Machine learning for Dynamic Social Network Analysis

Applications: Control

Manuel Gomez Rodriguez
Max Planck Institute for Software Systems

IJCAI Tutorial, August 2017

Outline of the Seminar

REPRESENTATION: TEMPORAL POINT PROCESSES

1. Intensity function
2. Basic building blocks
3. Superposition
4. Marks and SDEs with jumps

APPLICATIONS: MODELS

1. Information propagation
2. Information reliability
3. Knowledge acquisition

APPLICATIONS: CONTROL

1. Activity shaping
2. When-to-post

Next

Slides/references: learning.mpi-sws.org/ijcai-2017-tutorial

Applications: Control

1. Activity shaping
2. When-to-post

Activity shaping

Can we **steer users' activity** in a **social network in general**?

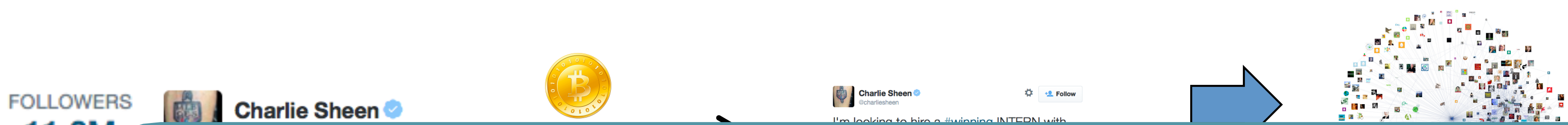
Why this goal?

TIME **Twitter Stock Tumbles After Drop in User Engagement**

HUFF POST **7 Ways to Increase Your Social Media Engagement**

Activity shaping vs influence maximization

Related to
Influence Maximization Problem

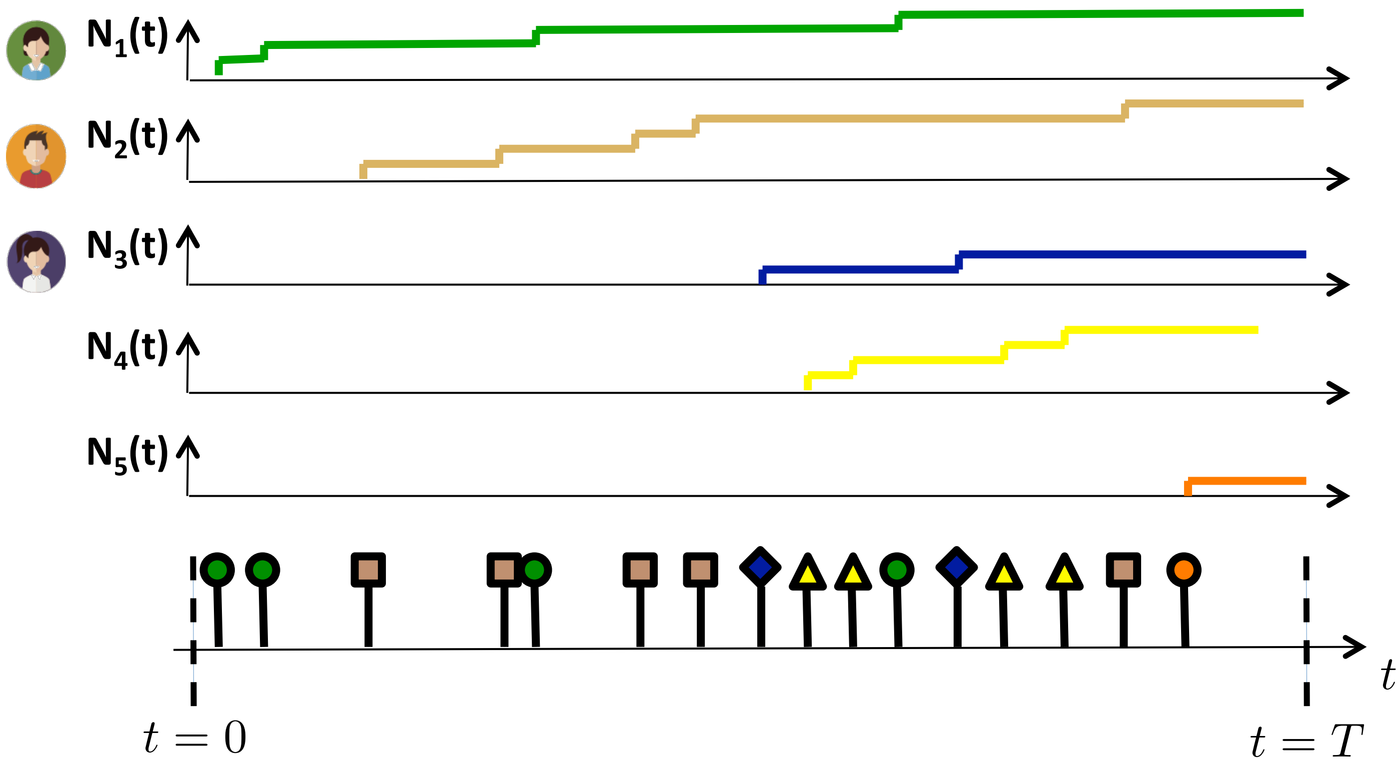


Activity shaping is a generalization of **influence maximization**

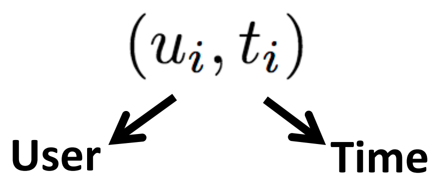
Influence Maximization	**incentive**	**same piece of information**	**maximizing adoption**
Activity Shaping	**Variable incentive**	**Multiple times multiple pieces, recurrent!**	**Many different activity shaping tasks**

Event representation

We represent messages using **nonterminating temporal point processes**:

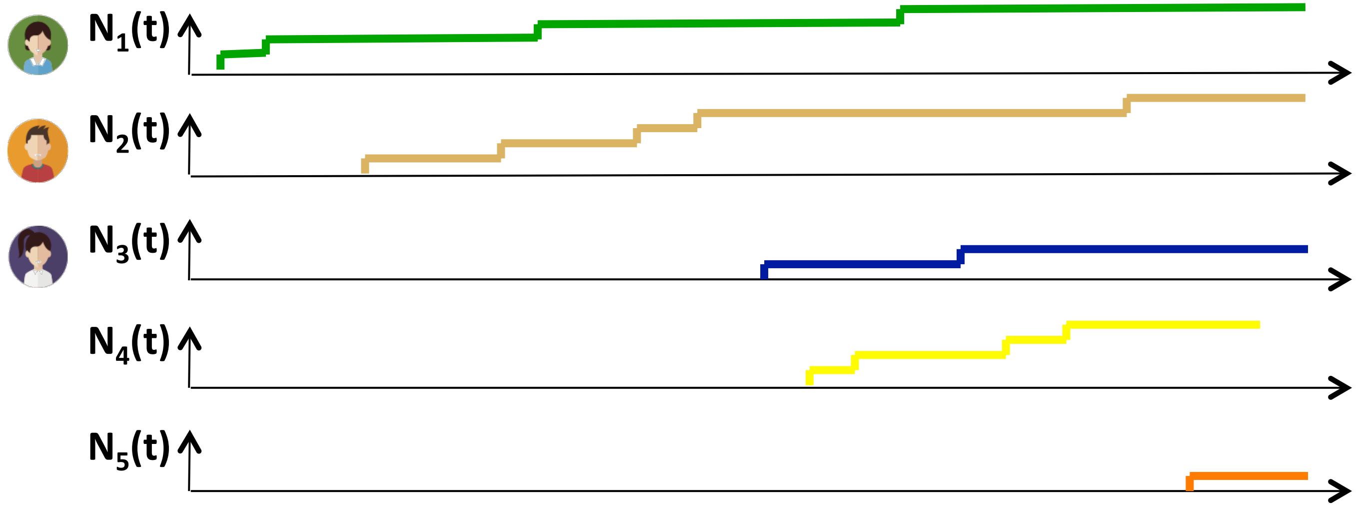


Recurrent event:

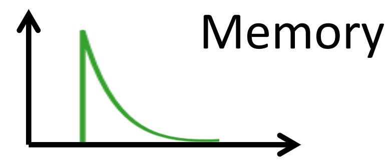


[Farajtabar et al., NIPS 2014]

Events intensity



Exogenous activity



$$\underbrace{\lambda_u^*(t)}_{\text{User's intensity}} = \underbrace{\boldsymbol{\lambda}^{(0)}(t)}_{\text{Messages on her own initiative}} + \sum_{i:t_i<t} a_{uu_i}\, g(t-t_i)$$

