Information Diffusion on Social Networks

SMART Summer School 2017

Sylvain Lamprier

LIP6 - UPMC

MLIA Team

1. Information Diffusion
   - Diffusion on Networks
   - Tasks
   - Challenges
   - Diffusion Models

2. The Independent Cascade Model
   - Learning
   - Limits
   - Extensions

3. Deep-Learning for Diffusion
   - Embedded IC
   - Predictive models
   - Recurrent Neural Networks for Diffusion

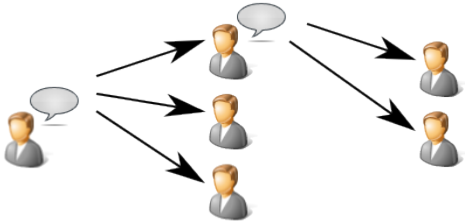
1. Information Diffusion
   - Diffusion on Networks
   - Tasks
   - Challenges
   - Diffusion Models
2. The Independent Cascade Model
3. Deep-Learning for Diffusion

Diffusion on Networks

Fundamental Process on Networks:

- Capture of the dynamics
- How information transits on the network ?

Diffusion = Iterative message passing process



$\Rightarrow$ Defines a diffusion **Cascade**

- Tree structure

Diffusion on Networks

- Diffusion Items
  - Word of mouth / viral marketing
  - Virus or diseases
  - News, opinions, rumors, ..
  - Topics / videos / hashtags / links
  - Language models / expressions
  - Behaviors
  - Errors / Problems
  - ...
- Diffusion Episode = Set of linked events that occur on the network through time

Diffusion

The study of diffusion dynamics has a long history:

- Agricultural practices (1943)
  - Study about the adoption of a new kind of hybrid corn by 259 Iowa's farmers
  - Conclusion: the relationships network plays an important role for the adoption of new products
- Medical practices (1966)
  - Study about the adoption of new drugs by Illinois' doctors
  - Conclusion: Word of mouth is more effective than scientific studies in convincing the doctors
- Psychological effects of opinions on the entourage of persons (1958)
- Contagion of obesity (2007)
  - Having an overweight friend increases our probability of becoming obese by 57% !

Homophily vs. Influence

- Homophily
  - Two connected users tend to have similar behaviors
- Influence
  - The behavior of a user has an impact on the future behavior of his neighborhood

$\Rightarrow$ Temporality is crucial to distinguish influence (diffusion) from homophily (recommendation)

  - If one observe relations of precedence between events: influence

Diffusion vs. Recommendation

Consider a network of product reviewing by users:

- Object of the diffusion: a Product
  - Nodes = Users
  - Infection of a node = a user likes the product
  - Influence relationships between users

    $\Rightarrow$ When a product is liked by this user, it then tends to be liked by these other ones in the future

- Object of the diffusion: a User
  - Nodes = Products
  - Infection of a node = an item has been liked by the user
  - Temporal recommendation

    $\Rightarrow$ When somebody liked this product, she then tends to like these related others in the future

Diffusion Tasks

Buzz prediction - Will the content impact an important number of users ? [Chen et al.,2013]

$$\underbrace{\begin{pmatrix} 1 \\ 0 \\ 0 \\ \ldots \\ 0 \\ 1 \\ 0 \end{pmatrix}}^{\textit{Source Users}} + \overbrace{\begin{pmatrix} \omega_1 \\ \omega_2 \\ \ldots \\ \omega_{d-1} \\ \omega_d \end{pmatrix}}^{\textit{Content}} \xrightarrow{\quad f_\theta \quad} \{0, 1\}$$

Diffusion Tasks

Volume prediction - How many users will be eventually infected? [Tsur and Rappoport, 2012]

$$\begin{array}{c} \textit{Source Users} \\ \begin{pmatrix} 1 \\ 0 \\ 0 \\ \dots \\ 0 \\ 1 \\ 0 \end{pmatrix} \\ + \\ \textit{Content} \\ \begin{pmatrix} \omega_1 \\ \omega_2 \\ \dots \\ \omega_{d-1} \\ \omega_d \end{pmatrix} \end{array} \xrightarrow{\qquad\qquad f_\theta \qquad\qquad} \mathbb{N}$$

Diffusion Tasks

Infection prediction - Which users will be eventually infected? [Bourigault et al., 2016]

$$\begin{array}{ccc} \textit{Source Users} & & \\ \begin{pmatrix} 1 \\ 0 \\ 0 \\ \dots \\ 0 \\ 1 \\ 0 \end{pmatrix} & \xrightarrow{\quad f_\theta \quad} & \begin{array}{c} \textit{Final Users} \\ \begin{pmatrix} 1 \\ 1 \\ 0 \\ \dots \\ 1 \\ 1 \\ 1 \end{pmatrix} \end{array} \\ + & & \\ \textit{Content} & & \\ \begin{pmatrix} \omega_1 \\ \omega_2 \\ \dots \\ \omega_{d-1} \\ \omega_d \end{pmatrix} & & \end{array}$$

Diffusion Tasks

Spread prediction - How will evolve the spread of the content?

$$
\begin{array}{c}
\textit{Source Users} \\
\begin{pmatrix} 1 \\ 0 \\ 0 \\ \dots \\ 0 \\ 1 \\ 0 \end{pmatrix} \\
+ \\
\textit{Content} \\
\begin{pmatrix} \omega_1 \\ \omega_2 \\ \dots \\ \omega_{d-1} \\ \omega_d \end{pmatrix}
\end{array}
\xrightarrow{\quad f_\theta \quad}
\begin{array}{c}
\textit{Infected Users per Step} \\
\begin{array}{ccccc} 1 & 2 & \dots & T-1 & T \end{array} \\
\begin{pmatrix}
1 & 1 & \dots & 1 & 1 \\
0 & 0 & \dots & 0 & 1 \\
0 & 0 & \dots & 0 & 0 \\
\dots & \dots & \dots & \dots & \dots \\
0 & 1 & \dots & 1 & 1 \\
1 & 1 & \dots & 1 & 1 \\
1 & 1 & \dots & 1 & 1
\end{pmatrix}
\end{array}
$$

Cascade prediction - Which links will follow the content?

Source Users

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ \dots \\ 0 \\ 1 \\ 0 \end{pmatrix} \xrightarrow{f_\theta} \{0,1\}^{|R|}$$

$$+$$

Content

$$\begin{pmatrix} \omega_1 \\ \omega_2 \\ \dots \\ \omega_{d-1} \\ \omega_d \end{pmatrix}$$

with R the set of relationships

Diffusion Tasks

Source prediction - Who are the sources of a given content ? [Shah and Zaman, 2010].

