Hiding the Rumor Source in Anonymous Messaging

Sewoong Oh
University of Illinois at Urbana-Champaign

Giulia Fanti

Peter Kairouz

K. Ramchandran

P. Viswanath
Anonymous Social Media provide meta-data privacy.
Existing anonymous messaging apps

Yik Yak

whisper

"Attendance is not expected to be high today given the rain and hangovers."

"I tried to facetime campus police last night."
Threat is real
Anonymous messaging meets social filtering
Diffusion of rumor/contagion

Social network/contact network
Diffusion of rumor/contagion
Diffusion of rumor/contagion
Diffusion of rumor/contagion
Diffusion of rumor/contagion
Diffusion of rumor/contagion
Diffusion of rumor/contagion
Diffusion of rumor/contagion
Rumor source detection

can we locate the message author?
**Snapshot**

**Timing**

<table>
<thead>
<tr>
<th>time</th>
<th>from</th>
<th>to</th>
<th>control</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:12</td>
<td>Alice</td>
<td>Spy1</td>
<td>101101</td>
</tr>
<tr>
<td>10:25</td>
<td>Bob</td>
<td>Spy2</td>
<td>101001</td>
</tr>
<tr>
<td>11:01</td>
<td>Mary</td>
<td>Spy3</td>
<td>100100</td>
</tr>
</tbody>
</table>
Snapshot-based adversary
Snapshot reveals the source

- message author is likely to be in the “center”
Rumor Centrality [Shah, Zaman ’11]

Similar performance using Jordan centrality [Zhu, Ying ’13]
Our Goal (on infinite regular trees)

\[ \log \mathbb{P}(\text{detection}) \]

\[ \log N_T \]

\[ \mathbb{P}(\text{Detection}) = \frac{1}{N_T} \]
Setting

- Contact network is infinite, regular, tree
- Adversary has the contact network and the snapshot
- Protocol is allowed to infect any 1-hop neighbors at each time
Line graph
Line graph: diffusion

\[ T = 0 \]
Line graph: diffusion

\[ T = 1 \]
Line graph: diffusion

\[ T = 1 \]
Line graph: diffusion

\[ T = 2 \]
Line graph: diffusion

\[ T = 2 \]
Line graph: diffusion

$T = 3$
Line graph: diffusion

- equivalent to two independent random walks

$T = 3$
Adversary with snapshot

nodes with the message

can we locate the message author?
Maximum likelihood detection

Probability of detection \( \approx \frac{1}{\sqrt{N}} \)
Line graph: adaptive diffusion

$T = 0$
Line graph: adaptive diffusion

Each neighbor is infected w.p. 1/2
Line graph: adaptive diffusion

Node 1 receives message at $T = 1$
Adaptive diffusion prescribes to pass at adaptive rate

\[ p_{\text{infect}} = \frac{h + 1}{T + 1} \]
Line graph: adaptive diffusion

- node 2 receives the message

$T = 2$
Line graph: adaptive diffusion

\[ p_{\text{infect}} = \frac{h + 1}{T + 1} \]
Line graph: adaptive diffusion

- equivalent to Polya’s urn process
- adaptive and asymmetric: Nodes that are infected earlier, spread faster
Given snapshot

can we locate the message author?
Maximum likelihood detection

Likelihoods

diffusion

adaptive diffusion

probability of detection \sim \frac{1}{N}
$d$-regular trees
$d$-regular trees: diffusion

spread with fixed probability $p$

- [Shah & Zaman ’11]
Probability of detection using Jordan centrality

spread with probability

\[ p(h, t) = \frac{h + 1}{t + 1} \]
- **Strategy:**
  - Design the infection to be a symmetric ball of depth $T/2$
  - Such that the source is equally likely to be any node at any given time $T$
- Initially, the author is also the virtual source.
- And randomly selects a neighbor to be the next virtual source.

\[ T = 0 \]

\textbf{d-regular trees: adaptive diffusion}
$d$-regular trees: adaptive diffusion

\[ T = 1 \]

In addition to the message, $v^*$ passes $h = 1$ and $T = 1$ to the chosen neighbor

\[ h = \# \text{ of hops away from true source} \]
at $T=2$, the virtual source passes the message to all its neighbors
\( d \)-regular trees: adaptive diffusion

At \( T=2 \), the protocol has two options:

- Keeping the virtual source token
- Passing the virtual source token
$d$-regular trees: adaptive diffusion

- Keeping the virtual source token with probability $\alpha_{d,T,h}$
- Passing the virtual source token with probability $1 - \alpha_{d,T,h}$
Passing the virtual source token

\[ T = 2 \]
Passing the virtual source token

\[ T = 3 \]
Passing the virtual source token

\[ h = 2 \]
\[ T = 4 \]
Keeping the virtual source token

\[ T = 2 \]
Keeping the virtual source token

\[ T = 3 \]
Keeping the virtual source token

\[ h = 1 \]
\[ T = 4 \]
Adversary with snapshot

can the adversary locate the message author?
Maximum likelihood estimation

- If virtual source is kept with prob. \( \alpha_{d,T,h} = \frac{(d - 1)^{\frac{T}{2} + 1} - h}{(d - 1)^{\frac{T}{2} + 1} - 1} \)
- Nodes that receive message faster, spread faster
Theorem. [Fanti, Kairouz, Oh, Viswanath 2015]