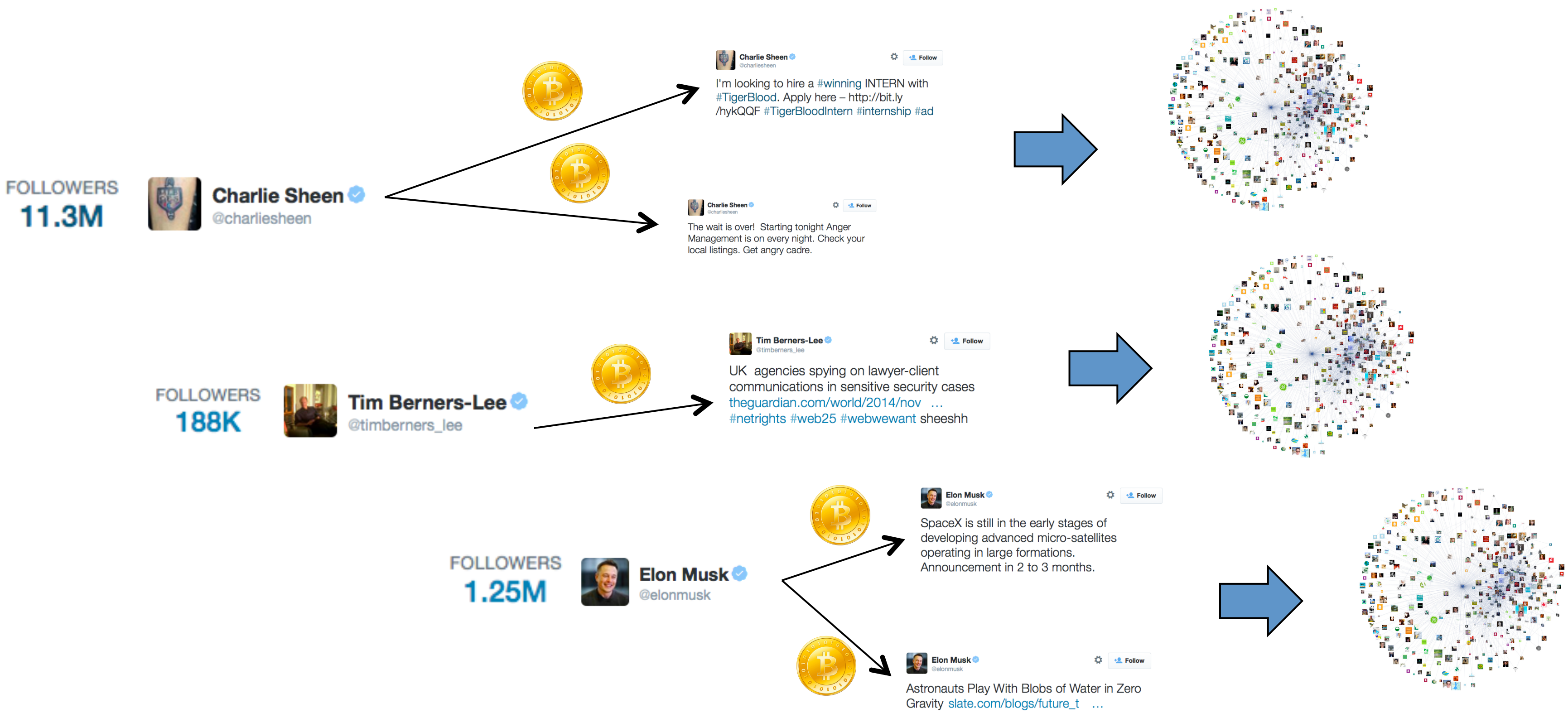
Influence from user u_i on user u

Hawkes process

[Farajtabar et al., NIPS 2014]

Activity shaping… how?

Incentivize
a few users to produce a given level
of overall users' activity



Exogenous activity

Endogenous activity

Activity shaping… what is it?

Activity Shaping:

Find exogenous activity $\boldsymbol{\lambda}^{(0)}(t)$ that results in a desired average overall activity at a given time:

$$\boldsymbol{\mu}(t) := \mathbb{E}_{\mathcal{H}_{t-}}\left[\boldsymbol{\lambda}(t)\right]$$

Average with respect to the history of events up to t!

[Farajtabar et al., NIPS 2014]

Exogenous intensity & average overall intensity

$$\boldsymbol{\lambda}^{(0)}(t) \quad \Rightarrow \quad \boldsymbol{\mu}(t) := \mathbb{E}_{\mathcal{H}_{t-}}\left[\boldsymbol{\lambda}(t)\right]$$

How do they relate?

Surprisingly…

linearly:

Convolution

$$\boldsymbol{\mu}(t) = \boldsymbol{\Psi}(t) \star \boldsymbol{\lambda}^{(0)}(t)$$

matrix that depends on

$g(t)$ and $\left(a_{uu'}\right)$

non negative kernel **influence matrix**

[Farajtabar et al., NIPS 2014]

Exact Relation

If the memory g(t) is exponential:

$$\boldsymbol{\Psi}(t) := \left(e^{(\boldsymbol{A}-\omega\boldsymbol{I})t} + \omega(\boldsymbol{A}-\omega\boldsymbol{I})^{-1}(e^{(\boldsymbol{A}-\omega\boldsymbol{I})t} - \boldsymbol{I})\right)$$

Matrix exponentials

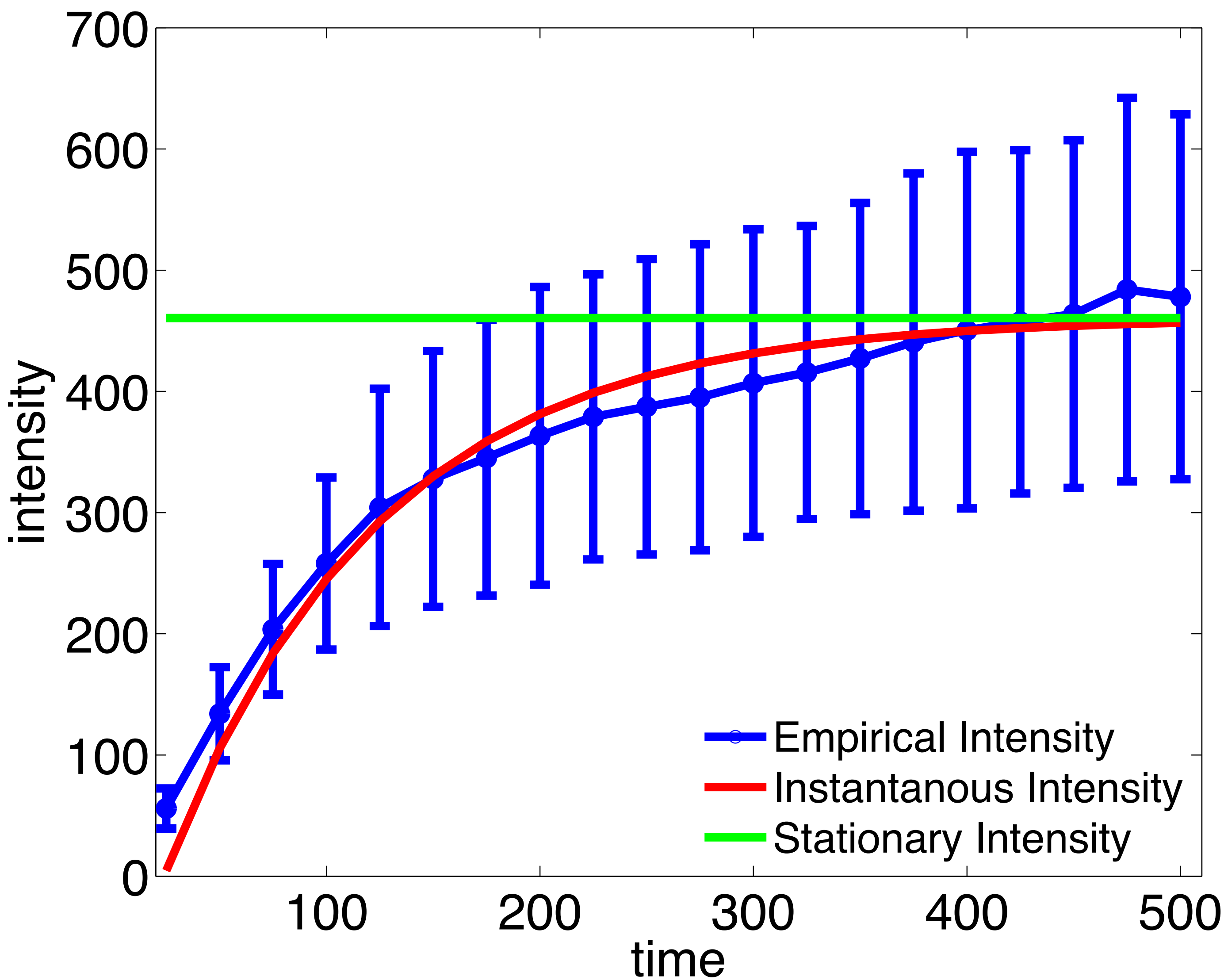
$$\boldsymbol{\mu}(t) = \boldsymbol{\Psi}(t) \star \boldsymbol{\lambda}^{(0)}(t)$$

$$\boldsymbol{\mu}(t) = \boldsymbol{\Psi}(t)\boldsymbol{\lambda}^{(0)}$$

Corollary
exogenous intensity
is **constant**

[Farajtabar et al., NIPS 2014]

Does it really work in practice?



