电子科技大学
University of Electronic Science and Technology of China

Information Diffusion and External Influence in Networks

Reporter : Zhongjing Yu

Data Mining Lab, Big Data Research Center, UESTC
Email：junmshao@uestc.edu.cn
http://staff.uestc.edu.cn/shaojunming

Outline:

- Background and Motivation
- Model
  - Proposed Model
  - Infer parameters
- Experiments
- Conclusion

Background and Motivation

数据挖掘实验室
Data Mining Lab

Background :

Social network : a role in the diffusion of the information(Fig.1).

Application : **recommend system(Fig.2), viral transmission,…**

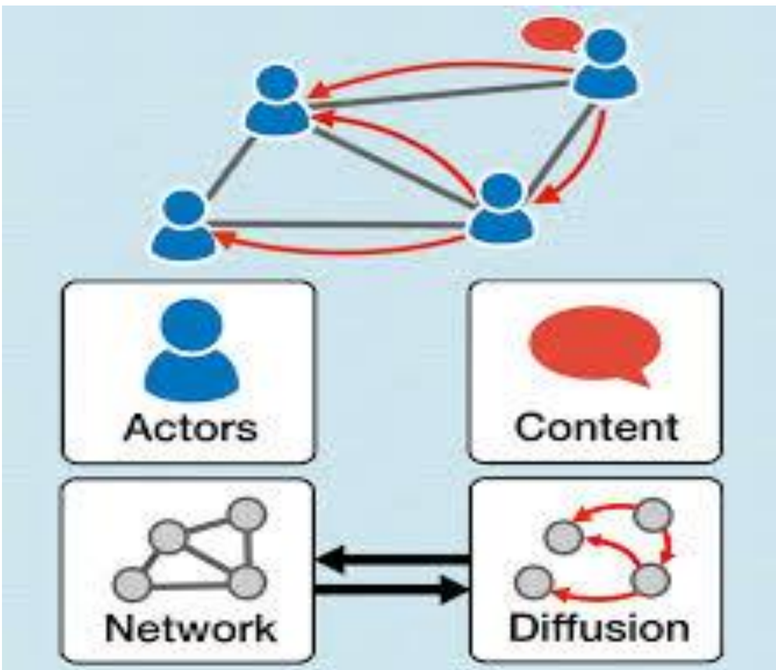


Fig.1

Fig.2

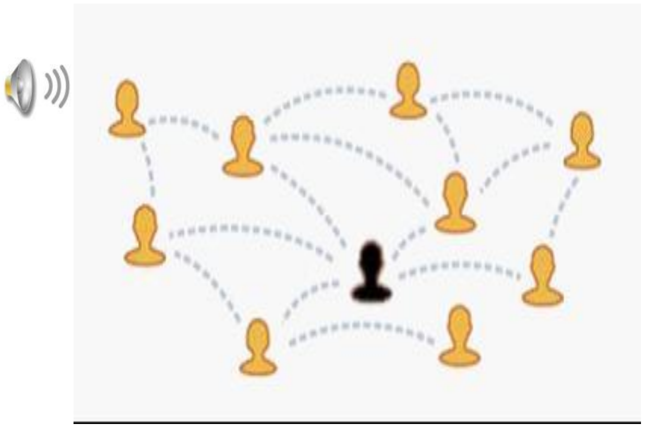
Background and Motivation

Background :

Message pass :

- From a node to another via edges (mostly)
- Through the influence of external sources.

Motivation



Study the process of information diffusion

From a node to another via edges (mostly)

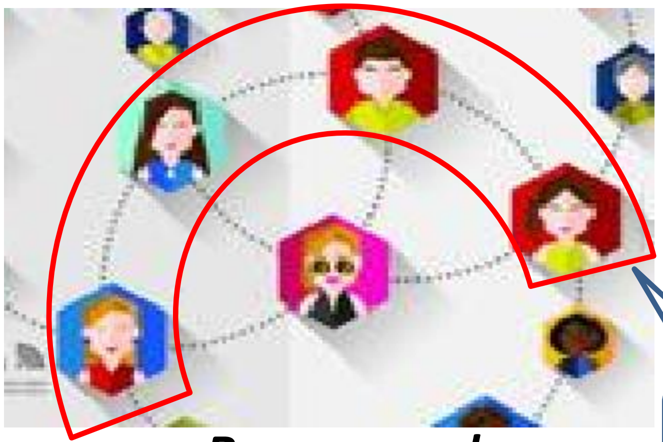
Through the influence of external sources.

Model (*target*)

Background and Motivation

Motivation

Advertisement (External exposures)



Recommend (Internal Exposures)

How to infect ?