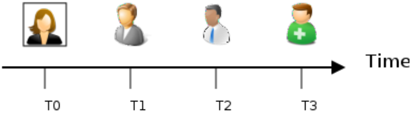
$$\underbrace{\begin{pmatrix} 1 \\ 1 \\ 0 \\ \dots \\ 1 \\ 1 \\ 1 \end{pmatrix}}_{}^{\textit{Infected Users}} + \overset{\textit{Content}}{\begin{pmatrix} \omega_1 \\ \omega_2 \\ \dots \\ \omega_{d-1} \\ \omega_d \end{pmatrix}} \xrightarrow{f_\theta} \overset{\textit{Source Users}}{\begin{pmatrix} 1 \\ 0 \\ 0 \\ \dots \\ 0 \\ 1 \\ 0 \end{pmatrix}}$$

Diffusion Tasks

- Other tasks
  - Link Detection - Which are the main diffusion links of the network? [Gomez-Rodriguez et al., 2011]
  - Opinion Leaders Detection - Who are the most influential users of the network ? [Kempe et al., 2003]
  - Diffusion Maximization - To whom should one give a content to maximize its spread ? [Kempe et al., 2003]
  - Firefighter Problem - How to stop the diffusion of a content ? [Anshelevich et al., 2009]
  - ...

Diffusion on networks

- Challenges
  - The diffusion cascade is usually hidden
    - We do not know who influenced whom
  - What we get is the dated (first) participation of users to the diffusion (diffusion **episode**)



$\Rightarrow$ We only know who participated in what and when

$\Rightarrow$ Model the diffusion dynamics of a network = Learning problem of influence relationships from incomplete data

- Challenges
  - Complex dynamics for rare events
    - Difficult learning
    - Stochastic models rather than deterministic ones
  - Influence distributions depend on the content
    - Different behaviors w.r.t. different contents
    - *e.g.*, Paul can have a strong influence on Pierre for sport but few for politics
  - Closed World Hypothesis rarely valid
    - Diffusion can take place on various media simultaneously
  - Inter-dependency / concurrency of diffusion processes
    - Some process can be impacted by others
  - Dynamicity of the network
    - New users / New relationships
    - Evolution of the influence relationships through time

Diffusion Models

- Models Macro : global statistics on the diffusion (size, speed)
  - Bass : adoption of a product
  - SIR : virus diffusion
- Models Micro : focus on users of the network *[Kempe et al., 2003]*
  - Linear Threshold (LT) : Receiver-centric
  - Independent Cascade (IC) : Transmitter-centric

The Bass model

Bass, 1969
Evolution of the rate of users $i(t)$ that have adopted a product a time t :

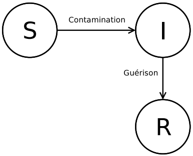
$$\frac{\partial i}{\partial t}(t) = \underbrace{p \times (1 - i(t))}_{\text{Spontaneous Adoptions}} + \underbrace{q \times (i(t) \times (1 - i(t)))}_{\text{Word of Mouth}}$$

- p : Probability that a user adopts a product from ads
- q : probability that a user adopts a product from a neighbor

  - Bass reports values $p = 0.03$ and $q = 0.38$ on average

The model SIR

Epidemiological model. Each user can be in 3 different states.



- *Susceptible* : not infected by the disease;
- *Infected* : infected by the disease;
- *Recovered* : cured and immunized.

Evolution of the system

$$\begin{cases} \frac{\partial S}{\partial t} = -p.SI \\ \frac{\partial I}{\partial t} = p.SI - r.I \\ \frac{\partial R}{\partial t} = r.I \end{cases}$$

- p : transmission probability
- r : probability of cure

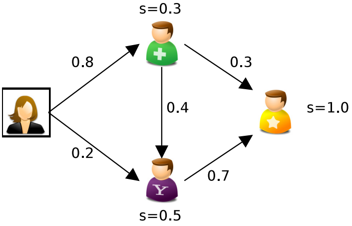
$\rightarrow$ Can also be applied on information diffusion on networks

The Linear Threshold Model *[Granovetter, 1973]*

Micro-model of diffusion

- Hypothesis: Additive Influence
- Links associated to influence weights $\theta_{i,j}$
- Nodes associated to (stochastic) thresholds γ_j
- Iterative model:

$\Rightarrow$ User *j* is infected at step *t* if: $\sum_{i \in Preds(j,t)} \theta_{i,j} \geq \gamma_j$

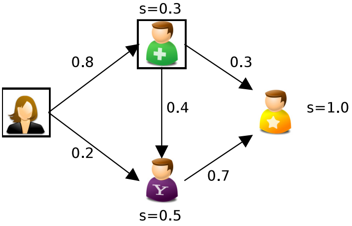


The Linear Threshold Model *[Granovetter, 1973]*

Micro-model of diffusion

- Hypothesis: Additive Influence
- Links associated to influence weights $\theta_{i,j}$
- Nodes associated to (stochastic) thresholds γ_j
- Iterative model:

$\Rightarrow$ User *j* is infected at step *t* if: $\sum_{i \in Preds(j,t)} \theta_{i,j} \geq \gamma_j$

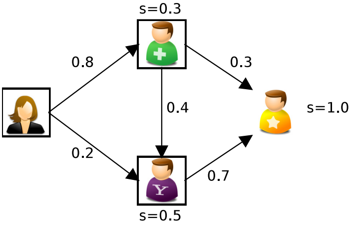


The Linear Threshold Model *[Granovetter, 1973]*

Micro-model of diffusion

- Hypothesis: Additive Influence
- Links associated to influence weights $\theta_{i,j}$
- Nodes associated to (stochastic) thresholds γ_j
- Iterative model:

$\Rightarrow$ User j is infected at step t if: $\sum_{i \in Preds(j,t)} \theta_{i,j} \geq \gamma_j$

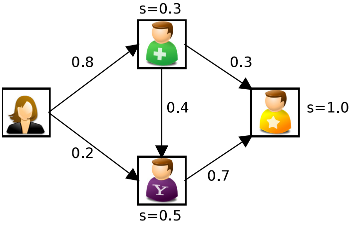


The Linear Threshold Model *[Granovetter, 1973]*

Micro-model of diffusion

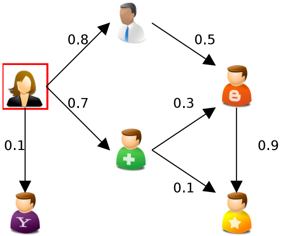
- Hypothesis: Additive Influence
- Links associated to influence weights $\theta_{i,j}$
- Nodes associated to (stochastic) thresholds γ_j
- Iterative model:

$\Rightarrow$ User j is infected at step t if: $\sum_{i \in Preds(j,t)} \theta_{i,j} \geq \gamma_j$



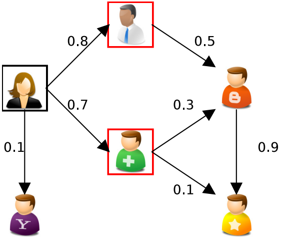
The Independent Cascade Model (IC)