[Farajtabar et al., NIPS 2014]

Activity shaping optimization framework

Once we know that $\boldsymbol{\mu}(t) = \boldsymbol{\Psi}(t)\boldsymbol{\lambda}^{(0)}$

we can find $\boldsymbol{\lambda}^{(0)}$ to satisfy **many different goals**:

ACTIVITY S

We can solve this problem **efficiently** for **a large family of utilities!**

maximize$_{\boldsymbol{\mu}(t)}$
subject to $\boldsymbol{\lambda}^{(0)} \geqslant 0$

Cost for incentivizing

[Farajtabar et al., NIPS 2014]

Capped activity maximization (CAM)

If our goal is **maximizing the overall number of events** across a social network:

Max feasible activity per user

$$
\begin{array}{ll}
\text{maximize}_{\boldsymbol{\mu}(t),\boldsymbol{\lambda}^{(0)}} & \sum_{u\in[m]} \min\left\{\mu_u(t), \alpha_u\right\} \\
\text{subject to} & \boldsymbol{\mu}(t) = \boldsymbol{\Psi}(t)\boldsymbol{\lambda}^{(0)}, \quad \boldsymbol{c}^{\top}\boldsymbol{\lambda}^{(0)} \leqslant C, \quad \boldsymbol{\lambda}^{(0)} \geqslant 0
\end{array}
$$

HUFF POST

7 Ways to Increase Your Social Media Engagement

[Farajtabar et al., NIPS 2014]

Minimax activity shaping (MMASH)

If our goal is **make the user with the minimum activity as active as possible**:

$$\text{maximize}_{\boldsymbol{\mu}(t), \boldsymbol{\lambda}^{(0)}} \;\; \min_u \;\; \mu_u(t)$$

$$\text{subject to} \qquad \boldsymbol{\mu}(t) = \boldsymbol{\Psi}(t)\boldsymbol{\lambda}^{(0)}, \quad \boldsymbol{c}^\top \boldsymbol{\lambda}^{(0)} \leqslant C, \quad \boldsymbol{\lambda}^{(0)} \geqslant 0$$

BUSINESS INSIDER

1% Of Players Provide The Biggest Chunk Of Zynga's Revenue

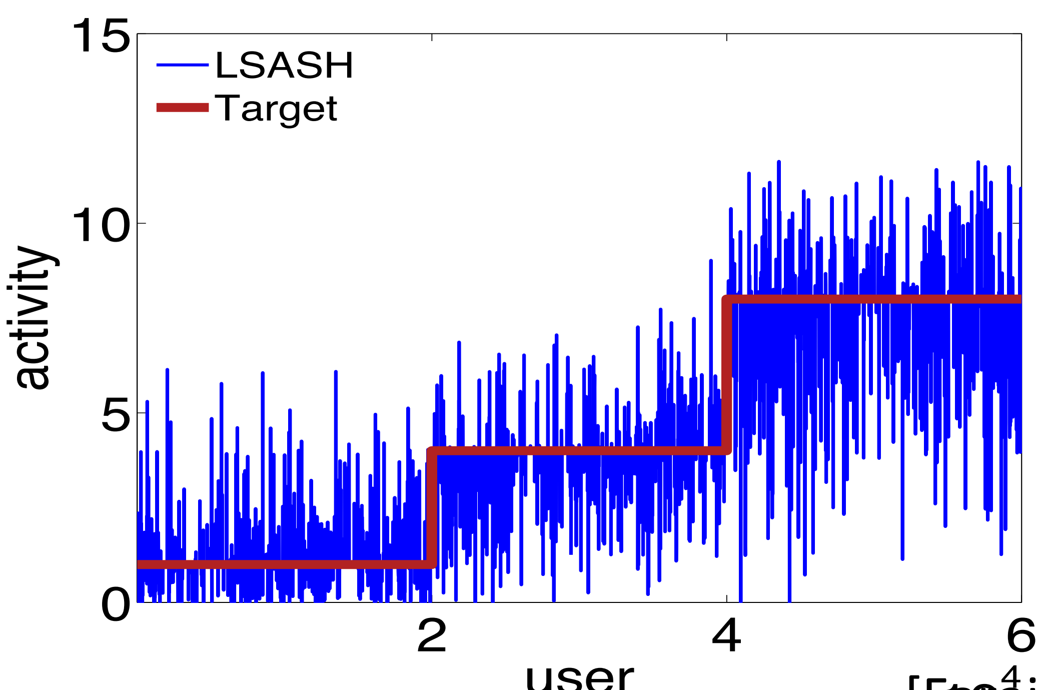
[Farajtabar et al., NIPS 2014]

Least-squares activity shaping (LSASH)

If our goal is **to achieve a pre-specified level of activity for each user or group of users**:

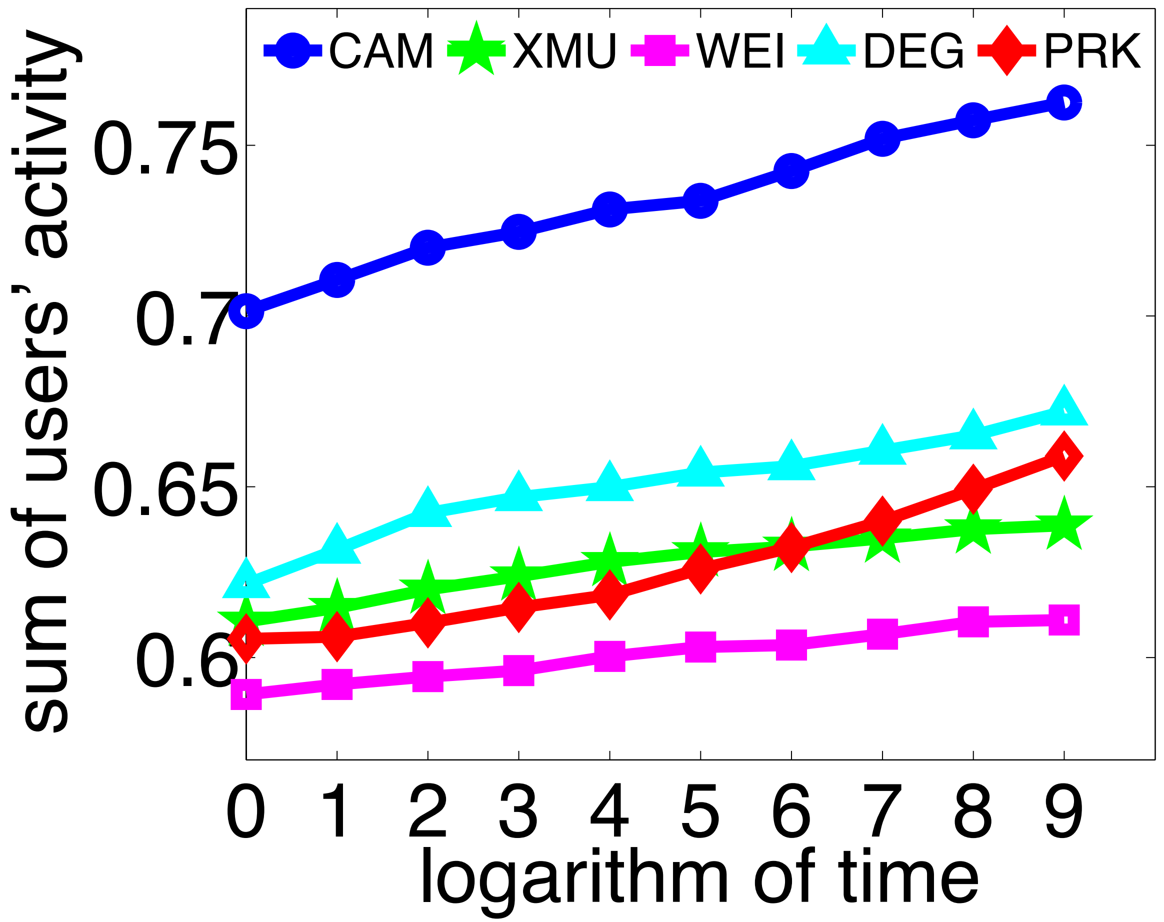
$$\text{maximize}_{\boldsymbol{\mu}(t),\boldsymbol{\lambda}^{(0)}} \quad -\|\boldsymbol{B}\boldsymbol{\mu}(t)-\boldsymbol{v}\|_2^2$$

$$\text{subject to} \qquad \boldsymbol{\mu}(t)=\boldsymbol{\Psi}(t)\boldsymbol{\lambda}^{(0)}, \quad \boldsymbol{c}^\top\boldsymbol{\lambda}^{(0)} \leqslant C, \quad \boldsymbol{\lambda}^{(0)} \geqslant 0$$



