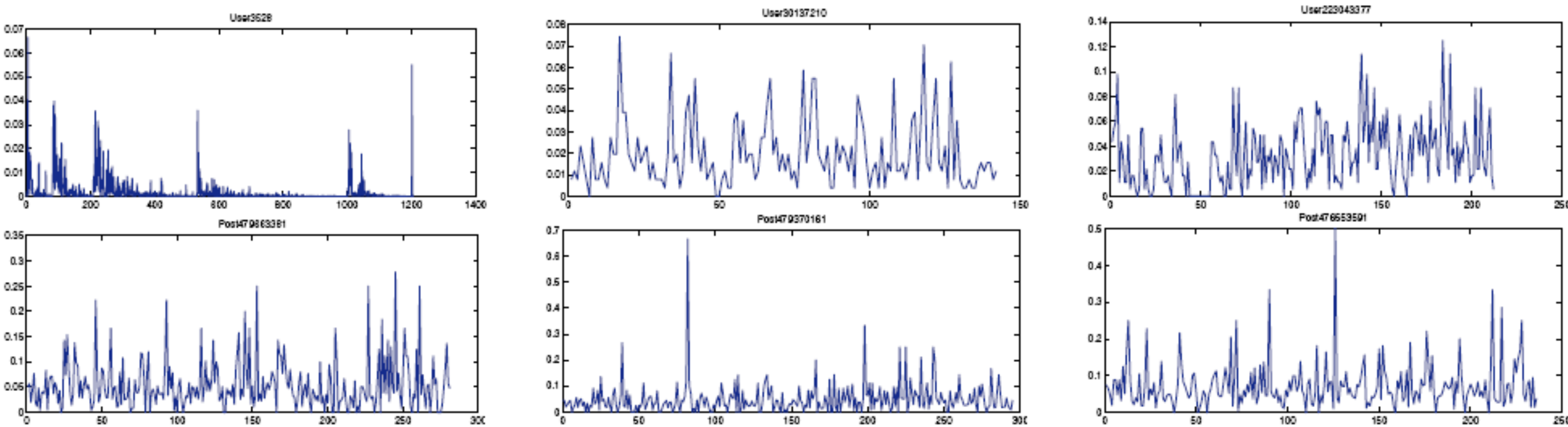
To predict the percentage of a user's followers that will retweet the microblog after the user retweet it.



The Dimensions

Are big users always trigger high forwarding numbers?

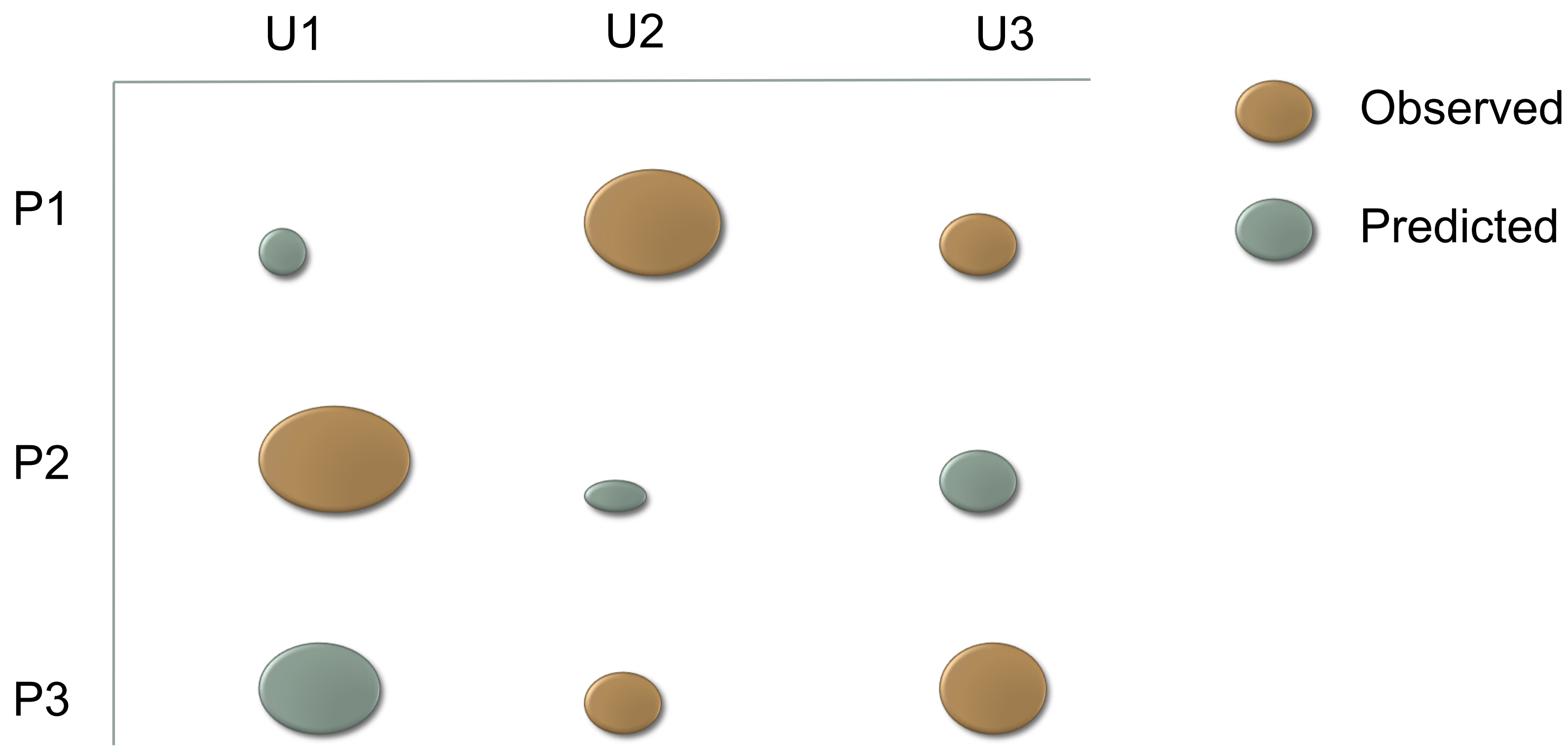
Post Variance



User Variance

Are popular tweets always trigger high forwarding numbers?

Problem Formulation



✓Given an user, rank the web posts to share

✓Given a web post, rank the users to target

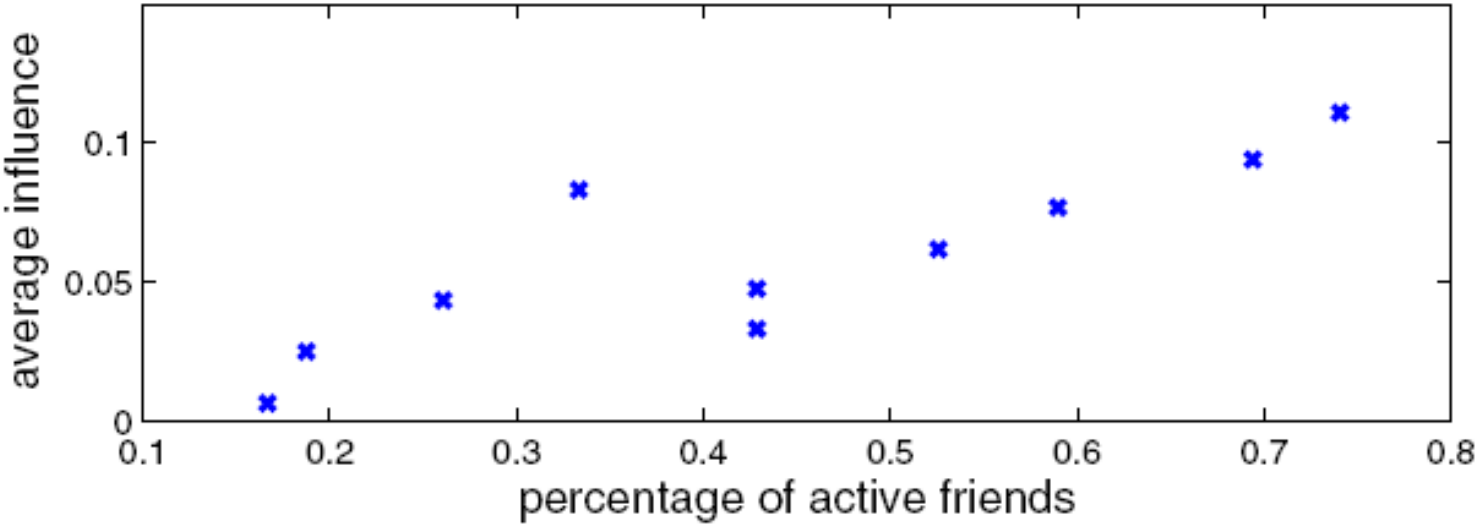
Density 0.1%

We need priors on users and posts.

