Using Crowdsourced Data in Location-based Social Networks to Explore Influence Maximization

Ji Li\textsuperscript{1} Zhipeng Cai\textsuperscript{1} Mingyuan Yan\textsuperscript{2} Yingshu Li\textsuperscript{1}

\textsuperscript{1}Department of Computer Science, Georgia State University

\textsuperscript{2}Department of Computer Science and Information Systems, University of North Georgia

IEEE INFOCOM 2016
Online social networks become a hot research topic

One of the most important research issues is the influence maximization problem

Many existing works: only the influence propagation in the online social network is considered

The influence can also propagate in the physical world
Introduction

Contribution

- Propose a network model and an influence propagation model

- Influence propagation in
  - Online social network
  - Physical world

- Propose an event activation position selection problem

- Designed a heuristic algorithm for position selection problem
Network Model

Two-layer graph \( G^t = (V, E_f, E_p^t) \)
- **\( V \):** set of \( n \) users deployed in \([0, 1)^2\)
- **\( E_f \):** friendship in the online social network, remains the same
- **\( E_p^t \):** neighbour in the physical world at time \( t \), changes over time

Influence propagation
- An event is activated in the physical world, users around the activation position may be influenced
- The influence propagates in both online social networks and the physical world simultaneously
Network Model

- $u_5 \notin IR^1$, $u_5$ may still be influenced since $u_3$, $u_6$ are $u_5$’s friends and $u_3$, $u_6 \in IR$
- $(u_1, u_7) \notin E_f$, influence may propagate $u_1 \rightarrow u_7$ since $(u_1, u_7) \in E^t_p$

\(^1 IR: \) influencing region
Measurements

Datasets
- Two actual datasets named Brightkite and Gowalla
- Dataset origin: Stanford Large Network Dataset Collection [1]

Findings
- Users’ positions in the physical world have high stability
- Influence of different users in an online social network varies a lot
- Convenient to propagate influence in the physical world
- High interdependency of geographical positions for friends in the online social networks
Influence Propagation Model

Basic Influence Propagation Model

- **Initial propagation period**: users around the activation position may be influenced
- **Additional propagation period**: influence propagates in online social networks and the physical world

\[ E.t_0 \quad E.t_0 + E.t_{init-pro} \quad E.t_0 + E.t_{init-pro} + E.t_{add-pro} \]
In initial propagation period, users in the influencing region may be influenced

Influencing probability: $p(E, v, init_{inf})$
A influenced user $v$ shares the event $E$ with its friends with probability $p(v, osn\_share)$.

$v$’s friend $u$ be influenced with probability $p(E, u, osn\_inf)$.
Influenced user $v$ propagates $E$ with probability $p(v, pw\_share)$

$v$ will choose neighbours to share event $E$

Neighbour $u$ be influenced with probability $p(E, u, phy\_inf)$
Influence Propagation Model

Cross Propagation

Online social networks → physical world
- User $u$ Online Social Network user $v$
- User $v$ Physical World user $w$

Physical world → online social networks
- User $u$ Physical World user $v$
- User $v$ Online Social Network user $w$
Problem Description

- A candidate position set $\mathcal{C}$
- Select $pos \in \mathcal{C}$ to activate event $E$ so that its influence can be maximized
- Other parameters of $E$ are fixed

Formalized Definition

- $F(pos)$: number of influenced users if activate event $E$ in $pos$
- Input: candidate position set $\mathcal{C}$, event $E$, graph $G^t = (V, E_f, E_p^t)$
- Output: $\arg \max_{pos \in \mathcal{C}} F(pos)$
Optimal Event Activation Position Selection
Heuristic Algorithm

- $N_s$ and $N_i$: predefined constant
- $P = \text{randomly selected } N_s \text{ positions in } C$
- Iterate for $N_i$ times
  - Replace $pos_1 \in P$ with a new position $pos_2$ if $F(pos_2) > F(pos_1)$
- Return $\arg\max_{pos \in P} F(pos)$
- $F'$-based algorithm