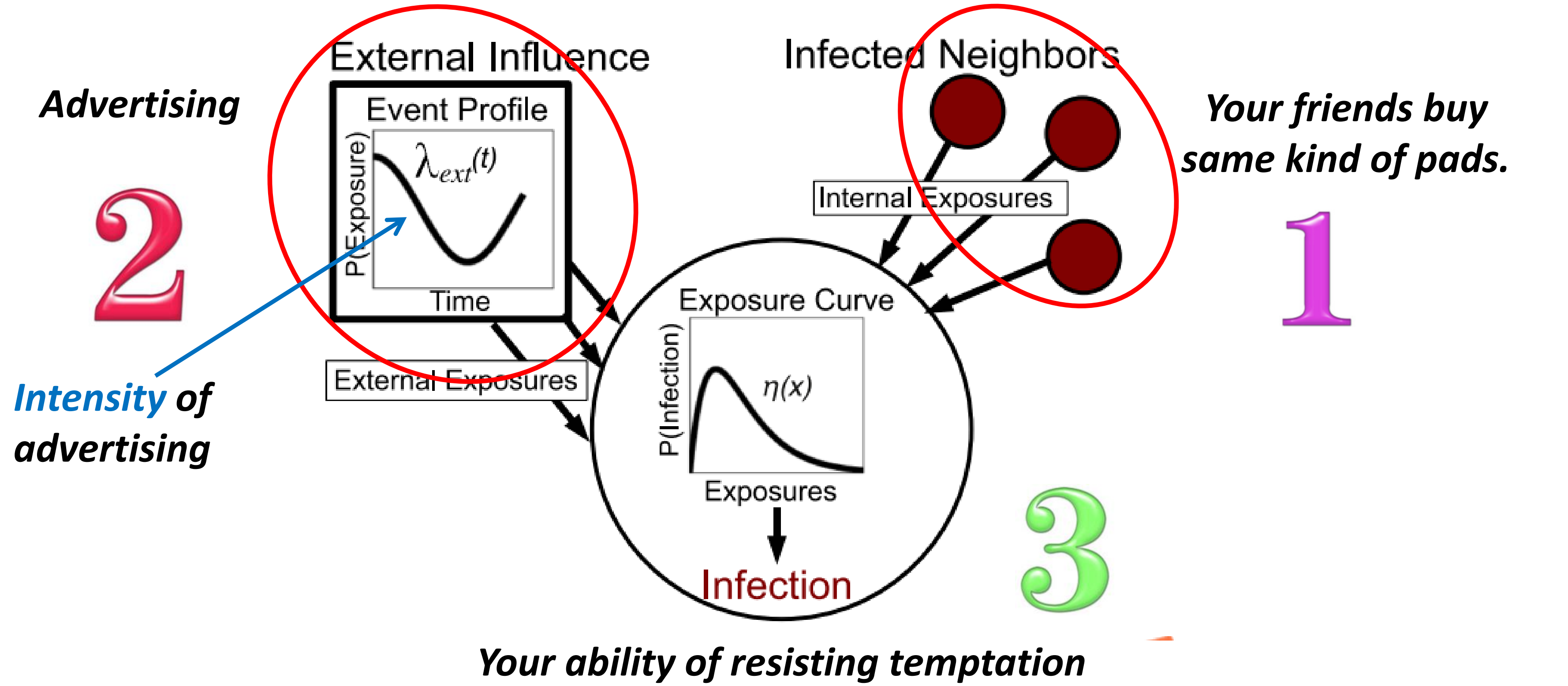
How to infect ?

Buy balabala (infection)

Background and Motivation

数据挖掘实验室
Data Mining Lab

Detail :



Key points:

- External Influence and Infected Neighbors both ***develop exposures***.
- A node is infected with ***different probability*** according to the number of exposures.

Proposed Model(Exposures)

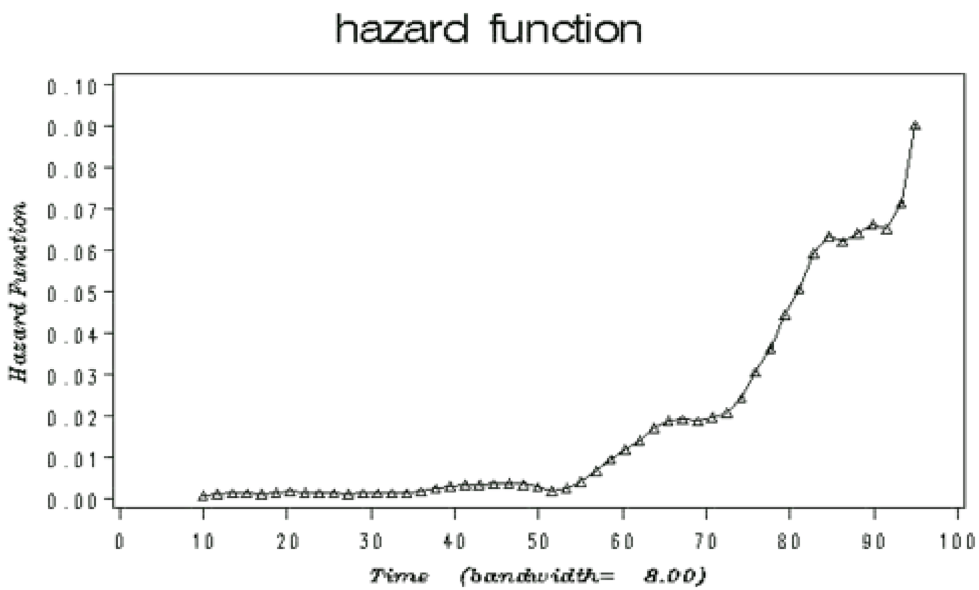
Internal Exposures (introduce a function):

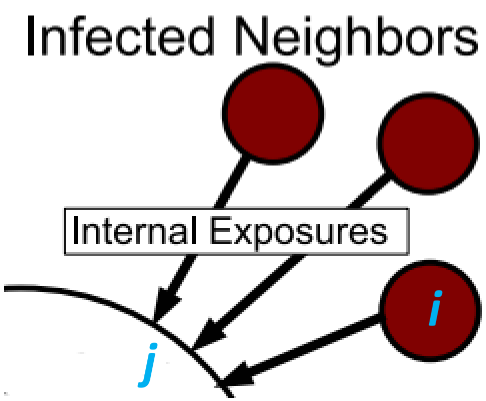
Hazard function : describe a ***distribution*** of the ***length of time*** it takes for an event to occur.

$$h(t) = \frac{N(t)}{M(t)}$$

N(t): *the number of people with* ***infection*** *at t.*

M(t): *the number of people with* ***no infected*** *at t.*





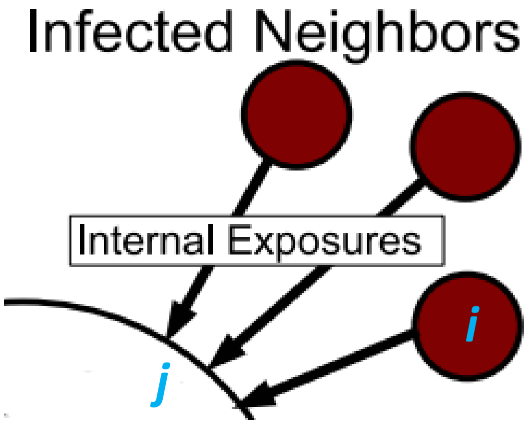
Proposed Model(Exposures)

λ_{int} : internal ***hazard*** function

$$\lambda_{int}(t)\,dt \equiv P\left(i \text{ exposes } j \in [t, t+dt) \middle| \, i \text{ hasn't exposed } j \text{ yet}\right)$$

λ_{int} :*function of the frequency* with which nodes ***check-up*** on each other.

t : the amount of time that has passed since node ***i*** was infected.