Predictive Factors

Percentage of active friends

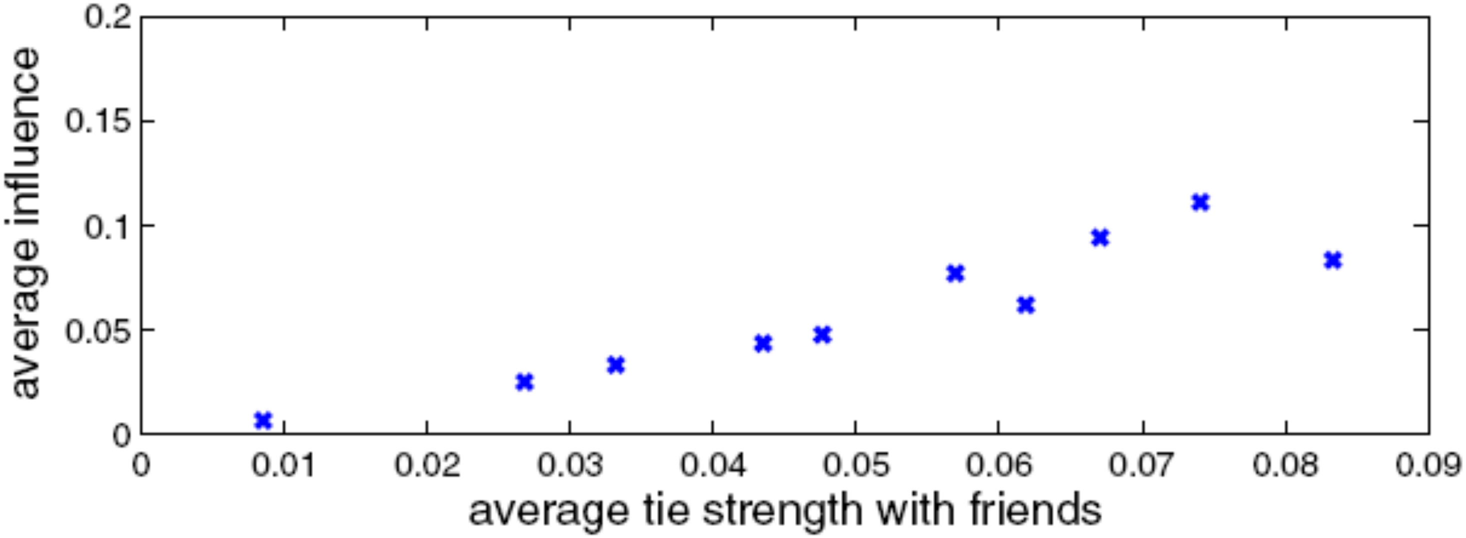
$$uf_1(u_i) = \frac{\sum_{u_r \in \mathcal{N}(u_i)} \delta(act(u_r) \geq \tau)}{|\mathcal{N}(u_i)|}$$



Predictive Factors

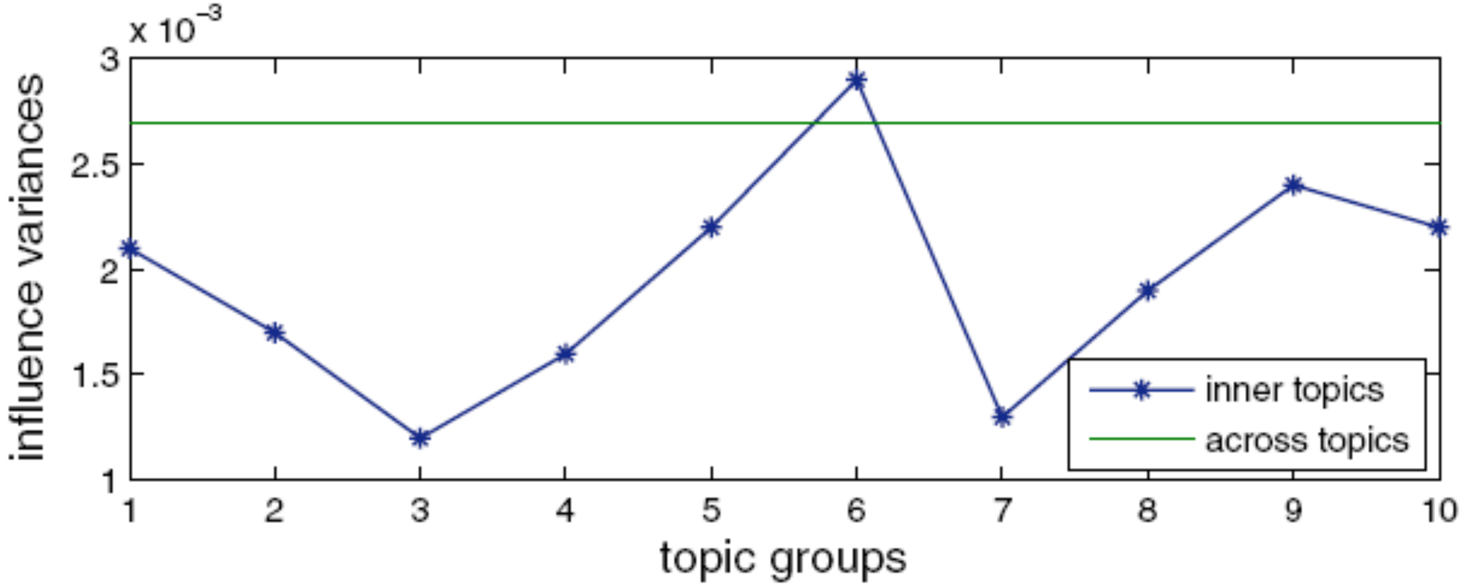
Average tie strength

$$uf_2(u_i) = \frac{\sum_{u_r \in \mathcal{N}(u_i)} \frac{tie(u_i, u_r)}{\sum_j Y_{ij}}}{|\mathcal{N}(u_i)|}$$



Predictive Factors

The introduction of post topic groups can reduce the variances of influences.