[Farajtabar et al., NIPS 2014]

Capped activity maximization: results



+10% more events than best heuristic

+34,000 more events per month than best heuristic for 2,000 Twitter users

[Farajtabar et al., NIPS 2014]

Applications: Control

1. Activity shaping
2. When-to-post

Social media as a broadcasting platform

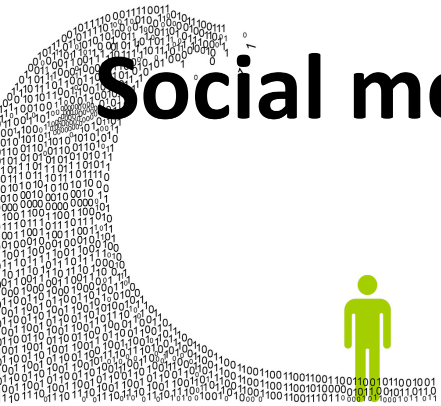
Everybody can build, reach and broadcast information to their own audience

Broadcasted content

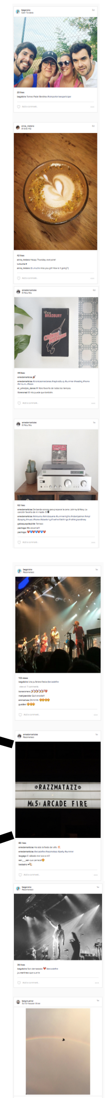
Audience reaction

Attention is scarce

Social media users follow many broadcasters

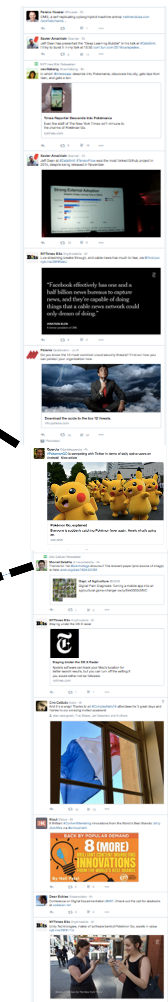


Instagram feed



Older posts

Twitter feed



Older posts

What are the best times to post?

THE HUFFINGTON POST

THE BLOG

The Best Times to Post on Social Media

Tech.Mic

Here Are the Be

So Your Picture

Can we design an algorithm that tell us when to post to achieve **high visibility?**

COMPANY

To Post On

HubSpot

The Best Tim

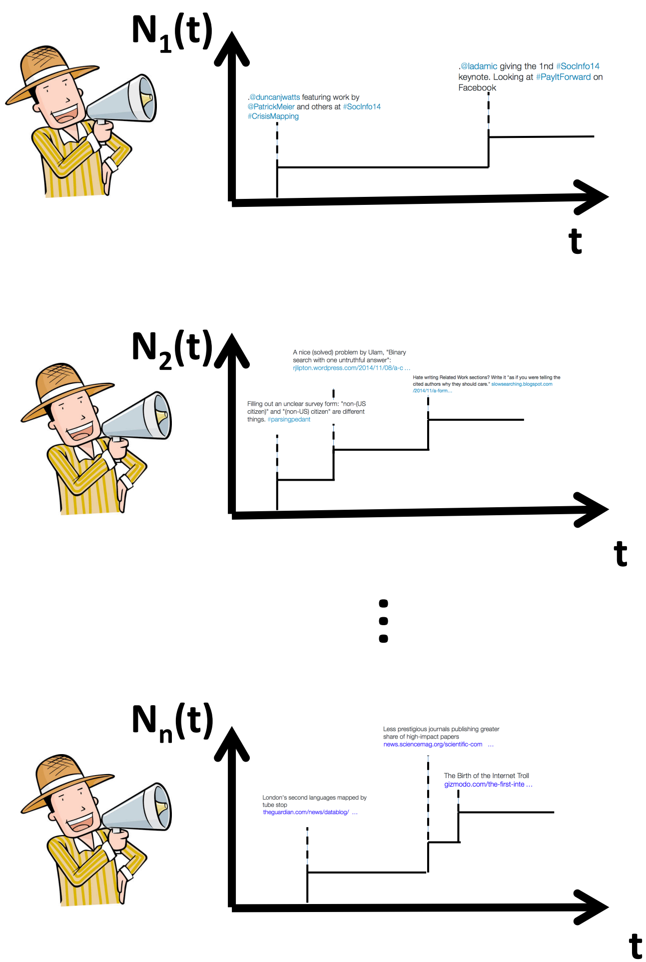
Twitter, LinkedIn & Other Social Media Sites [Infographic]

Forbes

For Brands And PR: When Is The Best Time To Post On Social Media?

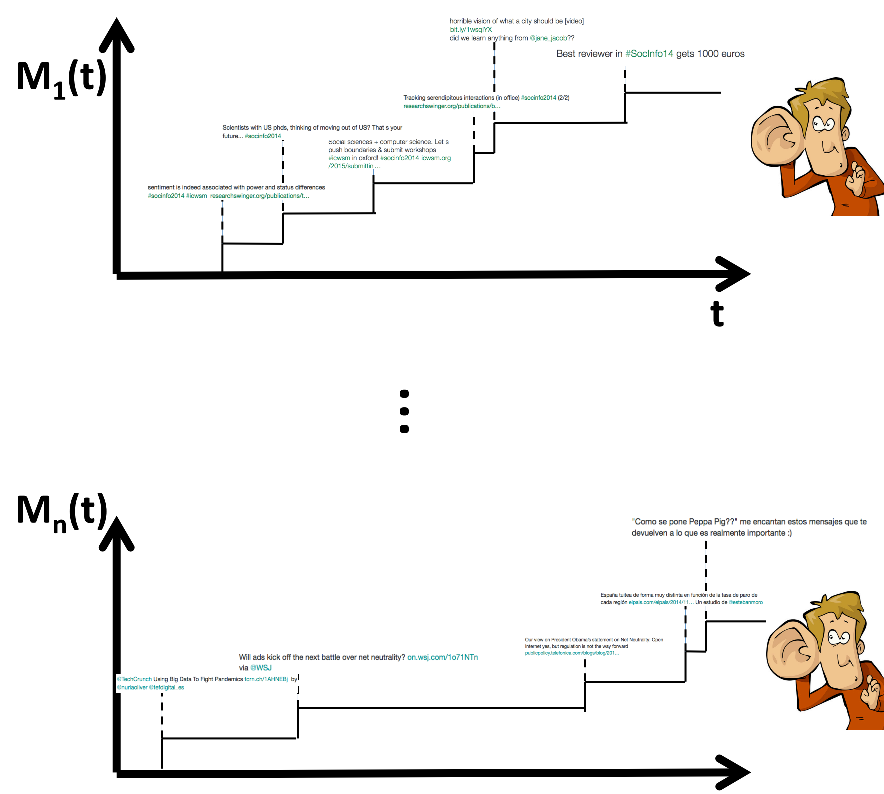
Representation of broadcasters and feeds

Broadcasters' posts as a counting process N(t)