Details

Algorithm code  New position selection  Why not other algorithms?
Figure 1: Comparison of different propagation models
Experimental Results

Distribution of Influenced Users

Figure 2: Comparison of different influence manners
Experimental Results
The Optimal Activation Position Selection Algorithm

Figure 3: Comparison between $F$-based and $F'$-based heuristic algorithms
Propose a new network model

A new event influence propagation model is proposed based on the measurement results of two actual datasets

An event activation position selection problem is defined

A heuristic algorithm for the position selection problem is designed
Questions and Answers
Thank you very much
Appendix: Measurements

Introduction to the Datasets

Basics

- Record formatte: \((user-id, login-time, latitude, longitude, location-id)\)
- Use the random way point model to estimate users’ positions [7]

Investigated login records period

- Brightkite: April 2008 through October 2010
- Gowalla: December 2009 through October 2010
Appendix: Measurements

Introduction to the Datasets

**Standardization**
- Users’ positions in the physical world $\rightarrow [0, 1)^2$
- Time stamps for the logins $\rightarrow [0, 1)$

**Data selection**
- To make sure the users’ movements are correctly detected, we employ part of the original Brightkite and Gowalla datasets
- The users are distributed in $400 km \times 400 km$ rectangle regions
- These regions include New York, Washington and Philadelphia where users are densely distributed
## Appendix: Measurements

### Introduction to the Datasets

<table>
<thead>
<tr>
<th>property</th>
<th>Brightkite</th>
<th>Gowalla</th>
</tr>
</thead>
<tbody>
<tr>
<td># users</td>
<td>3551</td>
<td>5231</td>
</tr>
<tr>
<td># edges</td>
<td>9317</td>
<td>10134</td>
</tr>
<tr>
<td>average degree</td>
<td>5.248</td>
<td>3.875</td>
</tr>
<tr>
<td># CC</td>
<td>569</td>
<td>1778</td>
</tr>
<tr>
<td># nodes in largest CC</td>
<td>2907</td>
<td>3114</td>
</tr>
<tr>
<td># logins</td>
<td>430657</td>
<td>297104</td>
</tr>
<tr>
<td>average login</td>
<td>121.278</td>
<td>56.797</td>
</tr>
<tr>
<td># triangles</td>
<td>6738</td>
<td>11580</td>
</tr>
<tr>
<td>average CC size</td>
<td>6.241</td>
<td>2.942</td>
</tr>
<tr>
<td># edges in largest CC</td>
<td>9228</td>
<td>9676</td>
</tr>
</tbody>
</table>

**Table 1: Dataset Details**
Appendix: Measurements

Users’ Positions and Number of Friends

Figure 4: The distribution of distances
Appendix: Measurements

Users’ Positions and Number of Friends

![Graph of the distribution of numbers of friends for Brightkite and Gowalla.](image)

(a) Brightkite  (b) Gowalla

**Figure 5:** The distribution of numbers of friends
Appendix: Measurements

Users’ Positions and Number of Friends

Figure 6: The distribution of number of neighbors

(a) Brightkite

(b) Gowalla
Appendix: Measurements

Positions and Friendships

Figure 7: The percentage of users within a given distance

(a) Brightkite

(b) Gowalla
Figure 8: The distribution of minimum distances
Appendix: Measurements

Positions and Friendships

Figure 9: The distribution of trajectory similarities

(a) Brightkite

(b) Gowalla
Appendix: Influence Propagation Model

Initial Influence Propagation

\[ p(E, v, \text{init\_inf}) = \min (p_1 I_1(E, v) I_2(E, v), 1) \]

\[ I_1(E, v) = I(J(E.\text{type}, v.\text{interest}), I_{max1}) \]

\[ I(x, I_{max}) = (I_{max} - 1) \sqrt{1 - (1 - x)^2} + 1 \]

\[ J(E.\text{type}, v.\text{interest}) = \frac{|E.\text{type} \cap v.\text{interest}|}{|E.\text{type} \cup v.\text{interest}|} \]

\[ I_2(E, v) = I(T(E, v), I_{max2}) \]