Modeling

Baseline objective function

$$\min_{\mathbf{U},\mathbf{V}} \left\|\mathbf{X} - \mathbf{U}\mathbf{V}^\top\right\|_F^2 + \gamma \|\mathbf{U}\|_F^2 + \delta \|\mathbf{V}\|_F^2$$
$$s.t. \mathbf{U} \geq 0, \ \mathbf{V} \geq 0$$

We suppose the users with similar observed predictive factors have similar distribution in latent space

$$\mathcal{J}_3 = \left\|\mathbf{W} - \mathbf{U}\mathbf{U}^\top\right\|_F^2$$

User similarity matrix

We constrain the latent post space by topic distributions

$$\mathcal{J}_4 = \left\|\mathbf{C} - \mathbf{V}\mathbf{G}^\top\right\|_F^2$$

Post content matrix

Topic matrix

Modeling

Hybrid Factor Non-Negative Matrix Factorization (HF-NMF)

$$\begin{aligned}\min_{\mathbf{U},\mathbf{V},\mathbf{G}} \quad & \left\|\mathbf{X}-\mathbf{U}\mathbf{V}^\top\right\|_F^2+\alpha\left\|\mathbf{W}-\mathbf{U}\mathbf{U}^\top\right\|_F^2+\beta\left\|\mathbf{C}-\mathbf{V}\mathbf{G}^\top\right\|_F^2 \\ & +\gamma\left\|\mathbf{U}\right\|_F^2+\delta\left\|\mathbf{V}\right\|_F^2 \\ s.t. \quad & \mathbf{U}\geq 0,\ \mathbf{V}\geq 0,\ \mathbf{G}\geq 0 \end{aligned} \tag{12}$$

Ranking Criterion

	User	Ranking	Post	Ranking
	η	ρ	η	ρ
HF-NMF	**0.8942**	**0.9389**	**0.8012**	**0.8697**
bNMF+UF	0.8739	0.9088	0.7423	0.8334
bNMF+PF	0.8236	0.8412	0.7654	0.8548
bNMF	0.813	0.8342	0.7358	0.7926
AvgU	0.7824	0.8056	0.7047	0.7583
AvgP	0.6973	0.7143	0.6746	0.736
CoxPH	0.6596	0.6893	0.659	0.6762
LR	0.6524	0.697	0.6328	0.6593

The advantages of HF-NMF is more apparent in ranking evaluations.

Examples

For a user, ranking the posts

PostIDs	8783	9993	6551	8169	3550	8698	1404	5655	7825	4459
RankOrder(groundtruth)	1	2	3	4	5	6	7	8	9	10
SocialInfluence(groundtruth)	73	53	53	33	13	13	13	13	6	6
RankOrder(Prediction)	1	3	2	4	9	6	7	8	5	10
SocialInfluence(Prediction)	65	43	44	31	12	20	15	14	25	9

For a post, ranking the users

UserIDs	2627	1287	2336	2952	4466	2764	3052	0893	7666	4909
RankOrder(groundtruth)	1	2	3	4	5	6	7	8	9	10
SocialInfluence(groundtruth)	33	26	19	19	13	13	6	6	6	6
RankOrder(Prediction)	4	1	2	3	5	6	7	8	9	10
SocialInfluence(Prediction)	16	27	19	17	13	11	7	6	6	6

Discussions

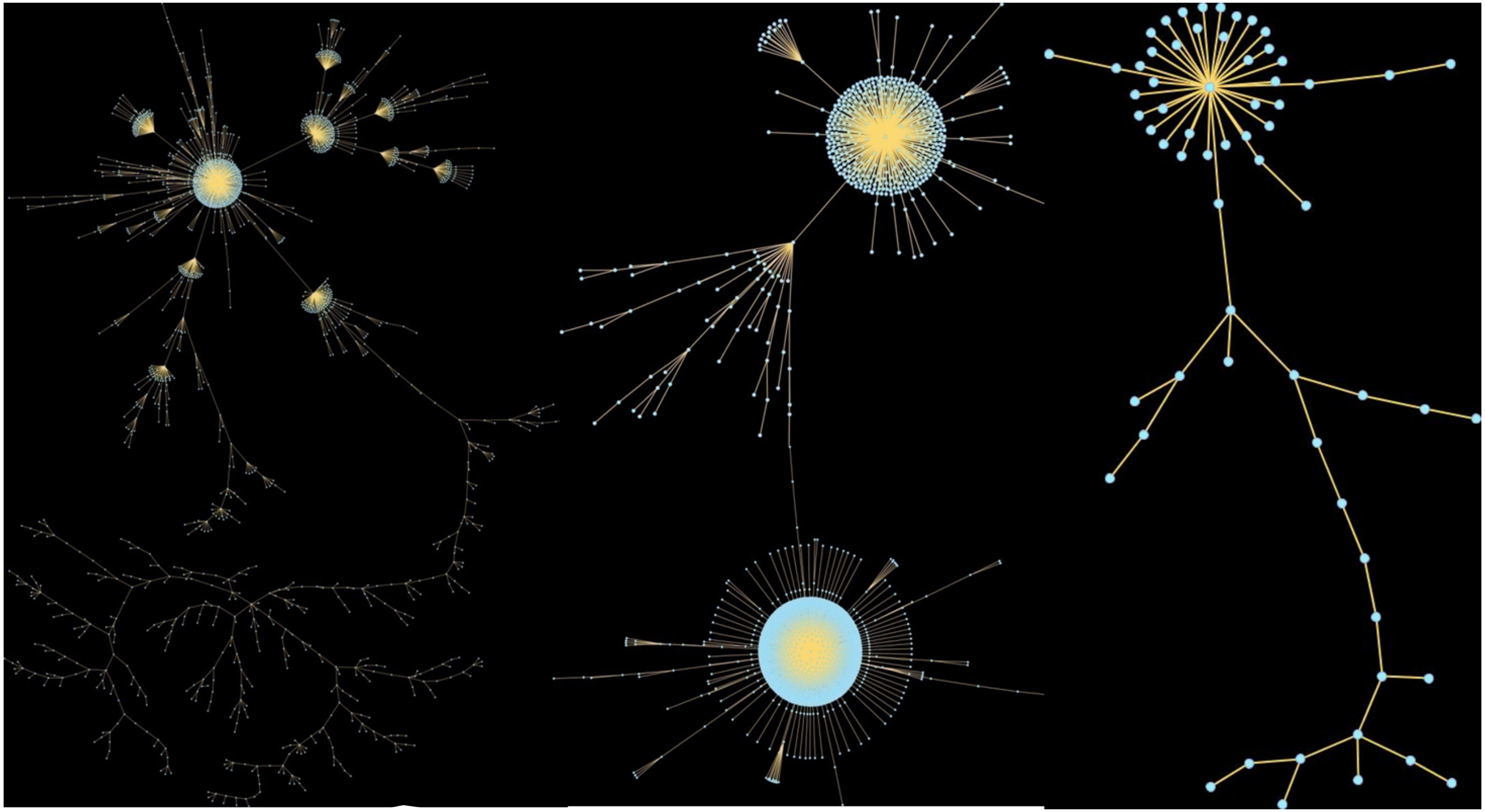
- The collective retweeting behaviors of a user's followers is predictable in fine granularity.
- Can we use the results of one-hop cascade prediction to predict the whole cascades? No!
  - Inapplicable in real applications
  - Error aggregation
- **Hint:** Different users play different roles in information spreading.

Predictive Modeling on Information Spreading

Ultimate Goal: Bridge the gap between macro phenomena of information spreading and micro behavioral mechanism.

One-Hop Cascade Prediction	Cascading Outbreak Prediction	Dynamic Process Prediction
Predict the collective response of a user's followers	Predict whether the information will break out in future	Predict the dynamic cascading process of a piece of information
SIGIR'11, AAAI'11	KDD'13	ICDM'15

Cascading Outbreak Prediction



Can we **predict** whether a tweet will be hot in future?

Outbreak prediction

- Basic Hypothesis: User behaviors cause outbreaks
- Experience: Different users play different roles in causing outbreaks
- How to identify the important users?
  - Topology measures
    - Indegree, centralities, etc.
  - Influential nodes
    - Suppose the cascading process

But does the real data follow the hypothesized cascading process and topology measures?