数据挖掘实验室
Data Mining Lab

Proposed Model(Exposures)

Excepted number of ***internal*** exposures:

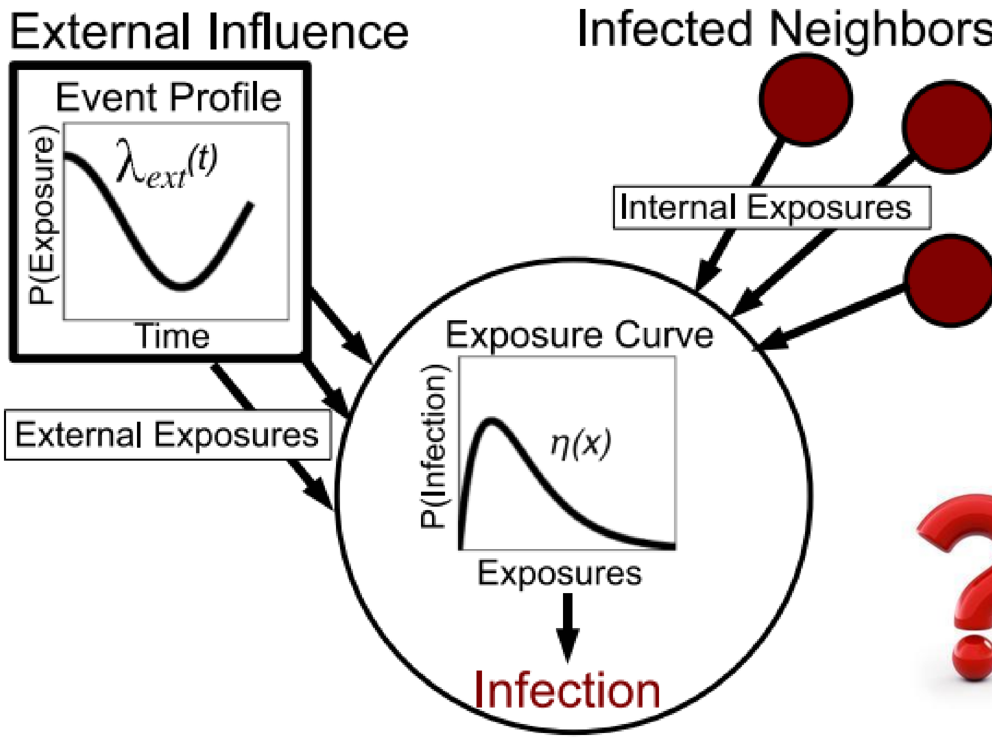
Node ***i*** has received by ***t*** :

$$\Lambda_{int}^{(i)}(t) = \sum_{j;j \text{ is } i\text{'s inf. neighbor}} P(j \text{ exposed } i \text{ before } t)$$

$$= \sum_{j;j \text{ is } i\text{'s inf. neighbor}} \left[1 - \exp\left(-\int_{\tau_j}^{t} \lambda_{int}(s - \tau_j)ds\right)\right]$$

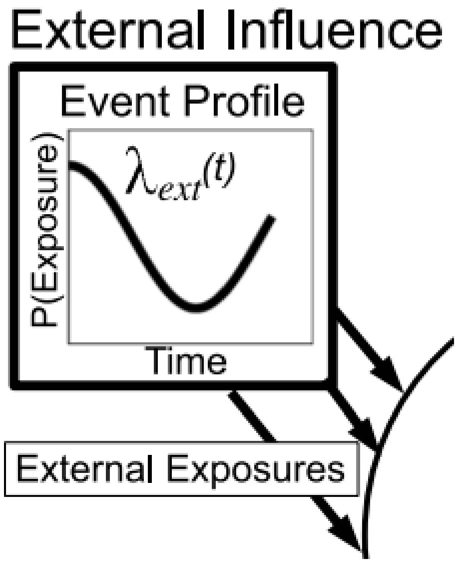
where τ_i is the ***infection time*** of node ***j***

Proposed Model(Exposures)



Proposed Model(Exposures)

External exposures:



$\lambda_{ext}(t)$: the source various in intensity over time called ***event profile***.

$$\lambda_{ext}(t)\,dt \equiv P(i \text{ receives exposure} \in [t, t+dt))$$

The ***probability :***

(receive ***n*** external exposures within $T \bullet \Delta t$)

$$P_{exp}(n; T \cdot \Delta t) = \binom{T}{n} (\lambda_{ext} \cdot \Delta t)^n \cdot (1 - \lambda_{ext} \cdot \Delta t)^{T-n}$$

数据挖掘实验室 Data Mining Lab

Proposed Model(Exposures)

The ***probability :***

(receive ***n*** external exposures within $T \bullet \Delta t$)

$$P_{exp}^{(i)}(n;t) \approx \binom{t/dt}{n} \left(\frac{\Lambda_{ext}(t)}{t} \cdot dt \right)^{n} \cdot \left(1 - \frac{\Lambda_{ext}(t)}{t} \cdot dt \right)^{t/dt-n}$$