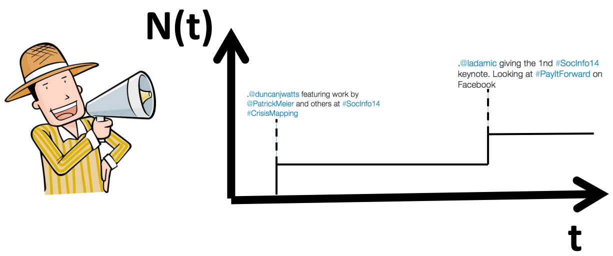


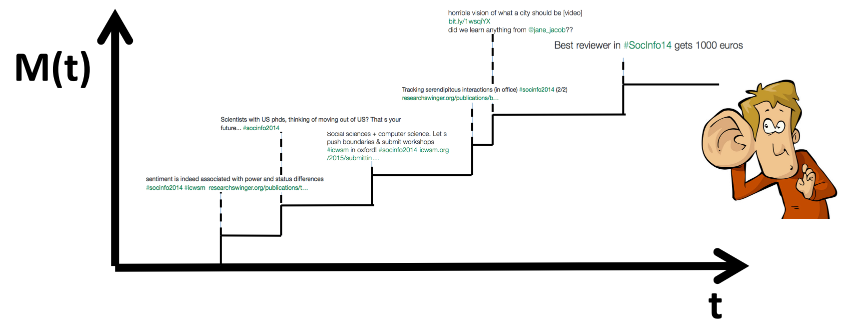
$$M(t) = A^T N(t)$$

Users' feeds as sum of counting processes M(t)



Broadcasting and feeds intensities





$$\mathbb{E}[\underbrace{d\boldsymbol{N}(t)}_{\in\{0,1\}} | \mathcal{H}(t)] = \underbrace{\boldsymbol{\mu}(t)}_{\text{Broadcaster intensity function (tweets / hour)}} dt$$

$$\mathbb{E}[\underbrace{d\boldsymbol{M}(t)}_{\in\{0,1\}} | \mathcal{H}(t)] = \underbrace{\boldsymbol{\gamma}(t)}_{A^T \boldsymbol{\mu}(t)} dt$$

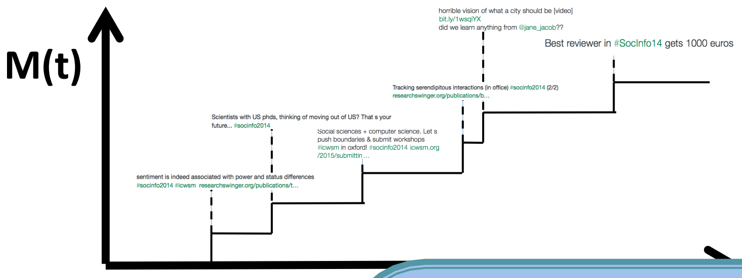
Feed intensity function (tweets / hour)

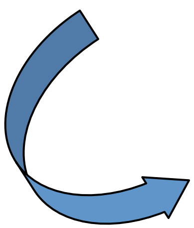
Given a broadcaster i and her followers

$$\boldsymbol{M}_{\setminus i}(t) = A^T \boldsymbol{N}(t) - A_i N_i(t)$$
$$\gamma_{j \setminus i}(t) = \gamma_j(t) - \mu_i(t)$$

Feed due to other broadcasters

Definition of visibility function





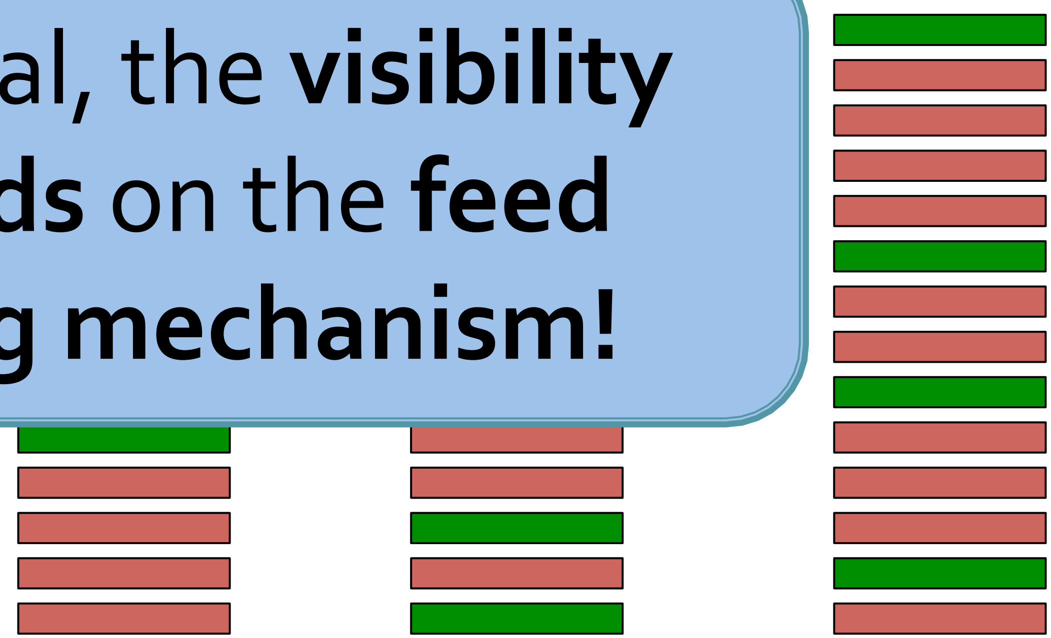
Feed ranking

Ranked stories

Visibility of broadcaster i at follower j

Position of the highest ranked tweet by broadcaster i in follower j's wall

$r_{ij}(t) = 0$ $r_{ij}(t') = 4$ $r_{ij}(t'') = 0$



In general, the **visibility depends** on the **feed ranking mechanism!**

Post by broadcaster u

Post by other broadcasters

Optimal control of temporal point processes

Formulate the when-to-post problem as a novel stochastic optimal control problem

(of independent interest)

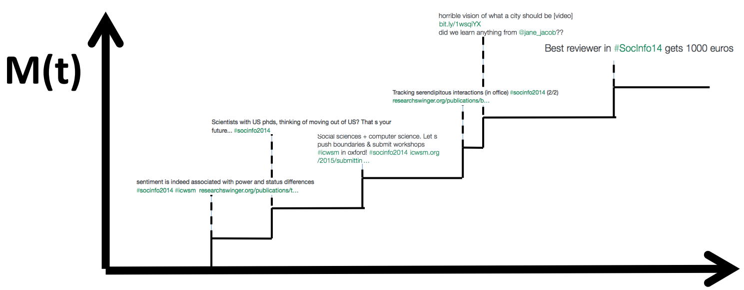
Visibility and feed dynamics → **Optimizing visibility** → **Experiments**

System of stochastic equations with jumps

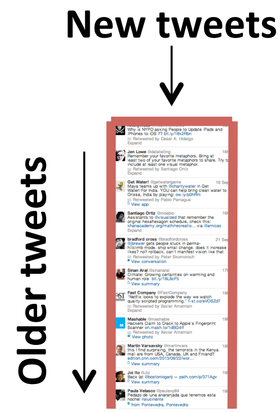
Optimal control of jumps

Twitter

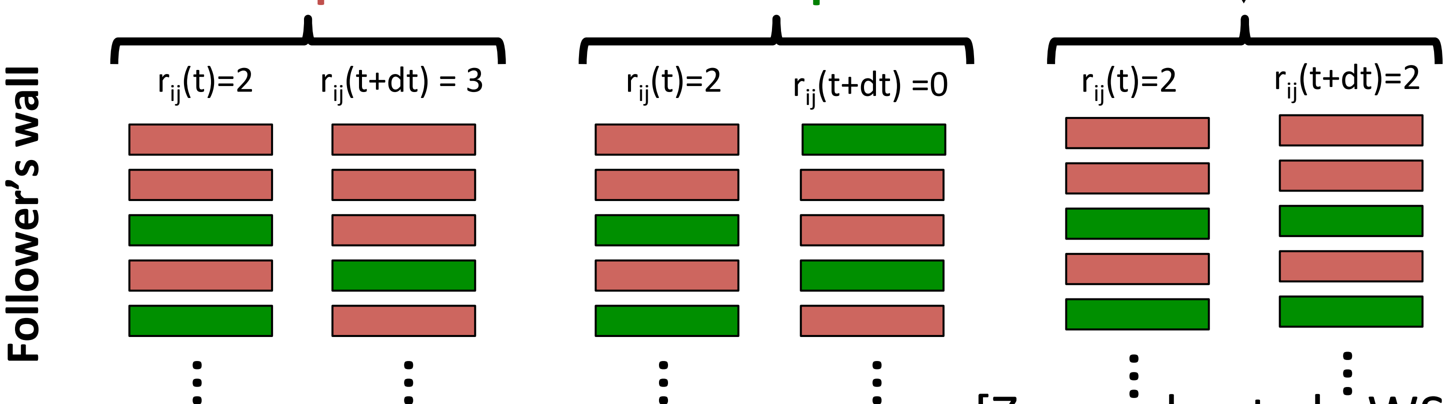
Visibility dynamics in a FIFO feed (I)