- Micro-model of diffusion
  - Hypothesis: influences are independent events
  - Infection probabilities $\theta_{u,v}$ are defined on every edge of the graph
  - After its infection, a user *u* gets a unique chance to infect each of its successors in the network for the next step



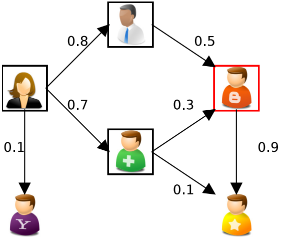
The Independent Cascade Model (IC)

- Micro-model of diffusion
  - Hypothesis: influences are independent events
  - Infection probabilities $\theta_{u,v}$ are defined on every edge of the graph
  - After its infection, a user *u* gets a unique chance to infect each of its successors in the network for the next step



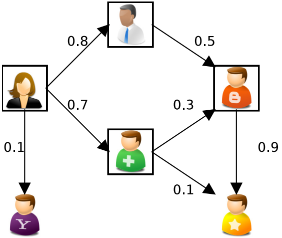
The Independent Cascade Model (IC)

- Micro-model of diffusion
  - Hypothesis: influences are independent events
  - Infection probabilities $\theta_{u,v}$ are defined on every edge of the graph
  - After its infection, a user *u* gets a unique chance to infect each of its successors in the network for the next step



The Independent Cascade Model (IC)

- Micro-model of diffusion
  - Hypothesis: influences are independent events
  - Infection probabilities $\theta_{u,v}$ are defined on every edge of the graph
  - After its infection, a user *u* gets a unique chance to infect each of its successors in the network for the next step



IC Extensions

- Continuous time
  - Saito 2009 (CTIC),Gomez-Rodriguez 2011 (NetRate)...
- Inclusion of content
  - Barbieri 2013 (TIC)
- Inclusion of users profiles
  - Guille 2012, Saito 2011...
- Concurrent Diffusions
  - Myers 2012, Bharathi 2007...
- etc...

IC: Learning the Influence Relationships

Which inputs ?

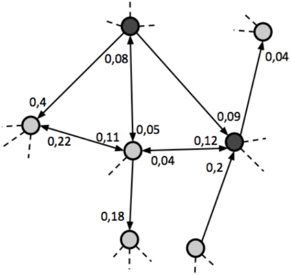
- Training set of episodes
  - Diffusion episode = List of timestamps of infection
- Graph of the network
  - Explicit relationships can help to drive the learning but...
    - Sometimes no available relationship
    - Explicit relations do not always correspond to the main influence relationships of the network [Ver Steeg et al., 2013]

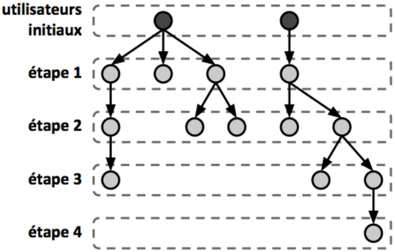
  $\Rightarrow$ Diffusion Link detection approaches: e.g., NetInf [Gomez Rodriguez et al., 2010]
    - Search of the maximum spanning tree for each episode
    - Selection of the n links the most used by the trees

  $\Rightarrow$ Or use the complete graph of the nodes if possible ($n \times (n-1)$ relations)
    - Can be restricted to links with at least one example of possible diffusion in the training set

IC: Learning the Influence Relationships

- Independent Cascade Model (IC)
  - Inference from an influence graph with probabilities defined on edges





- Infection probability for v at step t = Probability that at least one user infected at step $t-1$ succeeds in influencing v:

$$P_t(v) = 1 - \prod_{u \in Preds(v) \wedge t_u = t-1} 1 - \theta_{u,v}$$

IC: Learning the Influence Relationships

$\Rightarrow$ Find parameters $\theta_{u,v}$ maximizing the model likelihood according to a training set of diffusion episodes $\mathcal{D}$

[Saito et al., 2008] :

$$L(\mathcal{D};\theta) = \prod_{D\in\mathcal{D}} \prod_{u\in D} P_{t_u^D}(u) \prod_{\substack{(u,v),u\in D\wedge v\in Succs(u)\wedge \\ ((v\notin D)\vee(v\in D\wedge t_v^D>t_u^D+1))}} 1-\theta_{u,v}$$

with $P_{t_u^D}(u) = 1 - \prod_{v\in Preds(u)\wedge t_v^D=t_u^D-1} 1-\theta_{v,u}$

Or equivalently:

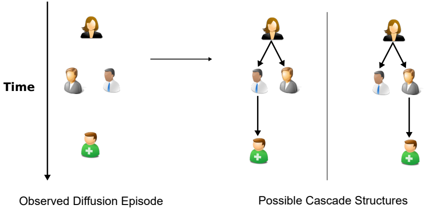
$$\log\left(L(\mathcal{D};\theta)\right) = \sum_{D\in\mathcal{D}} \sum_{u\in D} \log P_{t_u^D}(u) + \sum_{\substack{(u,v),u\in D\wedge v\in Succs(u)\wedge \\ ((v\notin D)\vee(v\in D\wedge t_v^D>t_u^D+1))}} \log\left(1-\theta_{u,v}\right)$$

$\Rightarrow$ Difficult to maximize

Missing Information

Diffusion Episodes

- One only know *when* each user was infected
- Missing information: *who* infected *whom*



- If this information was available, the maximization problem would be easy

$\Rightarrow$ An Expectation-Maximization algorithm (EM) was proposed by Saito in 2008 for solving the problem

Expectation-Maximization for IC [Saito et al., 2008]

1. Expectation (E) of the log-likelihood according to current parameters $\hat{\theta}$

$$Q\left(\boldsymbol{\theta};\hat{\boldsymbol{\theta}}\right) = E_{\mathbf{Z}|\mathbf{X},\hat{\boldsymbol{\theta}}}\left[L\left((\mathbf{X},\mathbf{Z});\boldsymbol{\theta}\right))|\hat{\boldsymbol{\theta}}\right]$$

with **Z** containing all hidden (binary) transmission outcomes.

$$P(z^D_{u,v}=1|D) = \frac{\hat{\theta}_{u,v}}{\hat{P}_{t^D_v}(v)} \quad \text{and} \quad P(z^D_{u,v}=0|D) = 1 - \frac{\hat{\theta}_{u,v}}{\hat{P}_{t^D_v}(v)}$$

with $\hat{P}_{t^D_u}(u) = 1 - \prod_{v\in Preds(u)\wedge t^D_v = t^D_u - 1} 1 - \hat{\theta}_{v,u}$