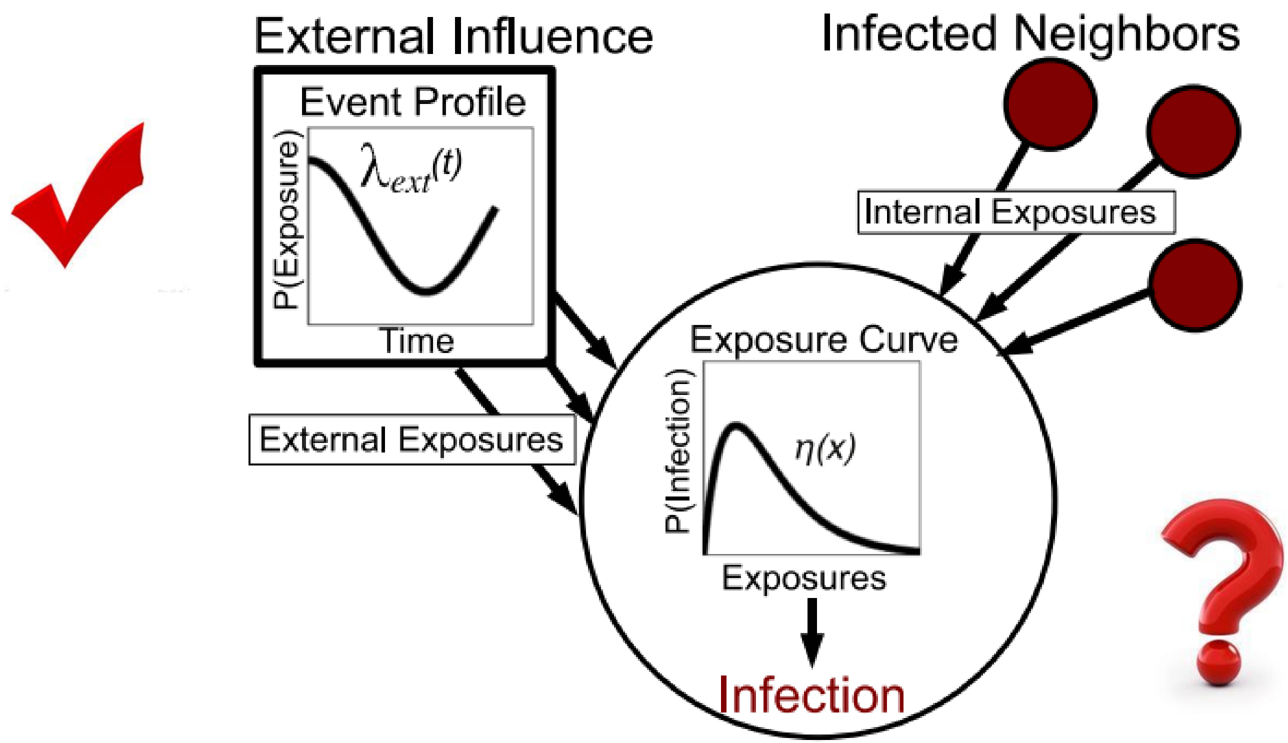
where $\Lambda_{ext}(t) \equiv \int_0^t \lambda_{ext}(s)ds.$ (average)

$\lambda_{ext}(t) + \lambda_{int}^{(i)}(t)$

$$P_{exp}^{(i)}(n;t) \approx \binom{t/dt}{n} \left(\frac{\Lambda_{int}^{(i)}(t) + \Lambda_{ext}(t)}{t} \cdot dt \right)^{n} \times \left(1 - \frac{\Lambda_{int}^{(i)}(t) + \Lambda_{ext}(t)}{t} \cdot dt \right)^{t/dt-n}$$

Exposures = *internal exposures* + *external exposures*

Proposed Model(Exposures)

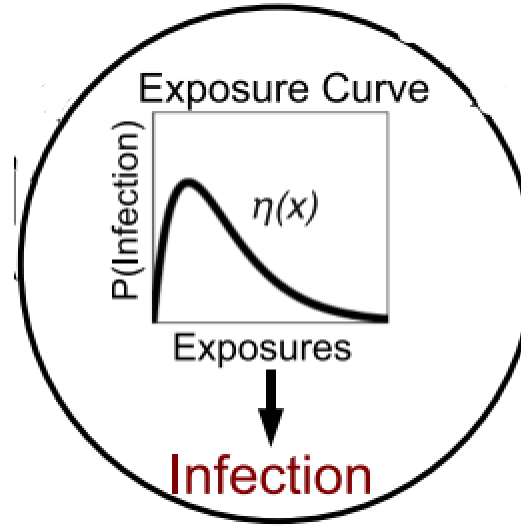


Proposed Model(infection)

Exposure curve:

$$\eta(x) \equiv$$
$$P(\text{node } i \text{ is infected immediately after } x^{th} \text{ exposure})$$

How to define the exposure curve ?



Some properties:

- *No exposure , no infection.*
- *Only one max infection rate.*
- ***Two factors:***
  - *Measure of how infectious a contagion.*
  - *Measure of the contagion's enduring relevancy.*

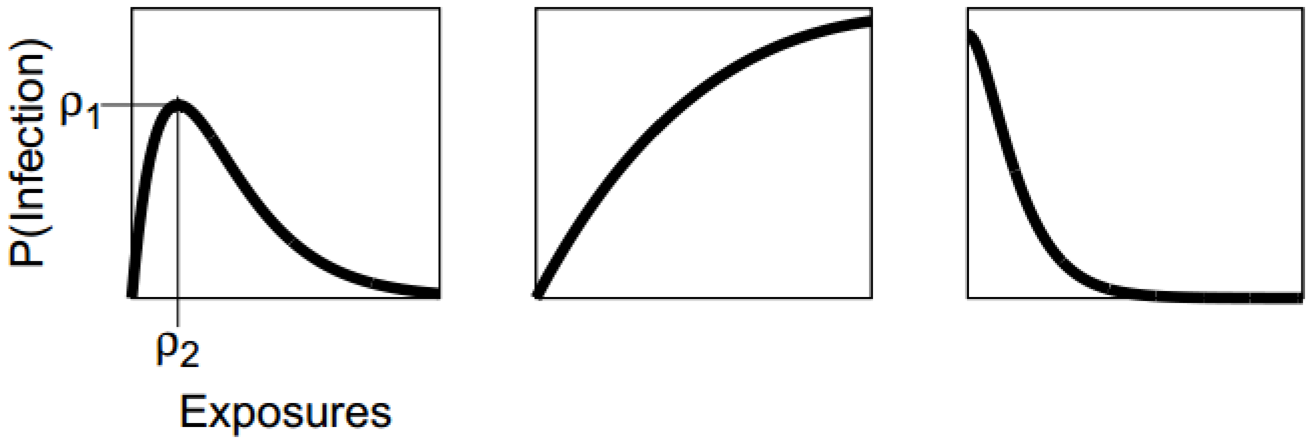
Model

Proposed Model(infection)

Exposure curve:

$$\eta(x) = \frac{\rho_1}{\rho_2} \cdot x \cdot \exp\left(1 - \frac{x}{\rho_2}\right).$$