Reverse chronological order



$$\underbrace{r_{ij}(t+dt)}_{\text{Rank at t+dt}} = \underbrace{(r_{ij}(t)+1)dM_{j\backslash i}(t)(1-dN_i(t))}_{\text{Other broadcasters post a story and broadcaster i does not post}} + \underbrace{0}_{\text{Broadcaster i posts a story and other broadcasters do not post}} + \underbrace{r_{ij}(t)(1-dM_{j\backslash i}(t))(1-dN_i(t))}_{\text{Nobody posts a story}}$$



[Zarezade et al., WSDM 2017]

Visibility dynamics in a FIFO feed (II)

$$r_{ij}(t+dt) = \quad (r_{ij}(t)+1)dM_{j\backslash i}(t)(1-dN_i(t)) + \; 0 + \; r_{ij}(t)(1-dM_{j\backslash i}(t))(1-dN_i(t))$$

Zero-one law $dN_i(t)dM_{j\backslash i}(t) = 0$

$$dr_{ij}(t) = -r_{ij}(t)\,dN_i(t) + dM_{j\backslash i}(t)$$

$r_{ij}(t+dt) - r_{ij}(t)$

Broadcaster i posts a story

Other broadcasters posts a story

Stochastic differential equation (SDE) with jumps

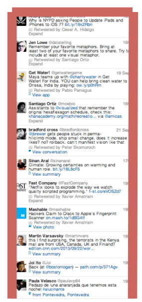
OUR GOAL:

Optimize $r_{ij}(t)$ over time, so that it is small, by controlling $dN_i(t)$ through the intensity $\mu_i(t)$

[Zarezade et al., WSDM 2017]

Feed dynamics

$N_i(t) \quad M_{j\setminus i}(t) = M(t) \rightarrow \gamma_{j\setminus i}(t) = \lambda(t)$



We consider **a general intensity:**

(e.g. Hawkes, inhomogeneous Poisson)

$$\lambda^*(t) = \underbrace{\lambda_0(t)}_{\substack{\textbf{Deterministic} \\ \textbf{arbitrary intensity}}} + \underbrace{\alpha \int_0^t g(t-s)\,dN(s)}_{\substack{\textbf{Stochastic} \\ \textbf{self-excitation}}}$$

Jump stochastic differential equation (SDE)

$$\left\{ d\lambda^*(t) = \left[\lambda_0'(t) + w\lambda_0(t) - w\lambda^*(t)\right] dt + \alpha dN_i(t) \right.$$

[Zarezade et al., WSDM 2017]

The when-to-post problem

$$\mu_i(t) = u(t) \rightarrow N_i(t)$$

$$N_i(t) \quad M_{j\setminus i}(t)$$

$$dr_{ij}(t) = -r_{ij}(t)\, dN_i(t) + dM_{j\setminus i}(t)$$

$$N_i(t) \quad M_{j\setminus i}(t)$$

$$dr_{ij}(t) = -r_{ij}(t)\, dN_i(t) + dM_{j\setminus i}(t)$$

...

$$N_i(t) \quad M_{j\setminus i}(t)$$

$$dr_{ij}(t) = -r_{ij}(t)\, dN_i(t) + dM_{j\setminus i}(t)$$

Optimization problem

$$\begin{aligned} \underset{u(t_0,t_f]}{\text{minimize}} \quad & \mathbb{E}_{(N_i, \boldsymbol{M}_{\setminus i})(t_0,t_f]}\left[\overbrace{\phi(\boldsymbol{r}(t_f))}^{\textbf{Terminal penalty}} + \int_{t_0}^{t_f} \overbrace{\ell(\boldsymbol{r}(\tau), u(\tau))}^{\textbf{Nondecreasing loss}} d\tau\right] \\ \text{subject to} \quad & u(t) \geq 0 \quad \forall t \in (t_0, t_f], \end{aligned}$$

Dynamics defined by Jump SDEs

$$dr(t) = -r(t)\, dN(t) + dM(t)$$

$$d\lambda(t) = [\lambda_0'(t) + w\lambda_0(t) - w\lambda(t)]\, dt + \alpha\, dM(t)$$

[Zarezade et al., WSDM 2017]

Bellman's Principle of Optimality

Lemma. The **optimal cost-to-go** satisfies **Bellman's Principle of Optimality**

$$J(r(t), \lambda(t), t) = \min_{u(t,t+dt]} \mathbb{E}\left[J(r(t+dt), \lambda(t+dt), t+dt)\right] + \ell(r(t), u(t))\, dt$$

$$J(r(t+dt), \lambda(t+dt), t+dt) = J(r(t), \lambda(t), t) + dJ(r(t), \lambda(t), t)$$

$$0 = \min_{u(t,t+dt]} \mathbb{E}\left[dJ(r(t), \lambda(t), t)\right] + \ell(r(t), u(t))\, dt$$

$$dr(t) = -r(t)\, dN(t) + dM(t)$$
$$d\lambda(t) = [\lambda_0'(t) + w\lambda_0(t) - w\lambda(t)]\, dt + \alpha\, dM(t)$$

Hamilton-Jacobi-Bellman (HJB) equation

Partial differential equation in J (with respect to r, λ and t)

[Zarezade et al., WSDM 2017]

Solving the HJB equation

Consider a quadratic loss

$$\ell(r(t), u(t)) = \frac{1}{2} s(t)\, r^2(t) + \frac{1}{2} q\, u^2(t)$$

Favors some periods of times (e.g., times in which the follower is online)

Trade-offs visibility and number of broadcasted posts

We propose $J(r(t), \lambda(t), t)$ and then show that the optimal intensity is:

$$u^*(t) = q^{-1} \left[J(r(t), \lambda(t), t) - J(0, \lambda(t), t) \right]$$

$$= \sqrt{s(t)/q}\, r(t)$$

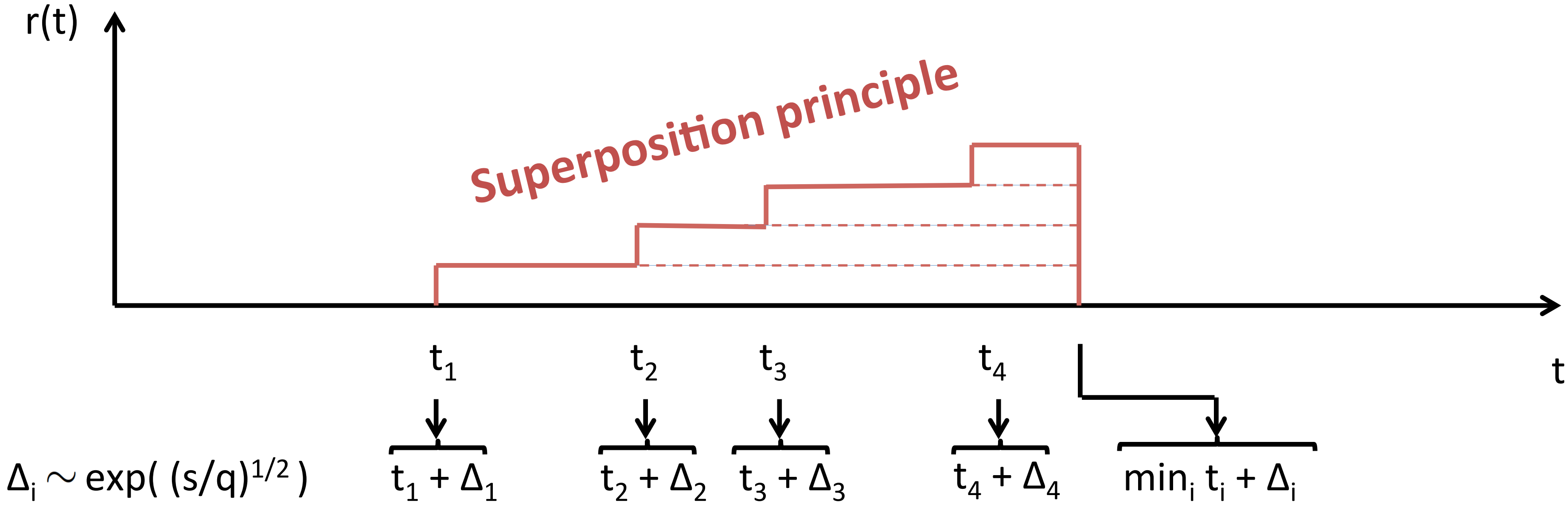
It only depends on the current visibility!