Thus:

$$Q\left(\theta;\hat{\theta}\right) = \sum_{D\in\mathcal{D}} \Phi^D\left(\theta;\hat{\theta}\right) + \sum_{\substack{(u,v),u\in D\wedge v\in Succs(u)\wedge\\ ((v\notin D)\vee(v\in D\wedge t^D_v > t^D_u+1))}} \log\left(1-\theta_{u,v}\right)$$

with

$$\Phi^D\left(\theta;\hat{\theta}\right) = \sum_{\substack{(u,v)\in D^2,\\ v\in Succs(u)\\ \wedge t^D_v = t^D_u+1}} \frac{\hat{\theta}_{u,v}}{\hat{P}_{t^D_v}(v)} \log(\theta_{u,v}) + \left(1 - \frac{\hat{\theta}_{u,v}}{\hat{P}_{t^D_v}(v)}\right) \log(1-\theta_{u,v})$$

2. Maximization (M) of the log-likelihood expectation :

$$\hat{\boldsymbol{\theta}} \leftarrow \arg\max_{\boldsymbol{\theta}} \left(Q\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}\right)\right)$$

$\Rightarrow$ By canceling $\dfrac{\partial Q\left(\theta; \hat{\theta}\right)}{\partial \theta}$, we get: $\theta^*_{u,v} = \dfrac{\sum\limits_{D \in \mathcal{D}^?_{u,v}} \dfrac{\hat{\theta}_{u,v}}{\hat{P}_{t^D_v}(v)}}{|\mathcal{D}^?_{u,v}| + |\mathcal{D}^-_{u,v}|}$

$$\mathcal{D}^?_{u,v} = \{D \in \mathcal{D} | (u,v) \in D^2 \wedge t^D_v = t^D_u + 1\}$$
$$\mathcal{D}^-_{u,v} = \{D \in \mathcal{D} | u \in D \wedge ((v \notin D) \vee (v \in D \wedge t^D_v > t^D_u + 1))\}$$

Limits of IC

- Closed-world Hypothesis
  - External world can be represented as an additional node [Gruhl et al., 2004]
- Diffused content not taken into account
  - Influence distributions do not depend on what is diffused
- Information on nodes not taken into account
  - User profiles
  - Current user activities
- Time Discretization
  - Diffusion proceeds in steps

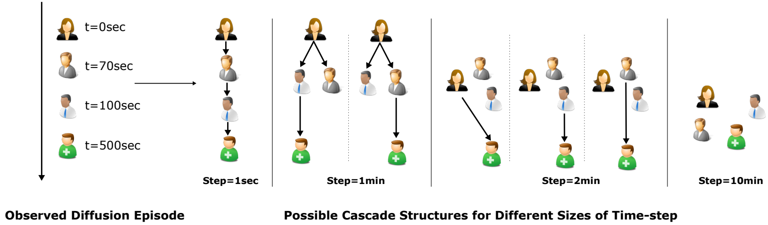
IC : Time Discretization

- IC learning requires to gather events per time period for learning
  - Infection of v by u only possible at step $t_u^D + 1$

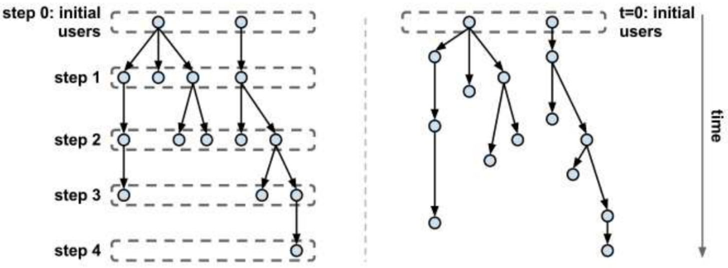
$\Rightarrow$ Too long steps : too many events in the same steps (no possible influence)

$\Rightarrow$ Too short steps : many "holes" in the diffusion process

  - Isolated users with no possible explanation
  - Even if we remove empty steps, very strong assumptions on the diffusion:



Continuous Time Diffusion



Modèle Independent Cascade

Processus de diffusion plus réaliste

- Two main variants of IC propose to consider continuous time delays of infection:
  - NetRate [Gomez-Rodriguez et al., 2011]
  - CTIC [Saito et al., 2009]

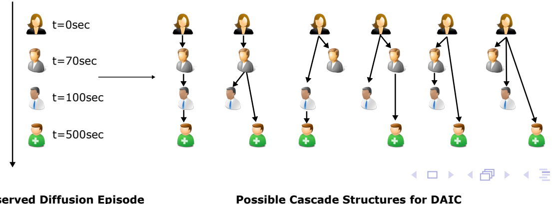
Continuous Time Diffusion

- NetRate [Gomez-Rodriguez et al., 2011]
  - Definition of probability distributions which decrease with time (Exponential, Power, Raighley, etc.)
    - *e.g.*, Exponential Distribution: $f(t_j|t_i;\theta_{i,j}) = \theta_{i,j}\exp^{-\theta_{i,j}(t_j-t_i)}$
  - Only one parameter per link to control:
    - Influence strength
    - Influence delay
  - $+$ Convex optimization problem
  - $-$ Every infection happens, some after a maximal time T
    - The choice of T can be difficult
  - $-$ A slower influence does not necessarily imply a less frequent one
- CTIC [Saito et al., 2009]
  - 2 types of parameters per link
    - Influence probability $k \in]0,1[$
    - Delay parameter $r \in \mathbb{R}^+$
  - Probability density that i infects j at time t_j^D:

    $f(t_j|t_i;k_{i,j};r_{i,j}) = k_{i,j}r_{i,j}\exp^{-r_{i,j}(t_j^D-t_i^D)}$
  - $+$ A more flexible model
  - $-$ But more complex to optimize $\Rightarrow$ EM algorithm

Learning Diffusion Models in Practice

- Continuous Time Diffusion
  - Very effective when infection delay regularities can be observed but...
  - Such regularities are rarely observed from social data
  - ⇒ The variability on delays can strongly limit the ability of extracting influence tendencies
- Relaxation of IC: Delay-Agnostic IC [Lamprier et al., 2015]
  - No time discretization
  - Uniform time delays
  - \+ More flexible than IC ($\times$10 more effective on social data)
  - \+ More realistic than continuous models (performs at least as well as CTIC on social data)
  - \+ Greatly simpler than CTIC
  - – Infection times cannot be predicted



DAIC

Log-likelihood of DAIC:

$$\mathcal{L}(\theta;\mathcal{D}) = \sum_{D\in\mathcal{D}} \left(\sum_{v\in D} \log P_{t_v^D}(v) + \sum_{v\notin D}\sum_{u\in D} \log(1-\theta_{u,v}) \right)$$

with $P_{t_v^D}(v) = 1 - \prod_{u\in Preds(v)\wedge t_u^D < t_v^D} 1-\theta_{u,v}$