- \( p_1 \): base influence probability
- \( I_{max1} \) & \( I_{max2} \): upper bound of increase
- \( T(E, v) \): time \( v \) stays in the influencing region
Appendix: Influence Propagation Model

Influence Propagation in Online Social Networks

\[ p(v, osn\_share) = \min(p_2 I_1(E, v), 1) \]

\[ p(E, u, osn\_inf) = \min(p_3 I_1(E, u)I_3(u, t), 1) \]

\[ I_3(u, t) = I(\min(\frac{n_r(E, u, t) - 1}{n_{max}}, 1), I_{max3}) \]

- \( p_2 / p_3 \): base sharing/influencing probability
- \( I_{max3} \): upper bound of increase
- \( n_r(E, u, t) \): the number of descriptions of event \( E \) received by user \( u \) at time \( t \)
- \( n_{max} \): predefined constant
Appendix: Influence Propagation Model

Influence Propagation in the Physical World

\[
p(v, pw\_share) = \min(p_4 I_1(E, v), 1)
\]

\[
p(E, u, phy\_inf) = \min(p_5 I_1(E, u) I_4(v, u), 1)
\]

\[
I_4(v, u) = \begin{cases} 
    c & \text{if } (v, u) \in E_f \\
    1 & \text{otherwise}
\end{cases}
\]

- \(p_4 / p_5\): base sharing/influencing probability
- \(v\_share\_phy\): set of different time instances for \(v\) to share \(E\) in the physical world
- \(c\): predefined constant which satisfies \(c > 1\)
- The roulette wheel method[8] is used to decide which neighbour to be chosen and \(I_4(v, u)\) is used to calculate the weight.
Appendix: Optimal Event Activation Position Selection

**Option 1**
- Idea: select the position with most users
- Problem: little interdependency between
  - The number of influenced users
  - The number of users in the initial influencing region

**Option 2**
- Idea: existing algorithm + influence propagation in the physical world
- Problem: distributions of top influencing users may be dispersive
Option 3

- **Idea:** use the monotonicity of function $F(pos)$
- **Problem:** values of function $F(pos)$ may distribute highly irregularly

Option 4

- **Idea:** test each position in $C$
- **Problem:** does not work if $C$ is an infinite set
Figure 10: The number of influenced users vs. the number of users in the initial influencing region.
Appendix: Optimal Event Activation Position Selection

Figure 11: User distribution

(a) Brightkite

(b) Gowalla
Appendix: Optimal Event Activation Position Selection

Figure 12: Function $F(pos)$’s values

(a) Brightkite

(b) Gowalla
Appendix: Optimal Event Activation Position Selection

- $N_c[pos_1]$ records how many times $pos_1$ is selected

- $\Delta_{init}$ and $\alpha$ ($\alpha > 1$) are predefined constants

- Calculate the upper bound of $dis(pos_1, pos_2)$ by $\Delta = \frac{\Delta_{init}}{\alpha^{N_c[pos_1]}}$

- Randomly select $pos_2 \in \{pos \in C \mid dis(pos, pos_1) \leq \Delta\}$
Optimal Event Activation Position Selection

Heuristic Algorithm

- Problem: high computation cost
- Reason: complexity of the influence propagation model
- Solution: use another objective function $F'$
  - $F'$-based heuristic algorithm: use $F'$ as the objective to evaluate the event activation position
Appendix: Optimal Event Activation Position Selection

\[ F'(\text{pos}) = \sum_{u \in U} |u.\text{friends}| + \frac{\sum_{v \in \text{Neg}(u, t)} |v.\text{friends}|}{|\text{Neg}(u, t)|} \]