A Data Driven Approach

- ➢Mining from massive historical data

Early stage users as features

Paths as samples

$$M\begin{pmatrix}1 & 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0 & 0\\ 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 1 & 0\end{pmatrix} \rightarrow Result\begin{pmatrix}1\\0\\1\\0\\1\end{pmatrix}$$

Outbreak signals

Selected as sensors
to predict outbreaks

Challenges

- The outbreak prediction and node selection procedures need to be jointly optimized
- The node selection need to be parsimonious so that the monitoring over the selected sensors can be cost effective
- The node selection process need to be efficient so that the method can be applied into large realistic networks

Orthogonal Sparse LOgistic Regression (OSLOR)

$$L(\boldsymbol{\theta}) = h(\mathbf{X}_{i\cdot})^{y_i} \cdot (1 - h(\mathbf{X}_{i\cdot}))^{1-y_i}$$

$$\log L(\boldsymbol{\theta}) = -\sum_{i=1}^{m}(\log(1 + e^{\mathbf{X}_{i\cdot}\boldsymbol{\theta}})) + \mathbf{y}^\top \mathbf{X}\boldsymbol{\theta}$$

$$F(\boldsymbol{\theta}) = T_1(\boldsymbol{\theta}) + T_2(\boldsymbol{\theta}) + T_3(\boldsymbol{\theta})$$

$$T_1(\boldsymbol{\theta}) = -\log L(\boldsymbol{\theta})$$

$$T_2(\boldsymbol{\theta}) = \frac{\beta}{4}\sum_{i,j}(\theta_i \mathbf{X}_{\cdot i}^\top \mathbf{X}_{\cdot j}\boldsymbol{\theta}_j)^2$$

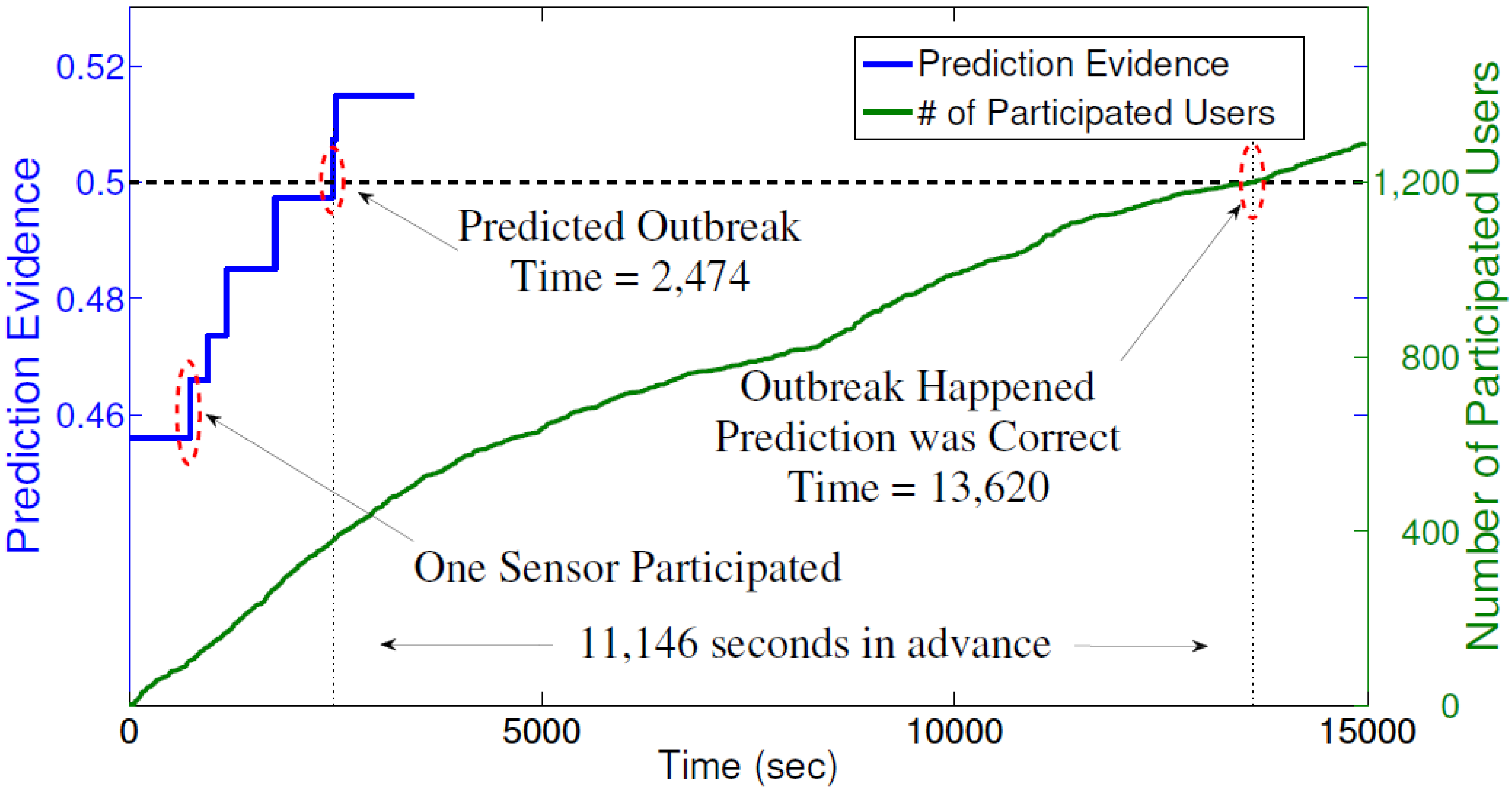
$$T_3(\boldsymbol{\theta}) = \gamma||\boldsymbol{\theta}||_1$$

Algorithm 1 Orthogonal Sparse LOgistic Regression (OSLOR)

Require: Tradeoff parameters $\beta > 0$, $\gamma > 0$, Radius $R > 0$, Cascade status matrix $\mathbf{X}$, Cascade outbreak indicator vector $\mathbf{y}$, Step size $c > 0$

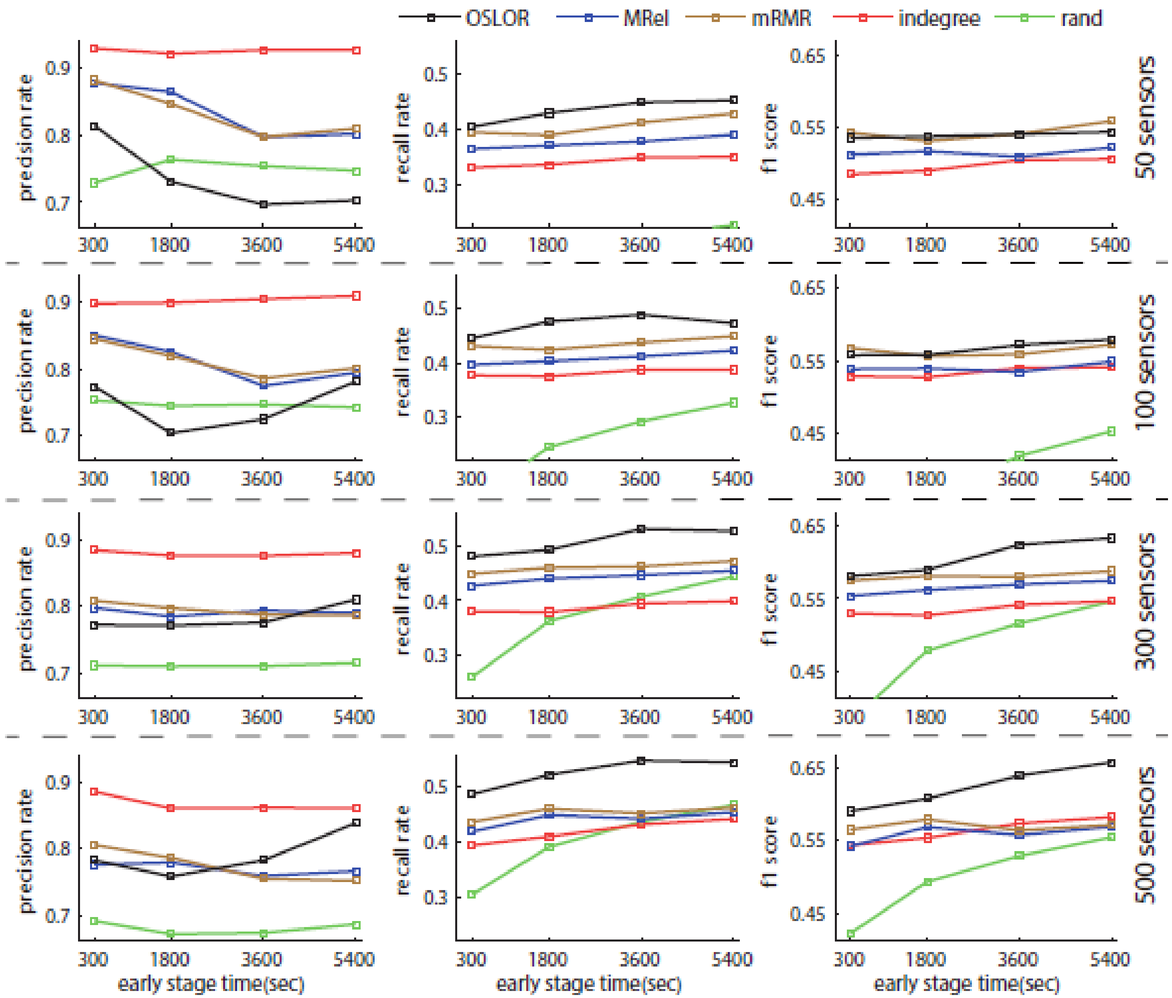
1: Calculate the inner product matrix $\mathbf{X}^\top \cdot \mathbf{X}$
2: Initialize the coefficient $\boldsymbol{\theta}^0 \leftarrow \mathbf{0}$
3: Calculate the current value of object function using Eq. (5) $F^0 \leftarrow F(\boldsymbol{\theta}^0)$
4: Initialize the iteration variable $k \leftarrow 0$
5: **repeat**
6: Calculate gradient $\nabla g(\boldsymbol{\theta}^k)$ using Eq. (9) and Eq. (10)
7: Update $\boldsymbol{\theta}^{k+1}$ using Eq. (17)
8: Update the value of object function $F^{k+1} = F(\boldsymbol{\theta}^{k+1})$
9: **if** $F^k \leq F^{k+1}$ **then**
10: $R \leftarrow R \cdot c$, continue;
11: **else**
12: $k \leftarrow k + 1$
13: **end if**
14: **until** converged
15: **Output:** The final coefficient $\boldsymbol{\theta}^k$