Update-rule for DAIC :

$$\theta^*_{u,v} = \frac{\sum_{D\in\mathcal{D}^+_{u,v}} \frac{\hat{\theta}_{u,v}}{\hat{P}_{t_v^D}(v)}}{|\mathcal{D}^+_{u,v}| \ + \ |\mathcal{D}^-_{u,v}|}$$

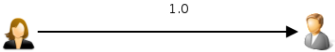
With:

$$\mathcal{D}^+_{u,v} = \{D\in\mathcal{D} | (u,v)\in D^2 \wedge t_v^D > t_u^D\}$$
$$\mathcal{D}^-_{u,v} = \{D\in\mathcal{D} | u\in D \wedge v\notin D\}$$

Avoid overfitting with IC and DAIC

Learning bias of IC (increased with DAIC):

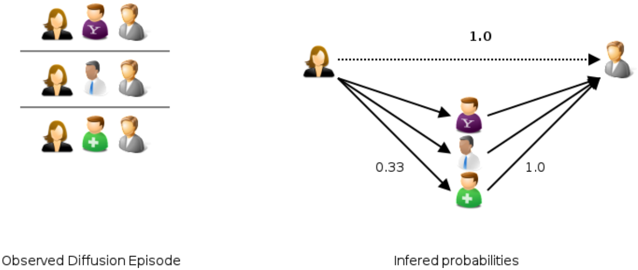
Observed Diffusion Episode



Infered probabilities

Avoid overfitting with IC and DAIC

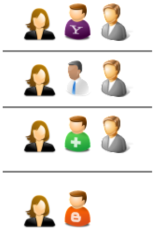
Learning bias of IC (increased with DAIC):



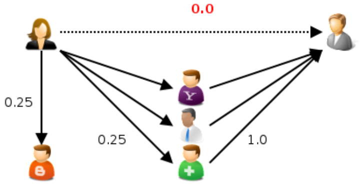
- Rare pairs (u, v) can easily obtain $\theta_{u,v} = 1.0$ ...

Avoid overfitting with IC and DAIC

Learning bias of IC (increased with DAIC):



Observed Diffusion Episode

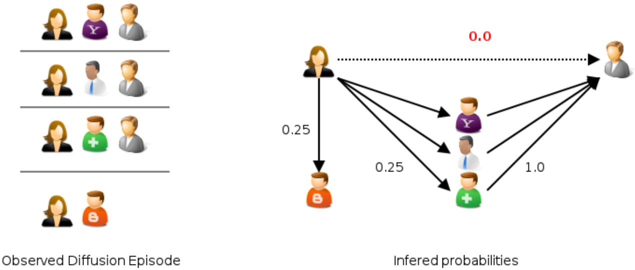


Infered probabilities

- Rare pairs (u, v) can easily obtain $\theta_{u,v} = 1.0$ ...
- .. and can make more frequent pairs (u, v) converge to $\theta_{u,v} = 0.0$

Avoid overfitting with IC and DAIC

Learning bias of IC (increased with DAIC):



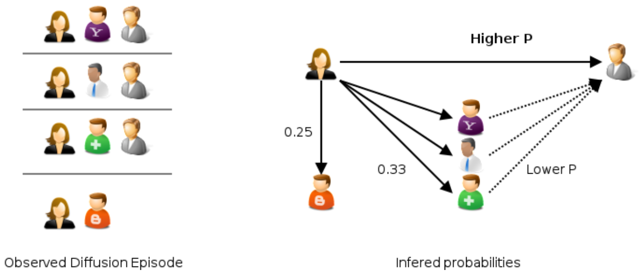
- Rare pairs (u, v) can easily obtain $\theta_{u,v} = 1.0$ ...
- .. and can make more frequent pairs (u, v) converge to $\theta_{u,v} = 0.0$

$\Rightarrow$ Maximum likelihood reached with several parameters set to 1 (overfitting)

$\Rightarrow$ Rare users have a great impact on the extracted relationships

Avoid overfitting with IC and DAIC

Learning bias of IC (increased with DAIC):



- Rare pairs (u, v) can easily obtain $\theta_{u,v} = 1.0$ ...
- .. and can make more frequent pairs (u, v) converge to $\theta_{u,v} = 0.0$

$\Rightarrow$ Maximum likelihood reached with several parameters set to 1 (overfitting)

$\Rightarrow$ Rare users have a great impact on the extracted relationships

DAIC Regularization

- Influence is a rare event
  - Very high probabilities for $\theta_{u,v}$ are unlikely

$\Rightarrow$ Introduction of an exponential prior [Lamprier et al., 2015]:

$$p(\theta) = \prod_{\theta_{u,v}} \lambda e^{-\lambda \theta_{u,v}}$$

- Maximum a Posteriori:

$$\theta^* = \arg\max_{\theta} \mathcal{L}(\theta; \mathcal{D}) - \lambda \sum_{\theta_{u,v}} \theta_{u,v}$$

- Favors sparse influence networks

$\Rightarrow$ Adaptation of the Saito's EM

1. Information Diffusion
2. The Independent Cascade Model
3. Deep-Learning for Diffusion
   - Embedded IC
   - Predictive models
   - Recurrent Neural Networks for Diffusion

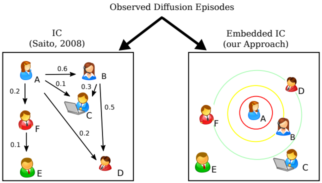
Embedded IC

Representation Learning

Project items in a continuous space in such a way that *relationships* between items are modeled by *distances* (or similarities) between their representations in this space

$\Rightarrow$ Obtain a more compact model

$\Rightarrow$ Infer new relationships

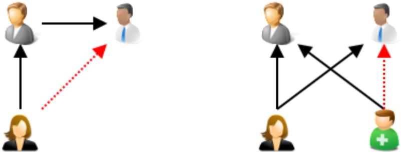


- Each user i is associated to a projection $z_i \in \mathbb{R}^d$
- The transmission probability $\theta_{i,j}$ becomes a function:

$$\theta_{i,j} = f(z_i, z_j)$$

Advantages

- Less parameters, $\mathcal{O}(N)$ rather than $\mathcal{O}(N^2)$
- Inclusion of correlations between links of the network:
  - Transitive relationships (cohesive communities)
  - Similar users tend to impact the same other users (bimodal communities)
  - $\rightarrow$ Naturally modeled by the use of a representation space