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(t)</td>
<td>a randomly selected time instance in ([E.t_0 + E.t_{\text{init}<em>\text{pro}}, \ E.t_0 + E.t</em>{\text{init}<em>\text{pro}} + E.t</em>{\text{add}_\text{pro}}])</td>
</tr>
<tr>
<td>(v.\text{friends})</td>
<td>the set of user (v)'s friends in the online social network</td>
</tr>
<tr>
<td>(\text{pos.x} / \text{pos.y})</td>
<td>the (x)-coordinate / (y)-coordinate of position (\text{pos})</td>
</tr>
<tr>
<td>(\text{Neg}(u, t))</td>
<td>(u)'s neighbors in the physical world at time (t)</td>
</tr>
<tr>
<td>(U)</td>
<td>Users in the influencing region</td>
</tr>
</tbody>
</table>

Table 2: Symbols in Function \(F'\)
Appendix: Optimal Event Activation Position Selection

**Algorithm 1 Optimal Activation Position Selection Algorithm**

**Input:** candidate position set $C$, event $E$, graph $G^t = (V, E_f, E_p^t)$

**Output:** position to activate event $E$

```
for $i = 1$ to $N_s$ do
    $P[i] =$ randomly selected element in $C$, $N_c[i] = 0$;
end for

for $i = 1$ to $N_i$ do
    find the minimum $j$ satisfying $rac{\sum_{k=1}^{j} F(P[k])}{\sum_{k=1}^{N_s} F(P[k])} > \text{rand}$;
    randomly select $pos' \in \{pos \in C | \text{dis}(pos, P[j]) \leq \frac{\Delta_{init}}{\alpha N_c[j]} \}$;
    replace $P[j]$ with $pos'$ if $F(pos') > F(P[j])$;
    $N_c[j]++$;
end for

return $\arg \max_{pos \in P} F(pos)$;
```
Appendix: Experimental Results

<table>
<thead>
<tr>
<th>parameter</th>
<th>value</th>
<th>parameter</th>
<th>value</th>
<th>parameter</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E.r_0$</td>
<td>0.01</td>
<td>$E.t_0$</td>
<td>0.5</td>
<td>$E.t_{init \cdot pro}$</td>
<td>0.02</td>
</tr>
<tr>
<td>$E.t_{add \cdot pro}$</td>
<td>0.2</td>
<td>$I_{max_1}$</td>
<td>3</td>
<td>$I_{max_2}$</td>
<td>1.5</td>
</tr>
<tr>
<td>$I_{max_3}$</td>
<td>6</td>
<td>$r_p$</td>
<td>0.01</td>
<td>$c$</td>
<td>5</td>
</tr>
<tr>
<td>$n_{max}$</td>
<td>10</td>
<td>$\alpha$</td>
<td>2</td>
<td>$\Delta$</td>
<td>0.1</td>
</tr>
<tr>
<td>$N_s$</td>
<td>10</td>
<td>$N_i$</td>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Parameters for the Experiments

<table>
<thead>
<tr>
<th>dataset</th>
<th>candidate position set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brightkite</td>
<td>{(x, y) \mid 0.5 \leq x \leq 0.6, 0.45 \leq y \leq 0.55}</td>
</tr>
<tr>
<td>Gowalla</td>
<td>{(x, y) \mid 0.7 \leq x \leq 0.8, 0.6 \leq y \leq 0.7}</td>
</tr>
</tbody>
</table>

Table 4: Candidate Position Set
Appendix: Experimental Results

Number of Influenced Users

Figure 13: The number of influenced users for different initial propagation time
Figure 14: The number of influenced users for different additional propagation time
Figure 15: The number of influenced users for different initial influence radius

[4] uses the Brightkite and Gowalla datasets to study friendships in online social networks and users’ movements in the physical world.

[5] proposes a friendship prediction approach by fusing the topology and geographical features in LBSNs.

[6] studies the impact of social relations hidden in LBSNs.
Appendix: References I