- *Measure of how infectious a contagion----- ρ_1*
- *Measure of the contagion's enduring relevancy----- ρ_2*



Proposed Model(from *exposures* to *infections*)

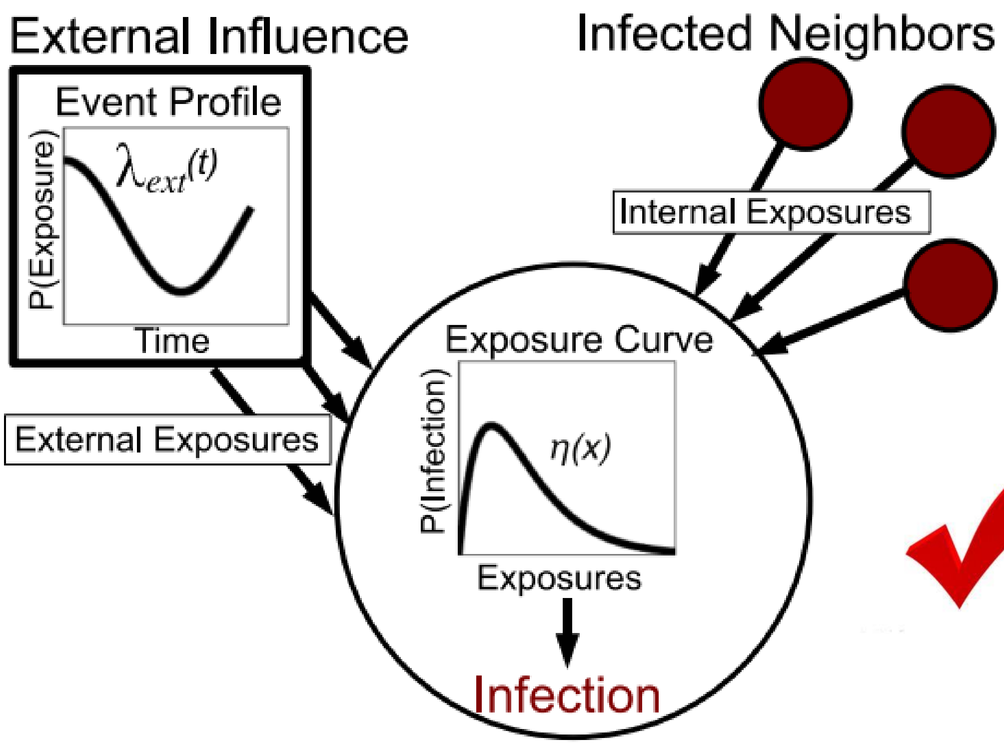
Construct the probability functions :

Idea : The probability of how long time between node i influenced(τ_i) and have infected (t).

$$F^{(i)}(t) \equiv P(\tau_i \leq t) \approx P_{\text{exp}}^{(i)}(n;t)$$

$$\begin{aligned} F^{(i)}(t) &= \sum_{n=1}^{\infty} P[i \text{ has } n \text{ exp. }] \times P[i \text{ inf. } | i \text{ has } n \text{ exp. }] \\ &= \sum_{n=1}^{\infty} P_{exp}^{(i)}(n;t) \times \left[1 - \prod_{k=1}^{n} [1-\eta(k)]\right]. \end{aligned}$$

Model



Finish model, any questions?

Infer parameters

Given :

- Network
- The ***infection times*** for each node

Infer parameters :

- Event profile $\lambda_{ext}(t)$
- Parameters of $\eta(x)$: ρ_1 , ρ_2 .

Idea :

- when the exposures curve is known.

  ⟹ ***infer*** the event profile $\lambda_{ext}(t)$

- when the event profile is known

  ⟹ ***infer*** the exposures curve $\eta(x)$

Infer parameters ($\lambda_{ext}(t)$)

Assume : given

$S(t)$: the number of nodes with uninfected at ***t***.

$$S(t_k) = \sum_{i=1}^{N} P(\text{ node } i \text{ not infected by time } t)$$

$$= \sum_{i=1}^{N} \sum_{n=1}^{\infty} P_{exp}^{(i)}(n; t_k) \prod_{k=1}^{n} [1 - \eta(k)]$$

$$\approx \ldots$$

$$\approx \sum_{i} \exp\left(- \int_{0}^{\Lambda_k + \Lambda_{int}^{(i)}(t_k)} \eta(y) dy \right)$$

Get from data

$$\Lambda_k = \Lambda_{ext}(t) \equiv \int_0^t \lambda_{ext}(s) ds$$

Model

Infer parameters ($\eta(x)$ -- ρ_1, ρ_2)

Assume : given Λ_{ext}

Idea : fix ρ_2, solve for a ρ_1 according to ***maximizes log-likelihood***.

$$F^{(i)}(t) = \sum_{n=1}^{\infty} P[i \text{ has } n \text{ exp. }] \times P[i \text{ inf. } | i \text{ has } n \text{ exp. }]$$

$$= \sum_{n=1}^{\infty} P_{exp}^{(i)}(n;t) \times \left[1 - \prod_{k=1}^{n} [1 - \eta(k)]\right].$$