Algorithm

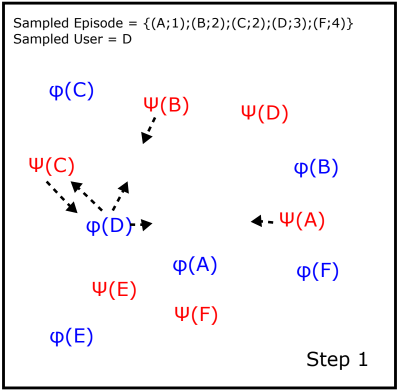
- Influence is an asymmetric relationship
  - Each user i is associated to *two* projections z_i (transmitter projection) and ω_i (receptor projection) in $\mathbb{R}^d$
  - The transmission probability $\theta_{i,j}$ becomes a function:

$$\theta_{i,j} = f(z_i, \omega_j) = \frac{1}{1 + \exp\left(z_i^{(0)} + \omega_j^{(0)} + ||z_i^{(1..d)} - \omega_j^{(1..d)}||^2\right)}$$

- Inter-dependent probability values: no analytic solution for the maximization step

$\rightarrow$ GEM: the maximization is replaced by a step of stochastic gradient ascent

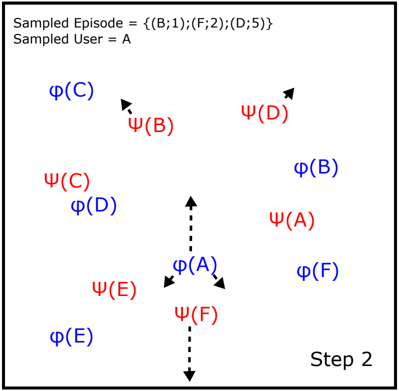
Embedded IC: Example



Iteration 1 :

- Episode : $\{(A, 1); (B, 2); (C, 2); (D, 3); (E, 3); (F, 4)\}$
- User : D (infected)
- Infected predecessors: $\{A, B, C\}$

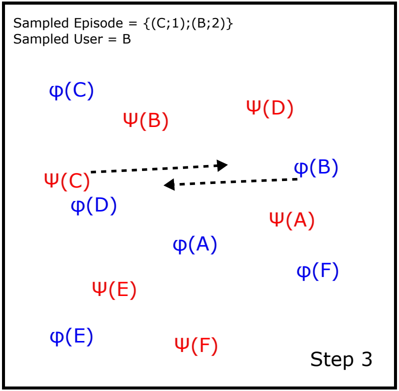
Embedded IC: Example



Iteration 2 :

- Episode : $\{(B, 1); (F, 2); (D, 5)\}$
- User : A (non infected)
- Infected predecessors : $\{B, F, D\}$

Embedded IC: Example

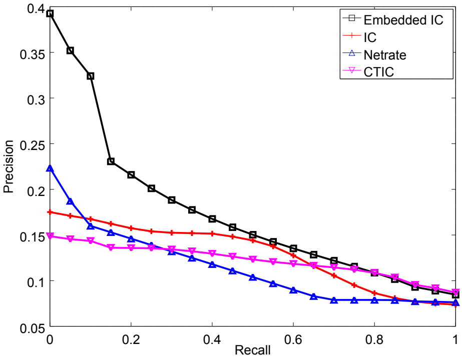


Iteration 3 :

- Episode : $\{(C, 1); (B, 2); \}$
- User : B (infected)
- Infected predecessors : $\{C\}$

Influence links detection

- On the Memetracker corpus:
  - Evaluation on the ability for assigning high transmission probabilities to known relationships
  - Ranking of the links (u_i, u_j) according to $f(z_i, \omega_j)$
  - Precision-Recall curves:



Limits of iterative approaches

- Iterative models effective to describe diffusion processes but...
- ... Low robustness w.r.t. network evolutions [Najar et al., 2012]
- ... Hard to learn for large networks
- ... Over-fitting risks
- ... Complex estimations of infection probabilities (Monte-Carlo simulations [Bòta et al., 2013] or Diffusion kernels [Rosenfeld et al., 2016])

$\Rightarrow$ Non-iterative approaches for diffusion prediction

- Focus on mapping final states from inital ones.
- $C^D = f_\theta(S^D)$, with S^D and C^D respectively the source and final contamination states for the episode D

A first non-iterative approach

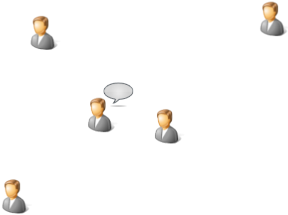
- Discriminative Model from a set of sources [Najar et al., 2012]
    - Input = Binary vector $S^D \in \{0;1\}^{|U|}$, with $S_i^D = 1$ if i is in the sources of D
    - Output = Binary vector $C^D \in \{0;1\}^{|U|}$, with $C_i^D = 1$ if i is in the final infection of D
- Logistic Regression

$$\theta^* = \arg\max_{\theta} \sum_{D \in \mathcal{D}} \sum_{i \in U} C_i^D \log\left(\frac{1}{1+e^{-f_\theta(i,S^D)}}\right) + (1 - C_i^D)\log\left(1 - \frac{1}{1+e^{-f_\theta(i,S^D)}}\right)$$

- Various possible functions f (dot product, neural network, etc...)

a Non-iterative Approach with Representation Learning

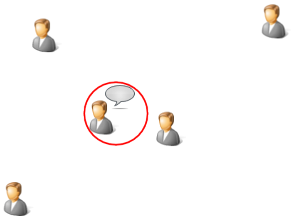
- Projection in a continuous latent space [Bourigault et al., 2014]
  - Diffusion modeled as a heat diffusion process in the space



- The temperature $T(u_i, t)$ of u_i at time t renders its propensity of infection
- The heat starts from the source

a Non-iterative Approach with Representation Learning

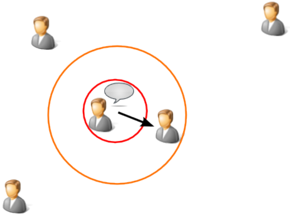
- Projection in a continuous latent space [Bourigault et al., 2014]
  - Diffusion modeled as a heat diffusion process in the space



- The temperature $T(u_i, t)$ of u_i at time t renders its propensity of infection
- The heat starts from the source

a Non-iterative Approach with Representation Learning

- Projection in a continuous latent space [Bourigault et al., 2014]
  - Diffusion modeled as a heat diffusion process in the space



Heat equation :

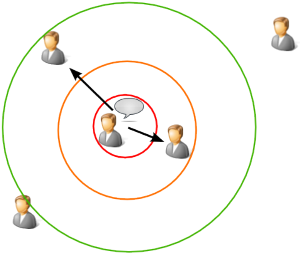
$$\begin{cases} \frac{\partial T}{\partial t} = \Delta_x T \\ f(x,0) = f_0(x) \end{cases}$$

Solution when the source is at x_0 :

$$T_{x_0}(x,t) = (4\pi t)^{-\frac{n}{2}} e^{-\frac{||x_0 - x||^2}{4t}}$$

a Non-iterative Approach with Representation Learning

- Projection in a continuous latent space [Bourigault et al., 2014]
  - Diffusion modeled as a heat diffusion process in the space



Heat equation :