[Zarezade et al., WSDM 2017]

The RedQueen algorithm

Consider $s(t) = s \longrightarrow u^*(t) = (s/q)^{1/2}\, r(t)$

How do we sample the next time?



It only requires sampling $M(t_f)$ times!

[Zarezade et al., WSDM 2017]

The RedQueen algorithm

RedQueen can be implemented in a few lines of code!

Algorithm 1: RedQueen for fixed s, q and one follower.

Input: Parameters q and s
Output: Returns time for the next post
$t \leftarrow \infty$; $\tau \leftarrow othersNextPost()$
while $\tau < t$ **do**
    $\Delta \sim \exp(\sqrt{s/q})$
    $t \leftarrow \min(t, \tau + \Delta)$
    $\tau \leftarrow othersNextPost()$
end
return t

[Zarezade et al., WSDM 2017]

When-to-post for multiple followers

Consider n followers and a quadratic loss:

$$\ell(\boldsymbol{r}(t), u(t), t) = \sum_{i=1}^{n} \frac{1}{2} s_i(t) r_i^2(t) + \frac{1}{2} q\, u^2(t)$$

umber

(e.g., times in

Then, w

$u^*(t)$

$i=1$

It only depends on the current visibilities!

We can easily adapt the efficient sampling algorithm to multiple followers!

[Zarezade et al., WSDM 2017]

Novelty in the problem formulation

The problem formulation is unique in two key technical aspects:

I. The control signal is a conditional intensity

Previous work: time-varying real vector

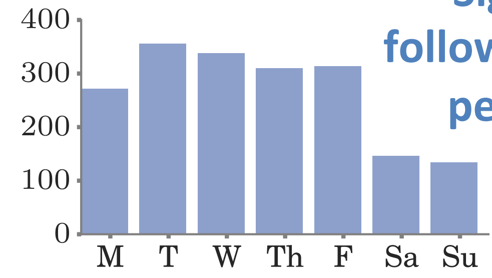
II. The jumps are doubly stochastic

Previous work: memory-less jumps

[Zarezade et al., WSDM 2017]

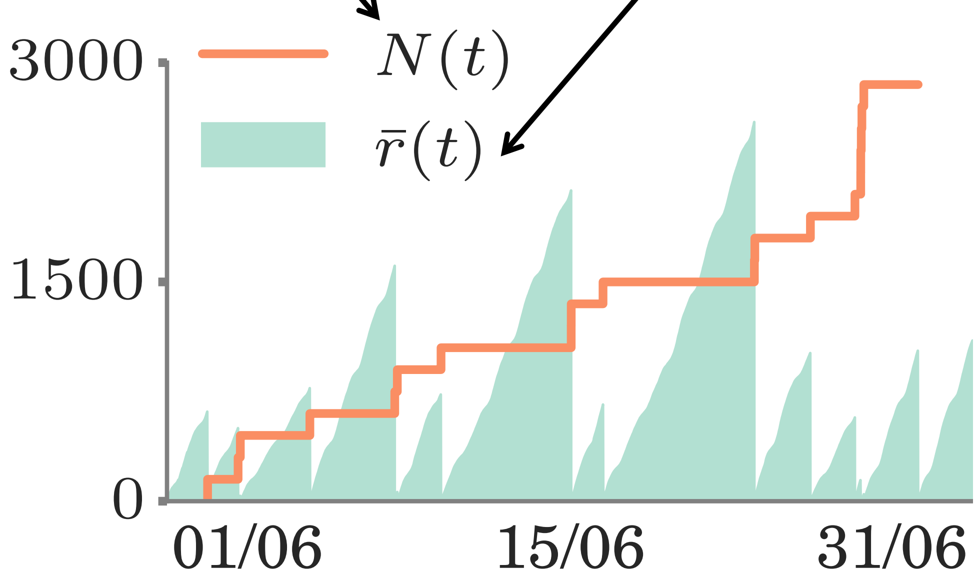
Case study: one broadcaster

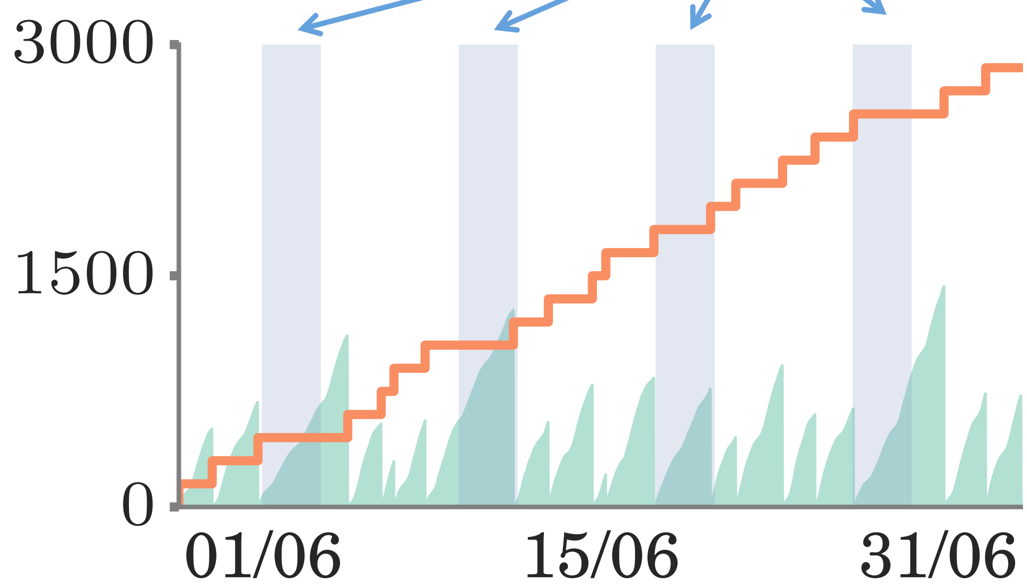
Significance: followers' retweets per weekday



Broadcaster's posts

Average position over time





True posts

$$\frac{1}{T}\int_0^T \bar{r}(t)dt = 698.04$$

40% lower!

RedQueen

$$\frac{1}{T}\int_0^T \bar{r}(t)dt = 425.25$$

[Zarezade et al., WSDM 2017]