数据挖掘实验室
Data Mining Lab

Infer parameters

Infer $\eta(x)$ ***through max log-likelihood***

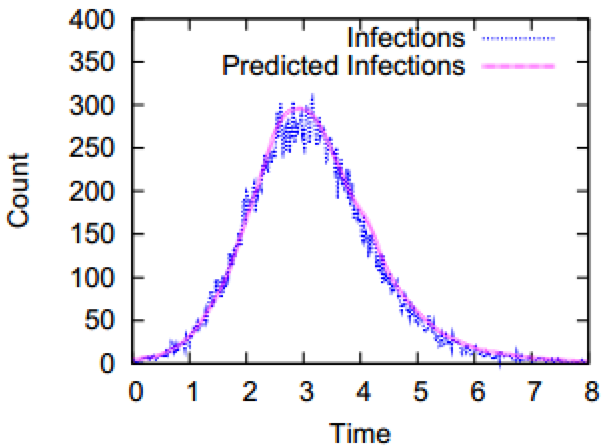
Algorithm 1 Model Parameter Inference

Initialize $\Lambda_{ext}(t)$, ρ^1_{final}, ρ^2_{final}, $\mathcal{L}_{max}$
for $\rho_2 = 1 \rightarrow \rho_{max}$ **do**
  Initialize ρ_1
  while not converged **do**
    $\rho_1 \leftarrow$ Solution to Eq. 12 using ρ_2, $\Lambda_{ext}(t)$
    $\Lambda_{ext}(t) \leftarrow$ Solution to Eq. 10 using ρ_1, ρ_2.
  end while
  $\mathcal{L} \leftarrow \textit{Log-Likelihood}(\Lambda_{ext}(t), \rho_1, \rho_2)$
  if $\mathcal{L} \geq \mathcal{L}_{max}$ **then**
    $\mathcal{L}_{max} \leftarrow \mathcal{L}$
    $\rho^1_{final} \leftarrow \rho_1$
    $\rho^2_{final} \leftarrow \rho_2$
  end if
end for
$\Lambda_{ext}(t) \leftarrow$ Solution to Eq. 10 using ρ^1_{final}, ρ^2_{final}.

Infer $\Lambda_{ext}(t)$ ***through*** $\eta(x)$ ***and*** $S(t)$

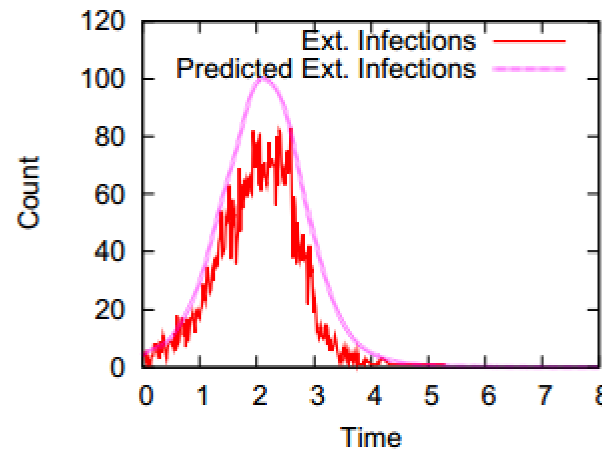
Experiments

Synthetic data ($\lambda_{\text{int}}(t)=t$)

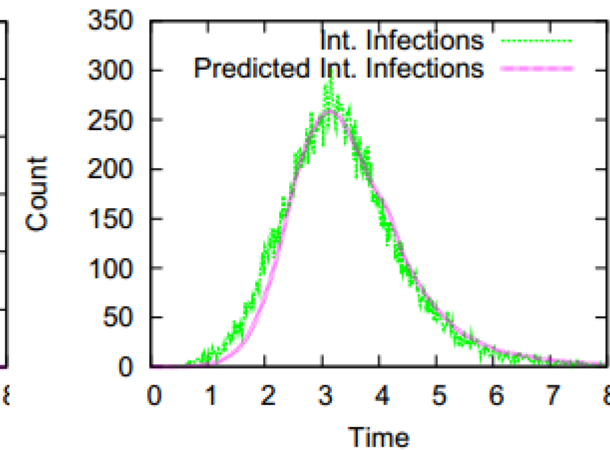


(a) All Infections

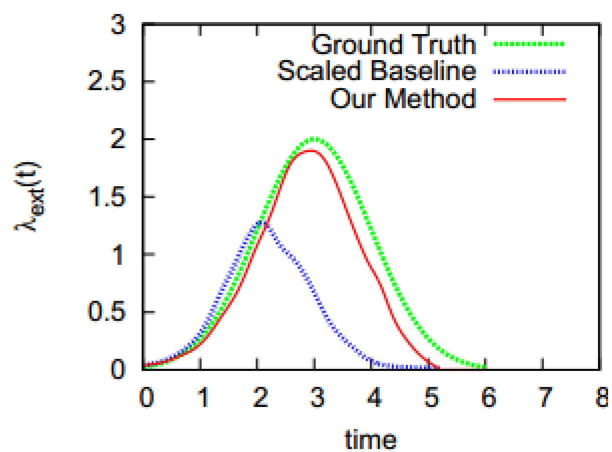
Given (a), infer (b)-(e)