A Showcase



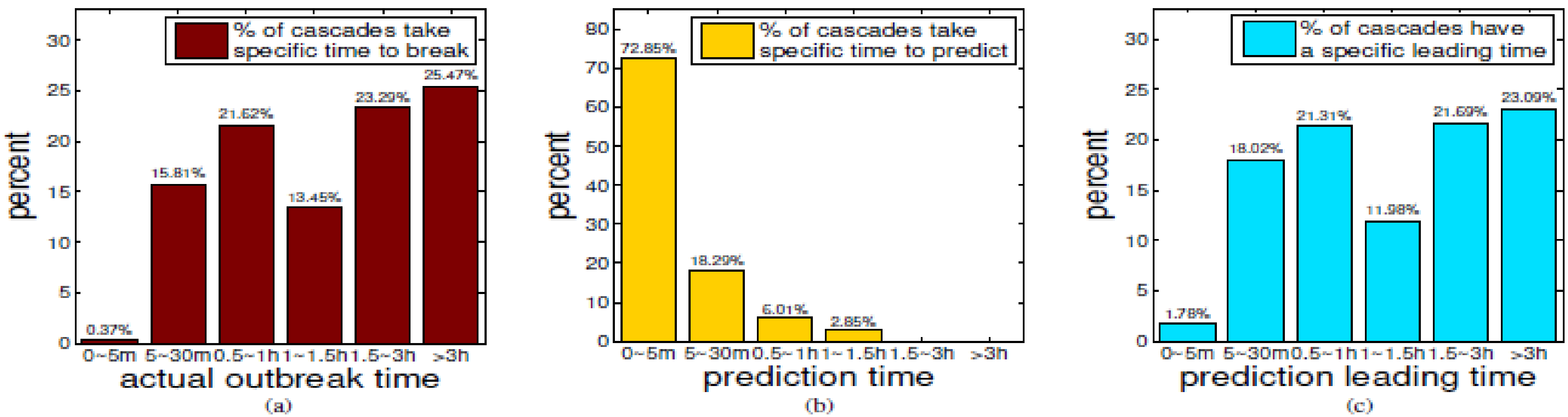
Prediction Evidence
# of Participated Users
0.52
0.5
0.48
0.46
Prediction Evidence
1,200
800
400
0
Number of Participated Users
Predicted Outbreak
Time = 2,474
Outbreak Happened
Prediction was Correct
Time = 13,620
One Sensor Participated
11,146 seconds in advance
0
5000
10000
15000
Time (sec)

Prediction Performance



- ➢ Our approach performs best
- ➢ Data driven approaches outperforms topology-based approaches
- ➢ Big nodes' participation will cause outbreaks in most cases
- ➢ Only a part of outbreaks are caused by big nodes

Prediction Leading Time



We only need 5 mins to predict the information outbreaks!

Peng Cui, Shifei Jin, Linyun Yu, Fei Wang, Shiqiang Yang, Cascading Outbreak Prediction in Networks: A Data-Driven Approach, ***ACM SIGKDD 2013***. (Full Paper)

Discussions

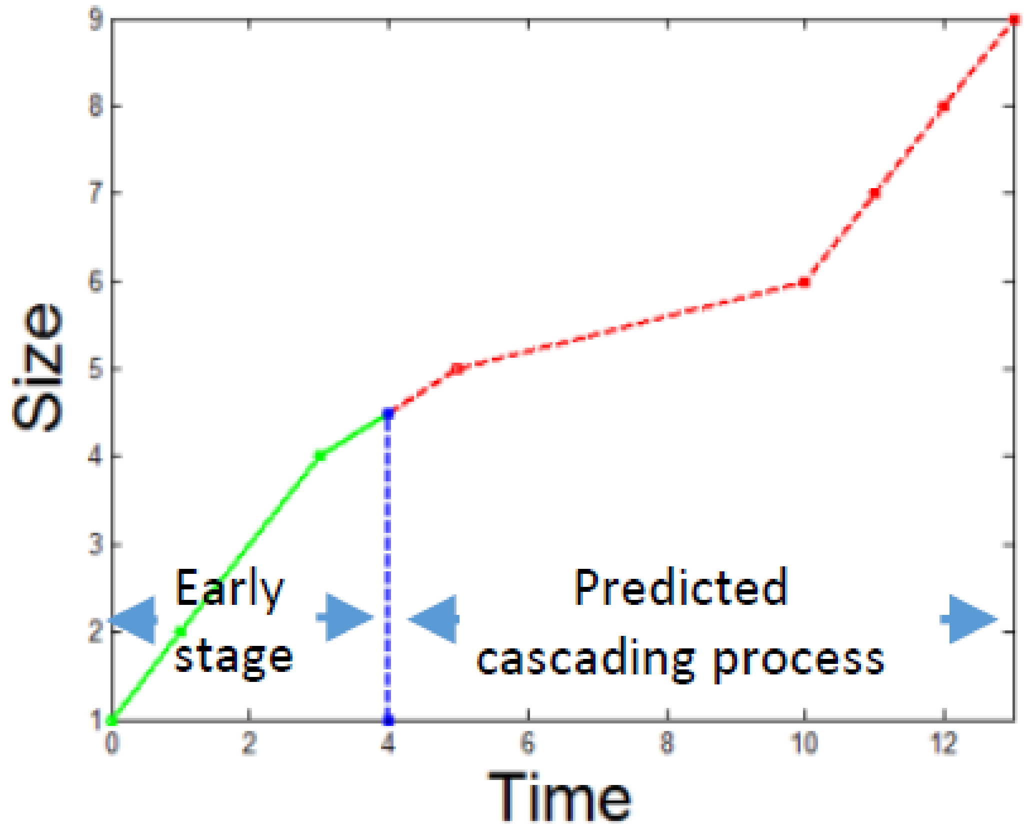
- Studying information spreading from user behavior angle is effective and promising.
- Many traditional hypothesis on the node importance and diffusion mechanism are not consistent with the real data.
- This is a one-shot study. Can we make continuous prediction on the information spreading?

