A Predictive Model for the Temporal Dynamics of Information Diffusion in Online Social Networks

Adrien Guille*, Hakim Hacid**
*ERIC Lab, Université Lumière Lyon 2
**Alcatel-Lucent Bell Labs France
Online social networks are very powerful tools for the spread of information.

- 2010 Arab spring (Howard et al., 2011)
- 2008 U.S. presidential elections (Hughes et al., 2009)
- etc.

The case of Twitter

- Explicit network
- Following & mentioning ties
A model capable of predicting the temporal dynamics

Input: a social network, an information topic
Output: the time-serie of the volume of tweets generated by the diffusion

Purpose of our proposal
### Predictive models for diffusion on Twitter

<table>
<thead>
<tr>
<th></th>
<th>basis network</th>
<th>prediction nature</th>
<th>object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bakshy et al., 2011</td>
<td>passive links</td>
<td>topological</td>
<td>URL</td>
</tr>
<tr>
<td>Yang &amp; Counts, 2010</td>
<td>active links</td>
<td>topological</td>
<td>topic</td>
</tr>
<tr>
<td>Galuba et al., 2010</td>
<td>passive &amp; active links</td>
<td>topological</td>
<td>URL</td>
</tr>
<tr>
<td>Yang &amp; Leskovec, 2010</td>
<td>non-graphical</td>
<td>temporal</td>
<td>hashtag</td>
</tr>
<tr>
<td><strong>Our approach</strong></td>
<td>passive &amp; active links</td>
<td>temporal</td>
<td>topic</td>
</tr>
</tbody>
</table>
• Raw data
  - Yang & Leskovec: $4.76 \cdot 10^8$ tweets
  - Kwak et al.: followers graph
    $(1.47 \cdot 10^9$ edges)

• Preprocessing
  - Extraction of the active ties
  - Identification of sub-networks

• Identification of cascades
• AsIC principle (Saito et al., 2010)
  ○ A network: the graph of followers
  ○ Diffusion probabilities & time-delays

• Asynchronous Independent Cascades

Basis of the proposed model
• Assumption
  ○ The dynamics of the spreading process at the **macroscopic** level is explained by interactions at a **microscopic** level between pairs of users and the topology of their interconnections.

• Dimensions
<table>
<thead>
<tr>
<th>user</th>
<th>I</th>
<th>H</th>
<th>dTR</th>
<th>hM</th>
<th>mR</th>
<th>hK</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>u1</td>
<td>0.78</td>
<td>0.12</td>
<td>0.7</td>
<td>0</td>
<td>0.65</td>
<td>1</td>
<td>0.125</td>
</tr>
<tr>
<td>u2</td>
<td>0.23</td>
<td>0.12</td>
<td>0.23</td>
<td>1</td>
<td>0.10</td>
<td>1</td>
<td>0.33</td>
</tr>
</tbody>
</table>
  
  = 13 attributes for each set (u1, u2, topic, time)

Our approach
Supervised classification task
Binary class: diffusion/non-diffusion
  - Bayesian Logistic Regression (BLR)

Results
  - Social dimension only
    precision rate of 79%
  - Social + temporal + semantic dimensions
    gain of 7%

Global precision: 85%

Estimating diffusion probabilities
• Input
  ◦ a network of users $U = \{u_1, u_2, \ldots, u_n\}$
  ◦ an information $i = \{k_1, \ldots, k_p\}$
  ◦ a subset of initially informed users

• extension of AsIC
  ◦ $f_{u_1, u_2}(i, t)$ based on BLR
  ◦ $r_{u_1, u_2} = (1 - l(u_2)) \cdot 10$

• $t$ simulates the course of each day

Building a prediction engine
• Visualization

Information about the release date of the new iPhone

Time-serie: • predicted • real

Evaluation on a prediction task
• **Gain of our approach over the 1-time lag predictor** (Yang & Leskovec, 2010)

• **Two aspects:**
  - Relative error on predicting **volume**
  - Relative error on **dynamics**

<table>
<thead>
<tr>
<th></th>
<th>E1</th>
<th>E2</th>
<th>E3</th>
<th>E4</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain over dynamics</td>
<td>25.19 %</td>
<td>39.23 %</td>
<td>29.21 %</td>
<td>3.22 %</td>
<td>24.21 %</td>
</tr>
<tr>
<td>Gain over volume</td>
<td>42.89 %</td>
<td>47.70 %</td>
<td>34.49 %</td>
<td>40.07 %</td>
<td>41.29 %</td>
</tr>
<tr>
<td>Overall gain</td>
<td>34.04 %</td>
<td>43.46 %</td>
<td>31.85 %</td>
<td>21.65 %</td>
<td>32.75 %</td>
</tr>
</tbody>
</table>
• Initial assumption confirmed by the experiments i.e. predicting diffusion from local properties

• Observation of a particular pattern

• Improve volume estimation

**Conclusion**

**Time-series:**
- adjusted prediction (c=1.6)
- predicted
- real
• P. N. Howard, A. Duffy, D. Freelon, M. Hussain, W. Marai, and M. Mazaid. Opening closed regimes, what was the role of social media during the arab spring?, 2011.


• J. Yang and S. Counts. Predicting the speed, scale, and range of information diffusion in twitter. ICWSM’10, 2010.


• Visualization

Information about the launch of Googlewave

Time-serie:  predicted  real

Annexes