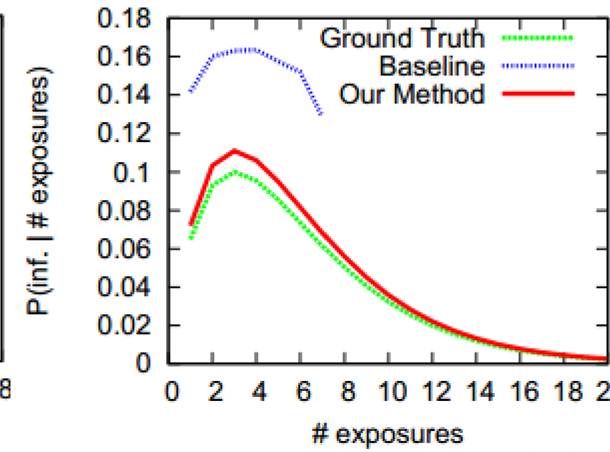
(b) External Infections



(c) Internal Infections



(d) Event Profile $\lambda_{ext}(t)$

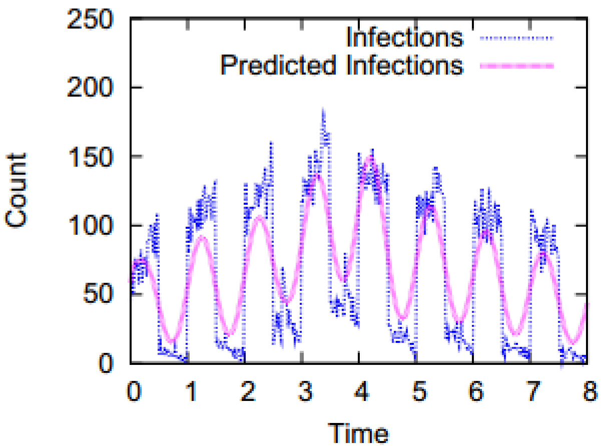


(e) Exposure Curve $\eta(x)$

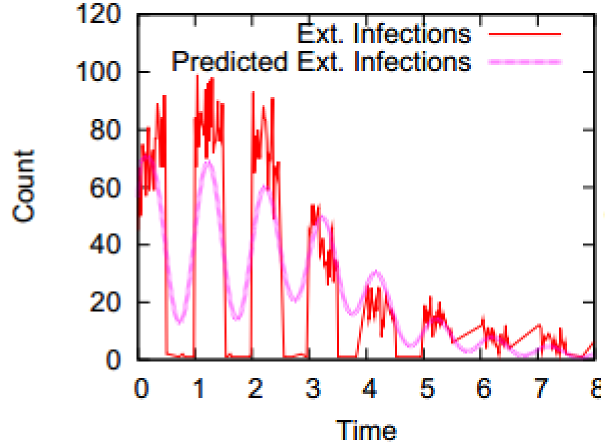
Experiments

数据挖掘实验室
Data Mining Lab

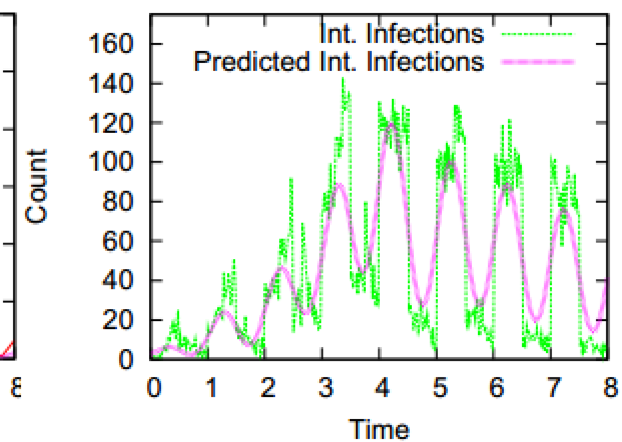
Synthetic data ($\lambda_{int}(t) = \frac{1}{t}$)



(f) All Infections

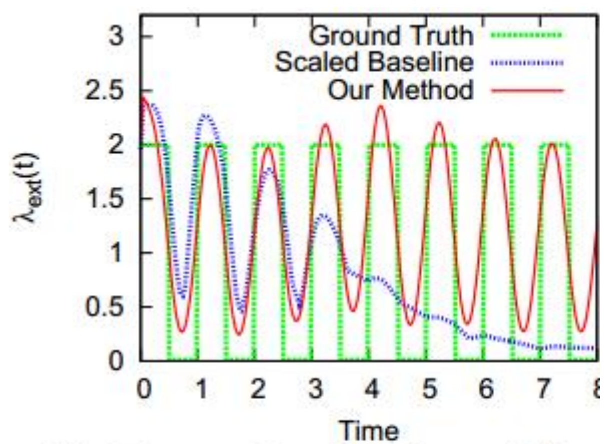


(g) External Infections

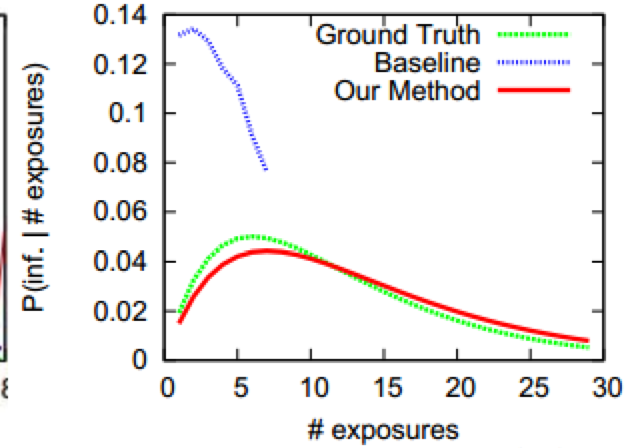


(h) Internal Infections

Given (f), infer (g)-(j)

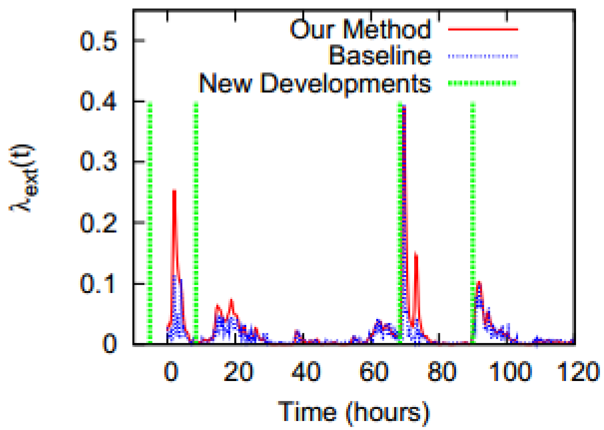


(i) Event Profile $\lambda_{ext}(t)$

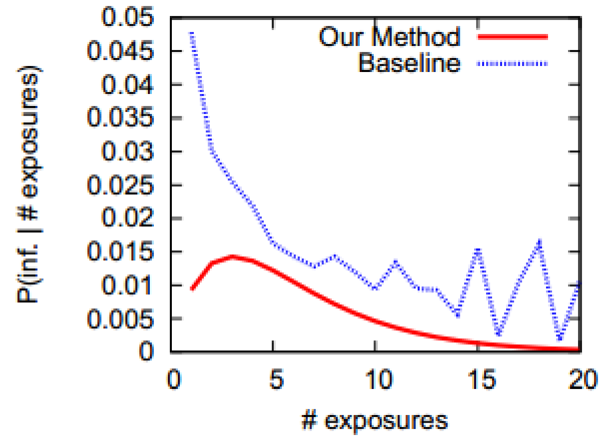


(j) Exposure Curve $\eta(x)$

Real data (Twitter)



(a) Event Profile $\lambda_{ext}(t)$



(b) Exposure Curve $\eta(x)$

Figure 4: The model fitted to a single contagion representing URLs related to the Tucson, Arizona shootings. The green vertical lines designate when four distinct developments related to the shooting event occurred.