Predictive Modeling on Information Spreading

Ultimate Goal: Bridge the gap between macro phenomena of information spreading and micro behavioral mechanism.

One-Hop Cascade Prediction	Cascading Outbreak Prediction	Dynamic Process Prediction
Predict the collective response of a user's followers	Predict whether the information will break out in future	Predict the dynamic cascading process of a piece of information
SIGIR'11, AAAI'11	KDD'13	ICDM'15

Beyond Cascade Size…



Time:
When will a cascade break out?

Size-Time:
How about the momentum of a cascade?

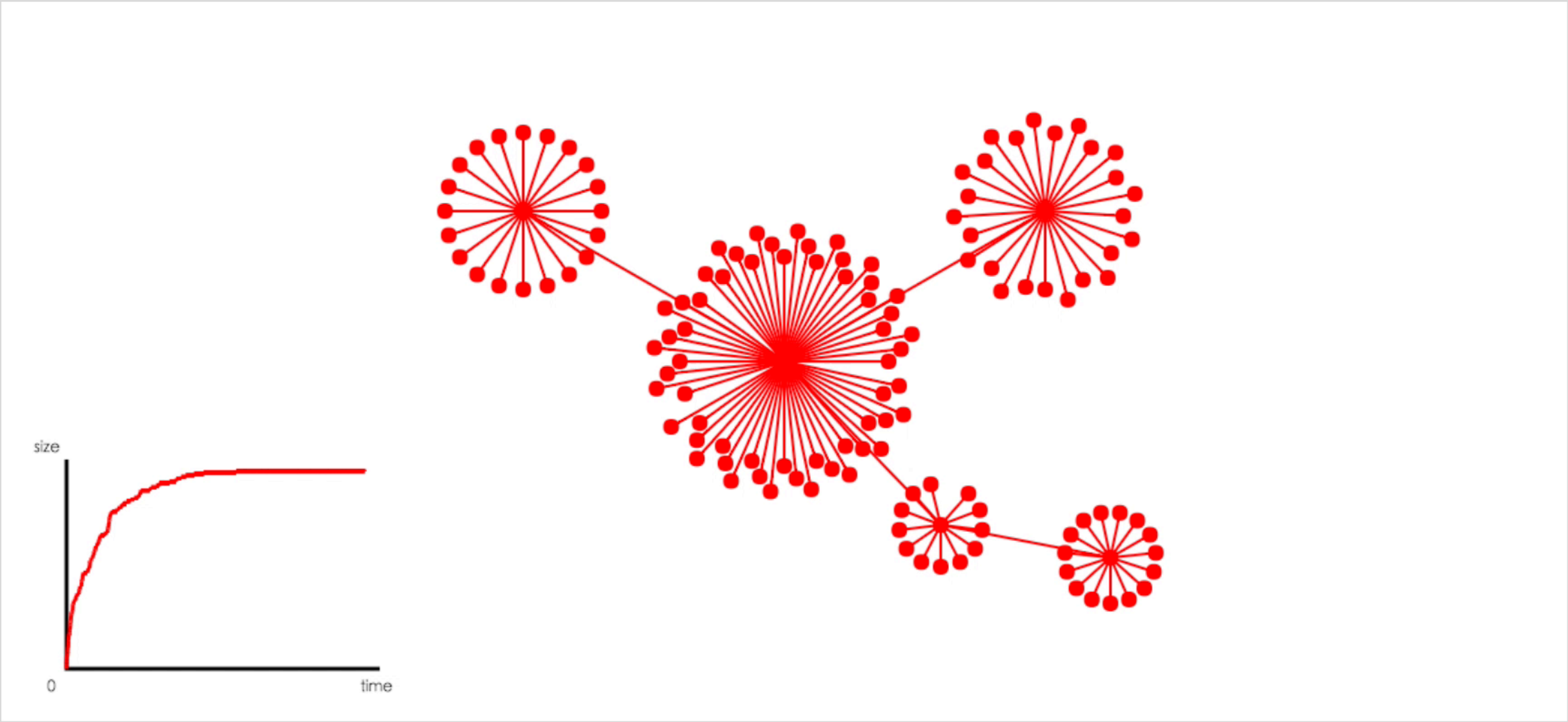
Cascading Process Prediction

Challenge: Cascade-level macro features do not work.
Content feature and structure feature
are not distinctive and predictive enough.

From Micro to Macro: Subcascades

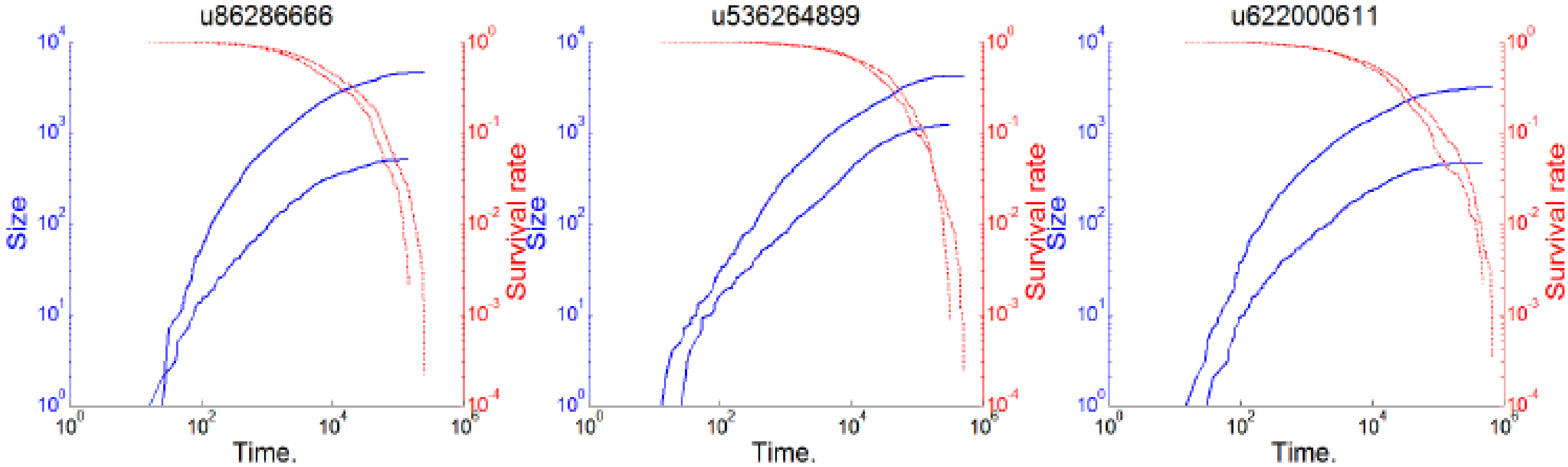
How to model subcascades?

How to connect subcascades and cascade?



Behavioral Dynamics

Behavioral Dynamics capture the changing process of the cumulative number of a user's followers retweeting a post after the user retweet the post.

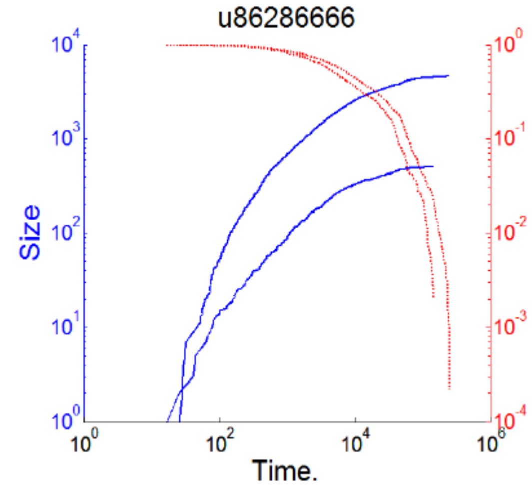


Survival Rate represent the percentage of nodes that has not been but will be infected.