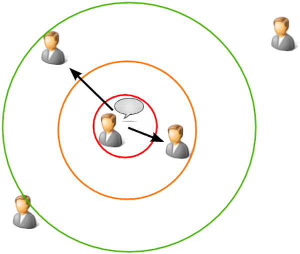
$$\begin{cases} \frac{\partial T}{\partial t} = \Delta_x T \\ f(x,0) = f_0(x) \end{cases}$$

Solution when the source is at x_0 :

$$T_{x_0}(x,t) = (4\pi t)^{-\frac{n}{2}} e^{-\frac{||x_0 - x||^2}{4t}}$$

a Non-iterative Approach with Representation Learning

- Projection in a continuous latent space [Bourigault et al., 2014]
  - Diffusion modeled as a heat diffusion process in the space



$\rightarrow$ Find a representation $\mathcal{Z}$ of the users such that observed diffusion can be explained as a heat kernel starting from the source

$$\mathcal{U} = (u_1, ..., u_N) \rightarrow \mathcal{Z} = (z_1, ..., z_N) \subset \mathbb{R}^D$$

Diffusion as a heat process

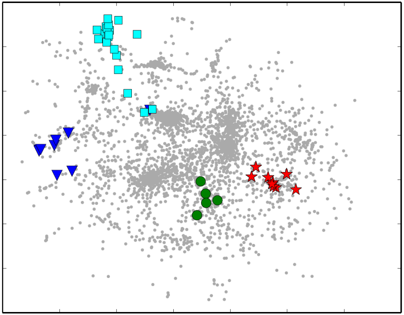
- Projection in a contiunous latent space [Bourigault et al., 2014]
  - For an episode D whose source is s^D:
    - $\forall (u,v), t_u^D < t_v^D \Rightarrow \forall t \; T_{s^D}(u,t) > T_{s^D}(v,t)$
  - In the latent space $\rightarrow$ geometric constraint:
    - $\forall (u,v), t_u^D < t_v^D \Rightarrow ||z_{s^D} - z_u|| < ||z_{s^D} - z_v||$

$\Rightarrow$ Loss Function:

$$\Delta_{rank}(\mathcal{Z}, \mathcal{D}) = \sum_{D \in \mathcal{D}} \sum_{\substack{u,v \\ t^{C}(u) < t^{C}(v)}} max(0, 1 - (||z_{s^D} - z_v||^2 - ||z_{s^D} - z_u||^2)) + \sum_{u,v \in D \times \bar{D}} max(0, 1 - (||z_{s^D} - z_v||^2 - ||z_{s^D} - z_u||^2))$$

Optimization by stochastic gradient descent on $\mathcal{Z}$.

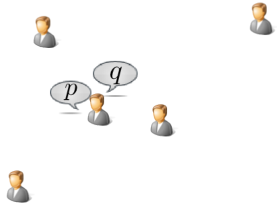
- Projection of users of **Digg** in 2 dimensions :



+ Capture of regularities between extracted relationships ⇒ better generalization
+ Possibility to include the propagated content

Heat Diffusion: Inclusion of the Content

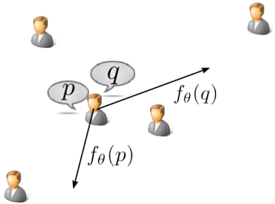
- Hypothesis : The diffused content impacts the diffusion dynamics
  - Translation of the source according to the content



- Similar learning:
  - Simultaneous learning of the translation function f_θ and the projections $\mathcal{Z}$
  - Optimization by stochastic gradient ascent

Heat Diffusion: Inclusion of the Content

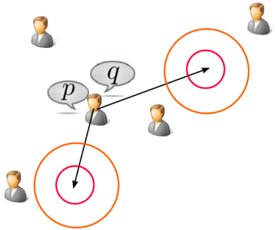
- Hypothesis : The diffused content impacts the diffusion dynamics
  - Translation of the source according to the content



- Similar learning:
  - Simultaneous learning of the translation function f_θ and the projections $\mathcal{Z}$
  - Optimization by stochastic gradient ascent

Heat Diffusion: Inclusion of the Content

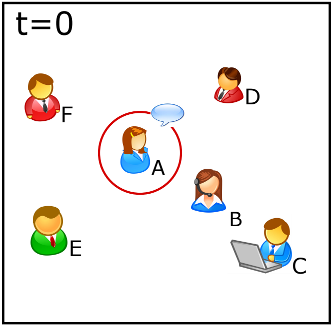
- Hypothesis : The diffused content impacts the diffusion dynamics
  - Translation of the source according to the content



- Similar learning:
  - Simultaneous learning of the translation function f_θ and the projections $\mathcal{Z}$
  - Optimization by stochastic gradient ascent

Heat diffusion: iterative version

Chain reaction : Each infected user starts emitting heat from its infection time



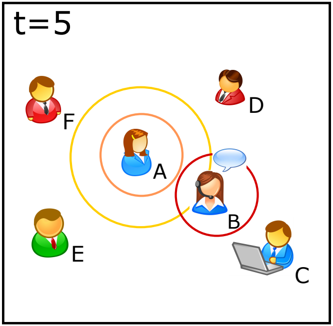
⇒ Dynamic model

⇒ Infection time prediction

⇒ Dealing with multiple sources

Heat diffusion: iterative version

Chain reaction : Each infected user starts emitting heat from its infection time



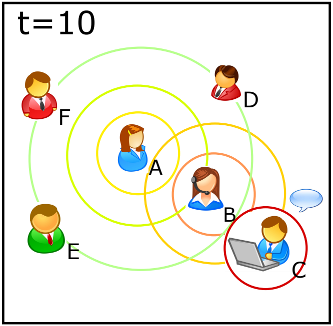
⇒ Dynamic model

⇒ Infection time prediction

⇒ Dealing with multiple sources

Heat diffusion: iterative version

Chain reaction : Each infected user starts emitting heat from its infection time



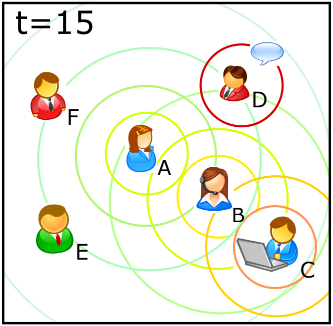
⇒ Dynamic model

⇒ Infection time prediction

⇒ Dealing with multiple sources

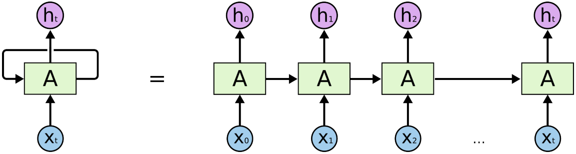
Heat diffusion: iterative version

Chain reaction : Each infected user starts emitting heat from its infection time



- ⇒ Dynamic model
- ⇒ Infection time prediction
- ⇒ Dealing with multiple sources