数据挖掘实验室
Data Mining Lab

Real data (Twitter)

-

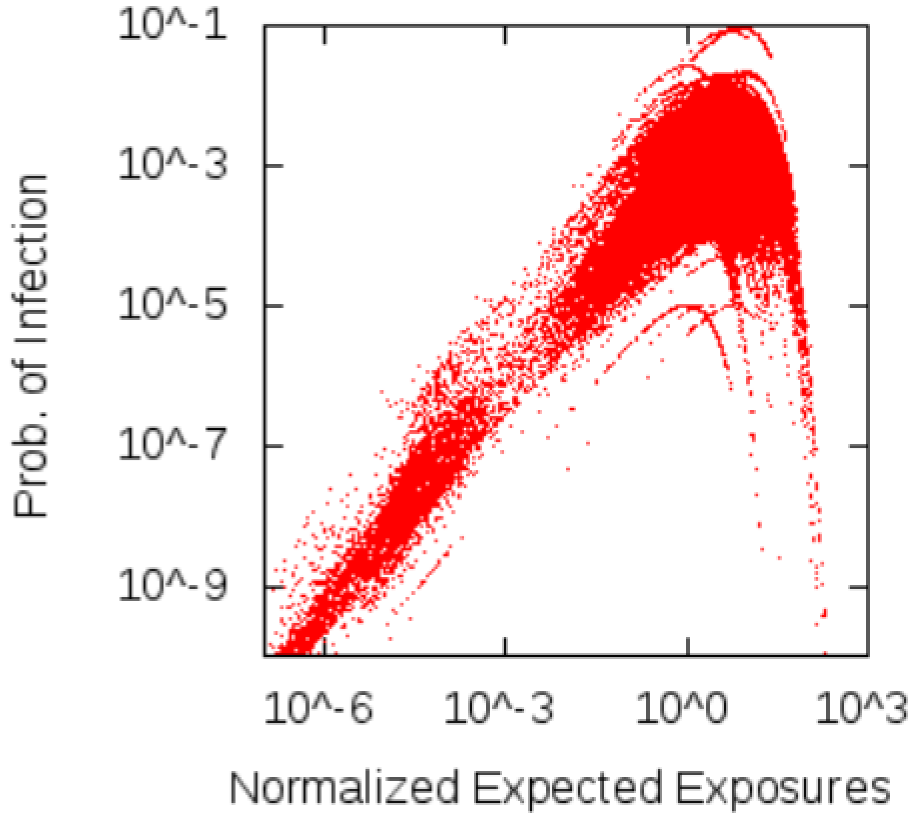
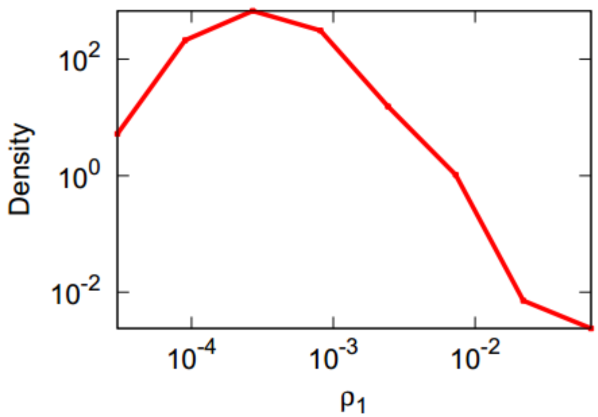
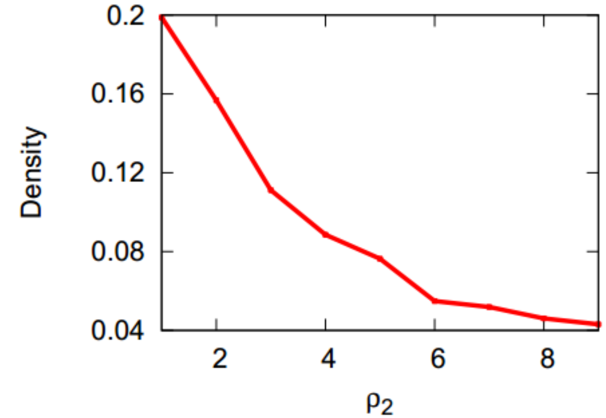


Figure 5: An aggregation of the exposure curves for each URL. Upon the infection of each user, the expected number of exposures received by the user divided by ρ_2 was plotted against inferred infection probability.

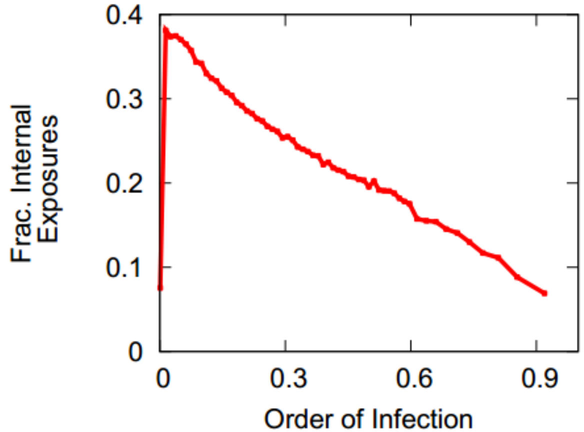
Real data



(a) Distribution of ρ_1



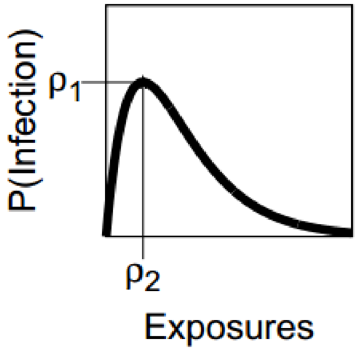
(b) Distribution of ρ_2



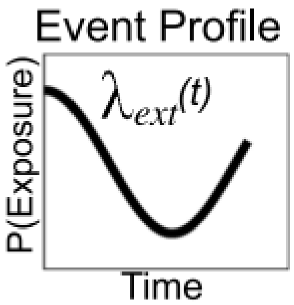
(c) Fraction of Internal Exposures

Conclusion

➢Infer the ***shape*** of influence functions.



➢Capture the ***external influence***.



➢More accurate ***description*** of the real network diffusion process

Thanks

Chongming Gao
Yingcai Experimental School
gchongming@126.com