Behavioral dynamics can be well represented by survival function.

Parameterize Behavioral Dynamics

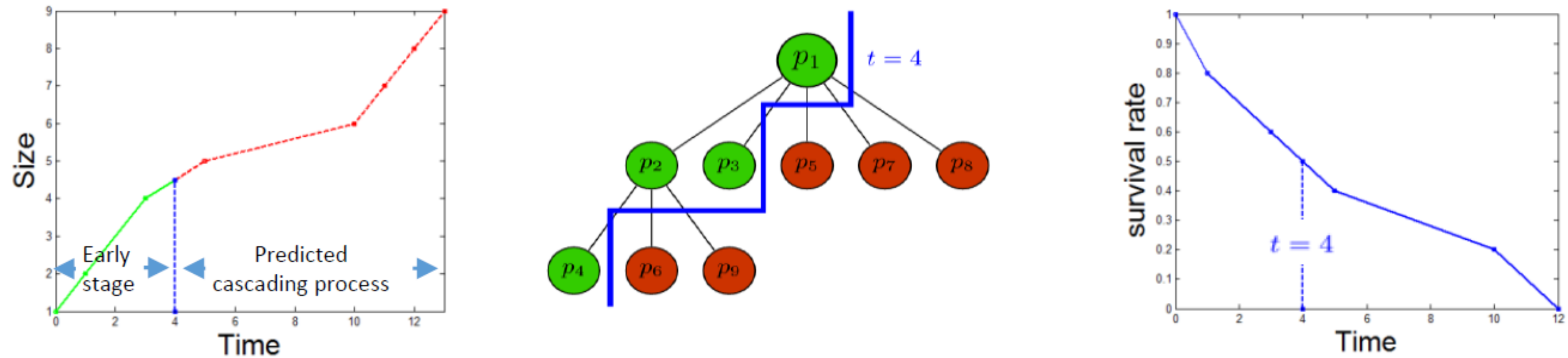
model	ks-statistic in Weibo
Exponential	0.2741
Rayleigh	0.7842
Weibull	0.0738



model	density function	survival function	hazard function
Exponential	$\lambda_i e^{-\lambda_i t}$	$e^{-\lambda_i t}$	λ_i
Rayleigh	$\alpha_i t e^{-\alpha_i \frac{t^2}{2}}$	$e^{-\alpha_i \frac{t^2}{2}}$	$\alpha_i t$
Weibull	$\frac{k_i}{\lambda_i}\left(\frac{t}{\lambda_i}\right)^{k_i-1} e^{-\left(\frac{t}{\lambda_i}\right)^{k_i}}$	$e^{-\left(\frac{t}{\lambda_i}\right)^{k_i}}$	$\frac{k_i}{\lambda_i}\left(\frac{t}{\lambda_i}\right)^{k_i-1}$

Characteristics of behavioral dynamics can be well captured by Weibull distribution.

From Behavioral Dynamics to Cascades



(a) Cascading Process

(b) Partially observed cascade at t

(c) Behavioral dynamics of p_1

Macro ← – – – – – – – – – – – – **Micro**

NEtworked WEibull Regression (NEWER)

$$F(\lambda, k, \beta, \gamma) = G_1(\lambda, k) + \mu G_2(\beta, \lambda) + \eta G_3(\gamma, k)$$

$$G_1(\lambda, k) = -\log L(\lambda, k)$$

$$G_2(\lambda, \beta) = \frac{1}{2N} \left\|\log \lambda - \log X \cdot \beta\right\|^2 + \alpha_\beta \left\|\beta\right\|_1$$

$$G_3(k, \gamma) = \frac{1}{2N} \left\|\log k - \log X \cdot \gamma\right\|^2 + \alpha_\gamma \left\|\gamma\right\|_1$$

- Theoretically proved to be lower-bounded.
- Coordinate Descent strategy is exploited with guaranteed convergence.

Algorithm 1 Basic Model

Input:
Set of users U involved in the cascade C before time t_{limit}, survival functions of users $S_{u_j}(t)$, predicting time t_e;

Output:
Size of cascade $size(C_{t_e})$;

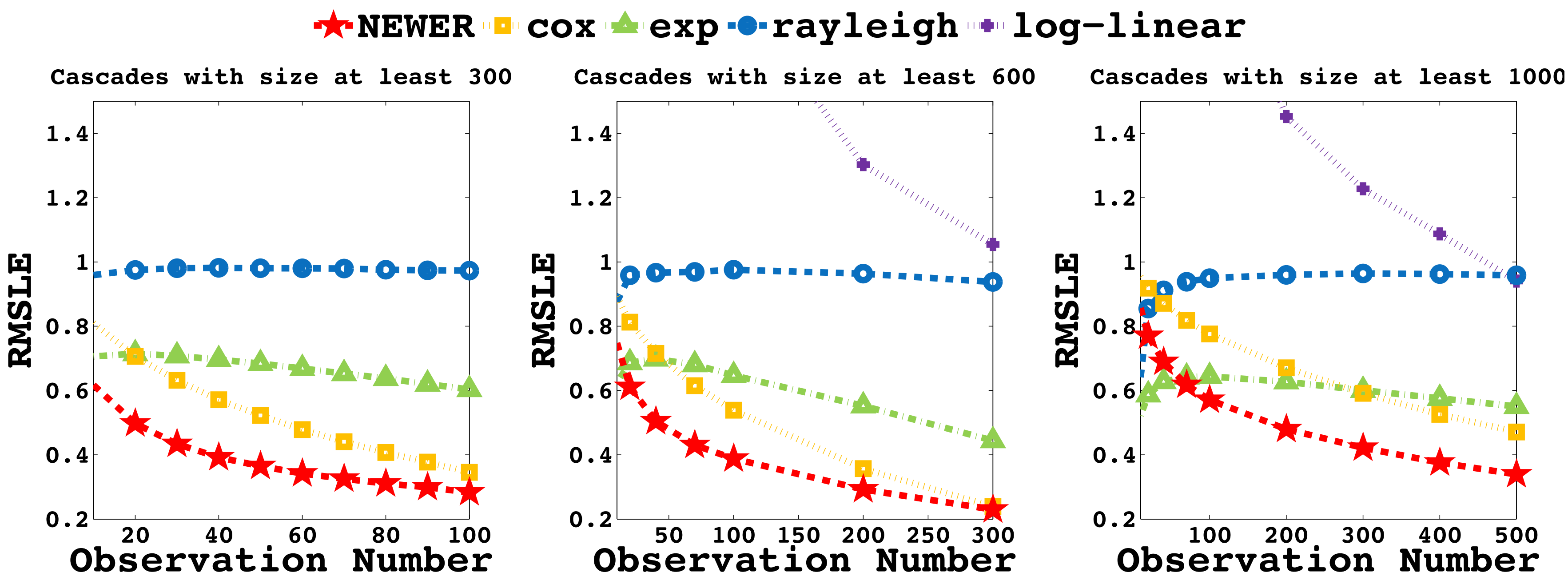
```
1: for all user u_i ∈ U do
2:   creates a subcascade process with replynum(u_i) = 0
3:   if u_i is not root node then
4:     replynum(rp(u_i)) = replynum(rp(u_i)) + 1
5:   end if
6: end for
7: sum = 1
8: for all user u_i ∈ U do
9:   deathrate(u_i) = max(1 − S_{u_i}(t_limit − t(u_i)), 1/|V|)
10:  fdrate(u_i) = max(1 − S_{u_i}(t_e − t(u_i)), 1/|V|)
11:  sum = sum + replynum(u_i)·fdrate(u_i) / deathrate(u_i)
12: end for
13: return size(C_{t_e}) = sum
```