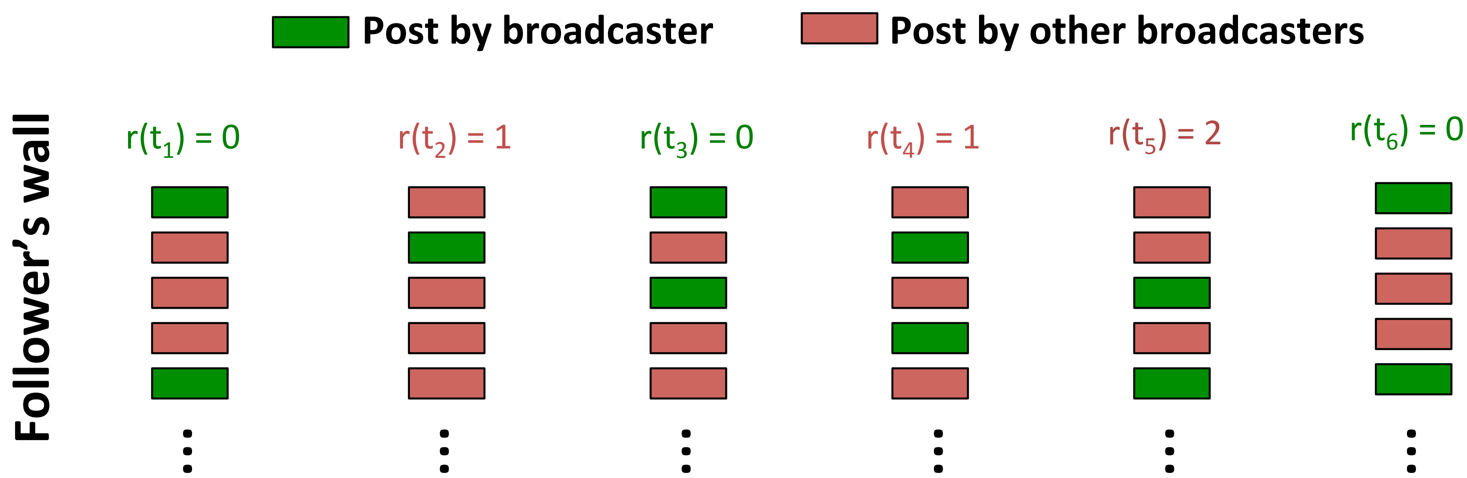
Evaluation metrics

Position over time

$$\int_0^T r(t)dt$$

Time at the top

$$\int_0^T \mathbb{I}(r(t)<1)dt$$

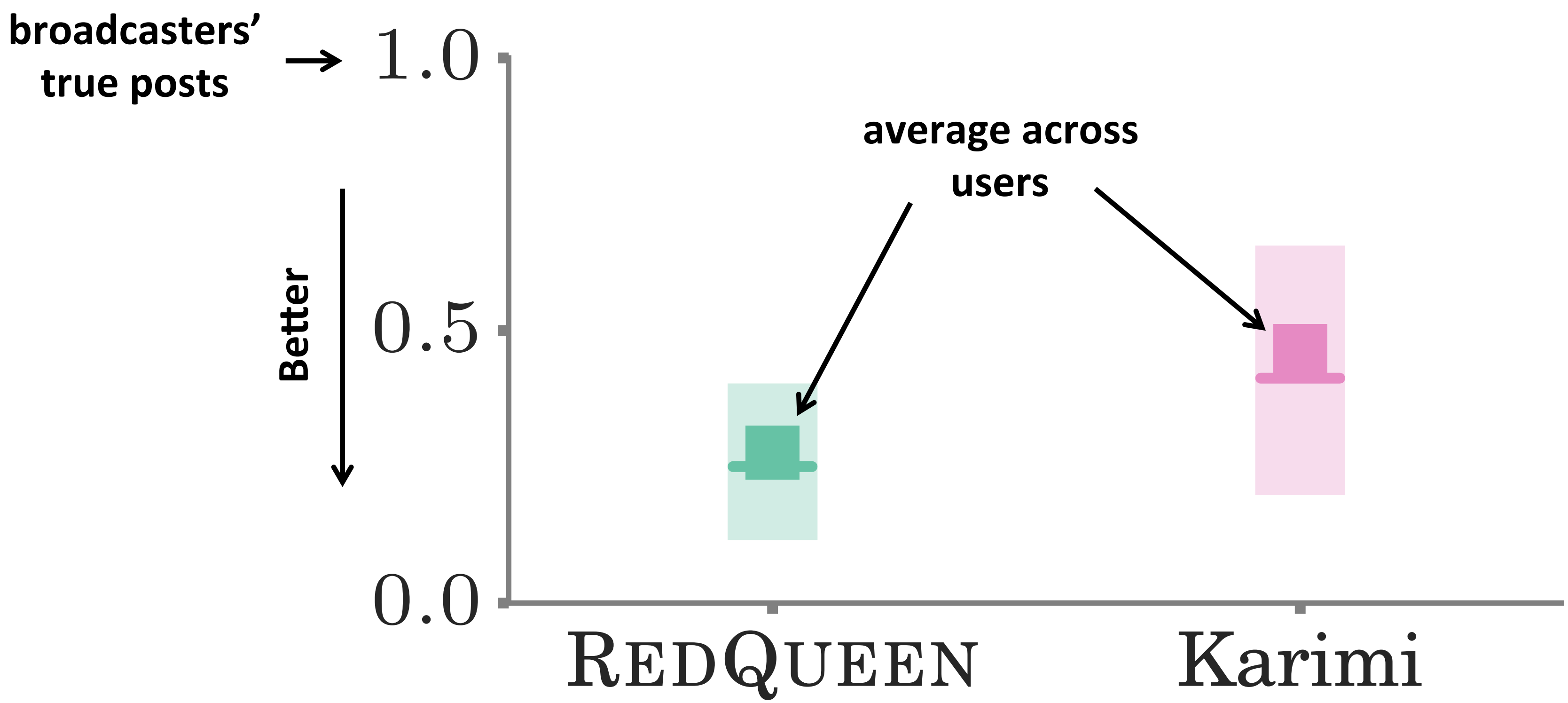


Position over time = $0 \times (t_2 - t_1) + 1 \times (t_3 - t_2) + 0 \times (t_4 - t_3) + 1 \times (t_5 - t_4) + 2 \times (t_6 - t_5)$

Time at the top = $(t_2 - t_1) + 0 + (t_4 - t_3) + 0 + 0$

[Zarezade et al., WSDM 2017]

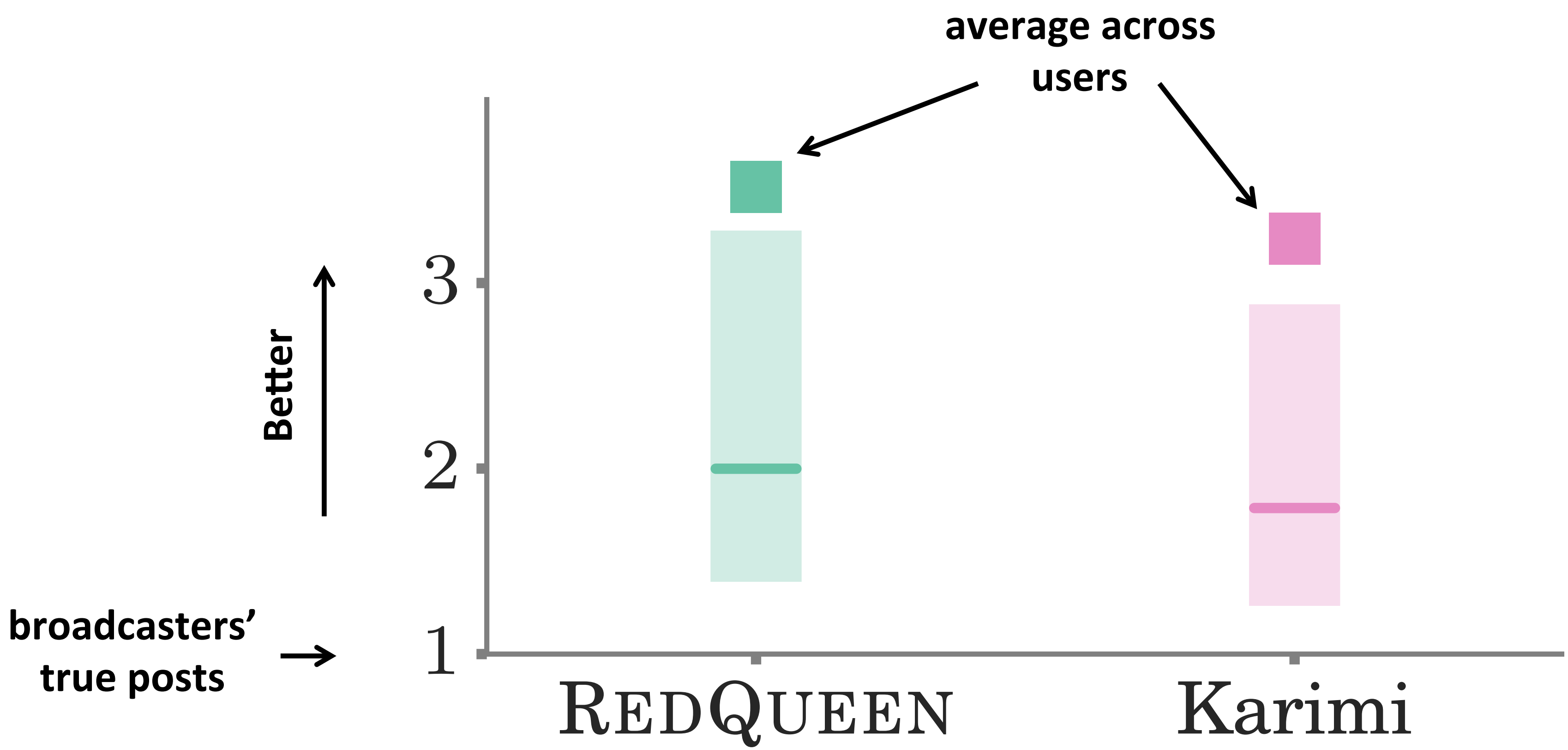
Position over time



It achieves **(i) 0.28x lower average position, in average, than the broadcasters' true posts** and **(ii) lower average position for 100% of the users.**

[Zarezade et al., WSDM 2017]

Time at the top



It achieves **(i) 3.5x higher time at the top, in average, than the broadcasters' true posts** and **(ii) higher time at the top for 99.1% of the users.**

[Zarezade et al., WSDM 2017]

REPRESENTATION: TEMPORAL POINT PROCESSES

1. Intensity function
2. Basic building blocks
3. Superposition
4. Marks and SDEs with jumps

APPLICATIONS: MODELS

1. Information propagation
2. Information reliability
3. Knowledge acquisition

APPLICATIONS: CONTROL

1. Activity shaping
2. When-to-post

Slides/references: learning.mpi-sws.org/ijcai-2017-tutorial