Recurrent Neural Networks for Diffusion

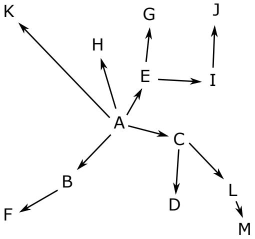


- Diffusion episodes could be seen as sequences
  - Recurrent Neural Networks are well suited for dealing with sequences

  $\Rightarrow$ RNN could be used to consider the history of events to predict the future events

- However, direct application of RNN does not well perform for diffusion

⇒ Not sequences but trees



- Should the prediction of J be impacted by the previous observation of D ?
  - ⇒ Cross-dependence of infections

Recurrent Neural Networks for Diffusion

- RNN based on Attention Learning
  - DeepCas [Li et al., 2017]

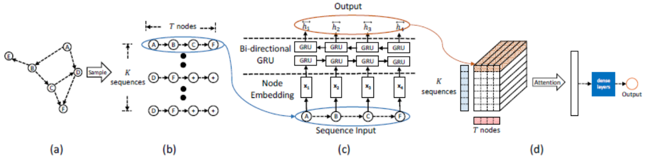
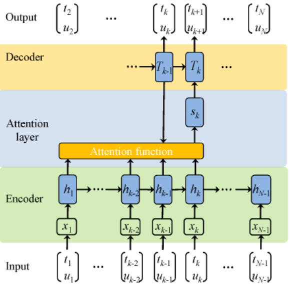


Figure 1: The end-to-end pipeline of DeepCas.

  - CYAN-RNN [Wang et al., 2017]



⇒ RNN IC with MCMC / Variational inference

References

- [Anshelevich et al., 2009] Elliot Anshelevich, Deeparnab Chakrabarty, Ameya Hate, and Chaitanya Swamy. 2009. Approximation Algorithms for the Firefighter Problem: Cuts over Time and Submodularity. ISAAC 2009: 974-983
- [Bourigault et al., 2014] Simon Bourigault, Cédric Lagnier, Sylvain Lamprier, Ludovic Denoyer, Patrick Gallinari: Apprentissage de représentation pour la diffusion d'Information dans les réseaux sociaux. CORIA-CIFED 2014: 155-170
- [Chen et al., 2013] Chen, G. H., Nikolov, S., and Shah, D. (2013). A latent source model for nonparametric time series classification. NIPS 2013: 1088–1096
- [Dempster et al., 1977] A.P. Dempster, N.M. Laird et Donald Rubin, " Maximum Likelihood from Incomplete Data via the EM Algorithm ", Journal of the Royal Statistical Society. Series B (Methodological), vol. 39, no 1, 1977, p. 1–38
- [Kempe et al., 2003] D. Kempe, J. Kleinberg, and E. Tardos, Maximizing the spread of influence in a social network, In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge
- [Gomez Rodriguez et al., 2010] Manuel Gomez Rodriguez, Jure Leskovec, and Andreas Krause. 2010. Inferring networks of diffusion and influence. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '10). ACM, New York, NY, USA, 1019–1028.
- [Gomez-Rodriguez et al., 2011] M. Gomez-Rodriguez, D. Balduzzi, and B. Schölkopf. Uncovering the temporal dynamics of diffusion networks. In Proceedings of the 28th International Conference on Machine Learning (ICML-11), ICML'11, pages 56–568. ACM, 2011.
- [Granovetter, 1973] Granovetter, M.S. The Strength of Weak Ties. The American Journal of Sociology 78 (6): 1360–1380. Discovery and Data Mining (KDD), 2003.
- [Guille et al., 2012] Adrien Guille, Hakim Hacid: A predictive model for the temporal dynamics of information diffusion in online social networks. WWW (Companion Volume) 2012: 1145-1152
- [Gruhl et al.] D. Gruhl , R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In WWW'04, pages 491–501, New York, NY, USA, 2004. ACM.
- [Lagnier et al., 2013] Cédric Lagnier, Ludovic Denoyer, Éric Gaussier, Patrick Gallinari: Predicting Information Diffusion in Social Networks Using Content and User's Profiles. ECIR 2013: 74-85

References

- [Lamprier et al., 2015] Sylvain Lamprier, Simon Bourigault, Patrick Gallinari: Extracting Diffusion Channels from Real-World Social Data: a Delay-Agnostic Learning of Transmission Probabilities. ASONAM 2015
- [Li et al., 2017] Cheng Li, Jiaqi Ma, Xiaoxiao Guo, Qiaozhu Mei: DeepCas: An End-to-end Predictor of Information Cascades. WWW 2017
- [Najar et al., 2012] Anis Najar, Ludovic Denoyer, Patrick Gallinari: Predicting information diffusion on social networks with partial knowledge. WWW (Companion Volume) 2012: 1197-1204
- [Saito et al., 2008] K. Saito, R. Nakano, and M. Kimura. Prediction of information diffusion probabilities for independent cascade model. In Proceedings of the 12th international conference on Knowledge-Based Intelligent Information and Engineering Systems, Part III, KES '08, pages 67–75. Springer-Verlag, 2008.
- [Saito et al., 2009] K. Saito, M. Kimura, K. Ohara, and H. Motoda. Learning continuous-time information diffusion model for social behavioral data analysis. In Proceedings of the 1st Asian Conference on Machine Learning: Advances in Machine Learning, ACML '09, pages 322–337, Berlin, Heidelberg, 2009. Springer-Verlag.
- [Saito et al., 2011] Kazumi Saito, Kouzou Ohara, Yuki Yamagishi, Masahiro Kimura, and Hiroshi Motoda. Learning diffusion probability based on node attributes in social networks. In Marzena Kryszkiewicz, Henryk Rybinski, Andrzej Skowron, and Zbigniew W. Ras, editors, ISMIS, volume 6804 of Lecture Notes in Computer Science, pages 153–162. Springer, 2011.
- [Shah and Zaman, 2010] Shah, D. and Zaman, T. (2010). Detecting sources of computer viruses in networks : Theory and experiment. SIGMETRICS 2010: 203–214
- [Tsur and Rappoport, 2012] Tsur, O. and Rappoport, A. What's in a hashtag ? : content based prediction of the spread of ideas in microblogging communities. WSDM 2012: 643–652
- [Ver Steeg et al., 2013] G. Ver Steeg and A. Galstyan. Information-theoretic measures of influence based on content dynamics. In Proceedings of the sixth ACM international conference on Web search and data mining, WSDM '13, pages 3–12, New York, NY, USA, 2013. ACM.
- [Wang et al., 2017] Yongqing Wang, Huawei Shen, Shenghua Liu, Jinhua Gao, Xueqi Cheng : Cascade Dynamics Modeling with Attention-based Recurrent Neural Network. IJCAI 2017.