Experiments

- Datasets: Tencent Weibo
  - All cascades generated between Nov 15th and Nov 25th in 2011.
  - retain all 0.59 million cascades that the cascades size are at least 5.
- Baseline:
  - Cox Proportional Hazard Regression Model (Cox)
  - Exponential/Rayleigh Proportional Hazard Regression Model (Exponential/Rayleigh)
  - log-Linear regression(Log-linear)
- Evaluation metric:
  - RMSLE: Root Mean Square Log Error
  - $\Delta\sigma$-Precision: Precision value that the predicted value within $(1+\sigma)\pm 1$groundtruth

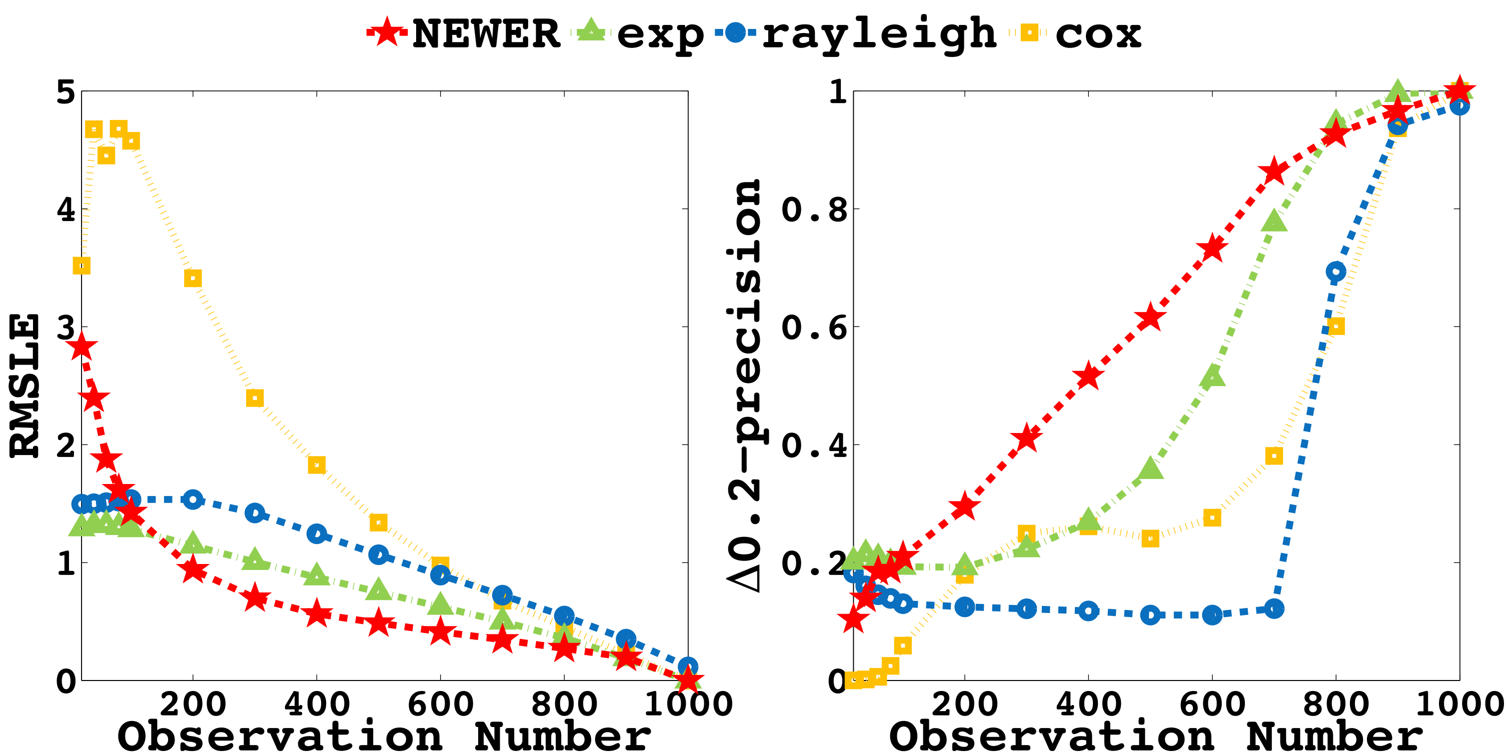
Cascade Size Prediction

- What is the final size of the cascade?



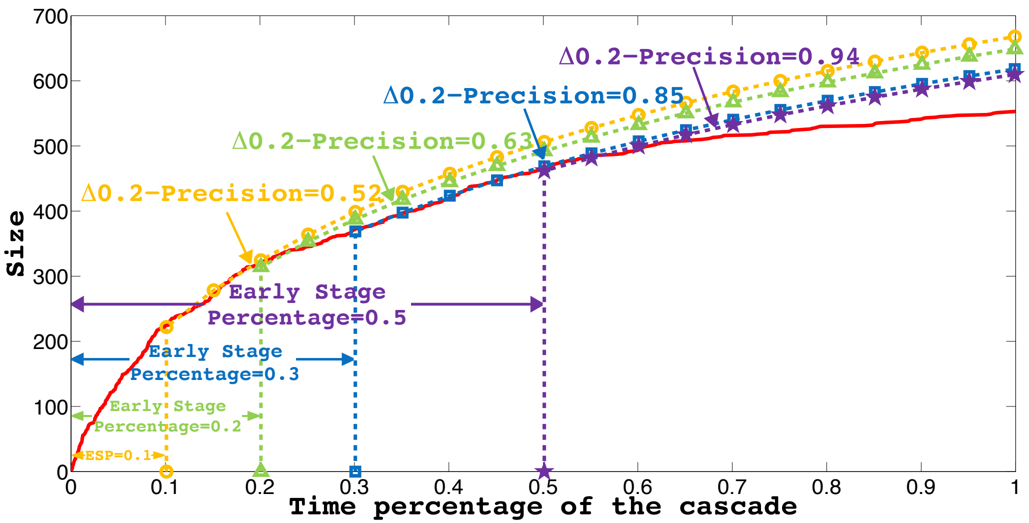
Outbreak Time Prediction

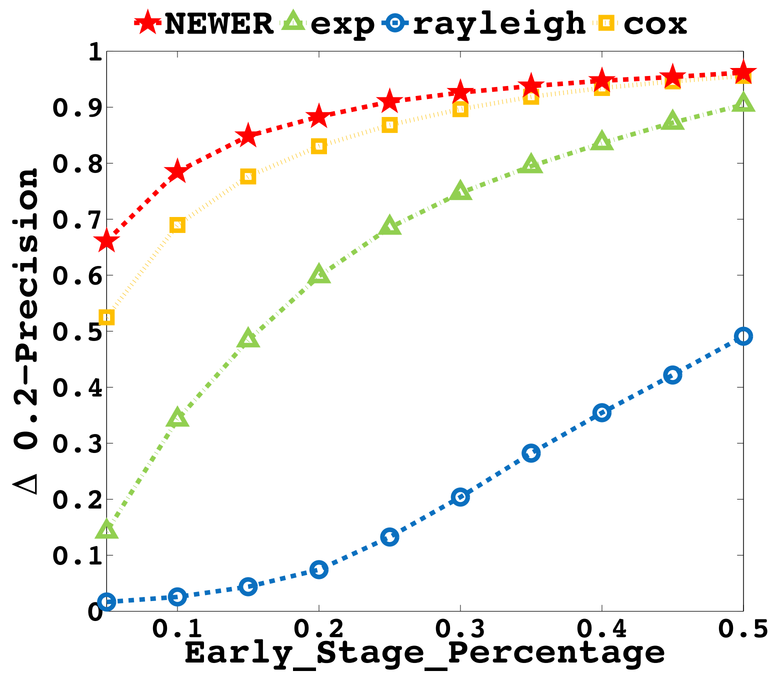
- When will the cascade break out?



Cascading Process Prediction

- What is the size of the cascade at any later point?





Conclusions

- Before predicting information spreading, understanding the ***behavioral mechanism*** is critical and fundamental.
- Behaviors can be modeled in different ***granularities***, which depends on the target problem.
- Modeling information spreading with ***continuous-time model*** is promising and demonstrated to be effective in our research.