On an infinite $d$-regular tree,
1. adaptive diffusion spreads fast,
   
   $$N_T \approx (d - 1)^{T/2}$$

2. achieves almost perfect obfuscation
   
   $$\mathbb{P}(\text{Detection}) = \frac{1}{N_T - 1}$$

3. The expected distance between the estimated and the true source is $T/2$
What if the tree is irregular?

\[ T = 0 \quad v^* \]

\[ d_v = \begin{cases} 
3 & \text{w.p. 0.5} \\
5 & \text{w.p. 0.5} 
\end{cases} 
\]
$T = 1$ \quad v^* \quad G_T$

$d_v = \begin{cases} 3 & \text{w.p. 0.5} \\ 5 & \text{w.p. 0.5} \end{cases}$
\[ T = 2 \]

\[ d_v = \begin{cases} 
3 & \text{w.p. 0.5} \\
5 & \text{w.p. 0.5} 
\end{cases} \]

\[ G_T \]
\[ T = 3 \]

\[ d_v = \begin{cases} 
3 & \text{w.p. 0.5} \\
5 & \text{w.p. 0.5}
\end{cases} \]

\( G_T \)
\[ T = 4 \]

\[ v^* \]

\[ G_T \]

\[ d_v = \begin{cases} 
3 & \text{w.p. 0.5} \\
5 & \text{w.p. 0.5} 
\end{cases} \]
\[ T = 4 \]

\[ d_v = \begin{cases} 
3 & \text{w.p. 0.5} \\
5 & \text{w.p. 0.5}
\end{cases} \]

\[ \hat{v}_{\text{ML}} = \arg \max_{v \in \partial G_T} \frac{1}{d_{v}s} \prod_{w \in \phi(v_s,v) \setminus \{v_s,v\}} (d_w - 1) \]

\[ F(G_T) \]
\[ T = 4 \]

\[ d_v = \begin{cases} 3 & \text{w.p. 0.5} \\ 5 & \text{w.p. 0.5} \end{cases} \]

\[ \hat{v}_{ML} = \arg \max_{v \in \partial G_T} \frac{1}{d_{vs}} \prod_{w \in \phi(vs,v) \setminus \{vs,v\}} (d_w - 1) \]

\[ F(G_T) = \frac{1}{5 \times 2} \]
Does adaptive diffusion still achieve perfect obfuscation?

\[
d_v = \begin{cases} 
3 & \text{w.p. 0.5} \\
10 & \text{w.p. 0.5}
\end{cases}
\]

\[
d_v = \begin{cases} 
3 & \text{w.p. 0.5} \\
5 & \text{w.p. 0.5}
\end{cases}
\]
Galton-Watson tree $G_T$

$$F(G_T) = \max_{v \in \partial G_T} \frac{1}{d_{vs}} \prod_{w \in \phi(vs,v) \setminus \{vs,v\}} (d_w - 1)$$

- Probability of detection:

$$\mathbb{P}(\hat{v}_{ML} = v^*) = \sum_{G_T} \mathbb{P}(G_T) \mathbb{P}(\hat{v}_{ML} = v^* | G_T) \underbrace{F(G_T)}_{F(G_T)} = \mathbb{E}(F(G_T))$$
Corollary

If \( d_v = \left\{ \begin{array}{c} d_{\text{min}} \quad \text{w.p.} \ p_{\text{min}} \\ \vdots \quad \vdots \\ d_{\text{min}} - 1 \quad \text{w.p.} \ p_{\text{min}} \end{array} \right\} \), and \((d_{\text{min}} - 1)p_{\text{min}} \geq 1\), then

\[
\mathbb{P}(\hat{v}_{\text{ML}} = v^*) = \mathbb{E}_{G_T} \left[ \frac{\max_{v \in \partial G_T} d_{v_s}}{d_{w} - 1} \prod_{w \in \phi(v_s,v) \setminus \{v_s, v\}} \right] = (d_{\text{min}} - 1)^{-T + o(T)}
\]

\[d_v = \left\{ \begin{array}{c} 3 \quad \text{w.p.} \ 0.5 \\ 10 \quad \text{w.p.} \ 0.5 \\ 5 \quad \text{w.p.} \ 0.5 \end{array} \right\} \]
Proof idea for

\[
\min_{v \in \partial G_T} \prod_{w \in \phi(vs,v) \setminus \{vs,v\}} (d_w - 1) = (d_{\min} - 1)^{T + o(T)}
\]

\[
d_v = \begin{cases} 
3 & \text{w.p. } 0.5 \\
5 & \text{w.p. } 0.5
\end{cases}
\]

\[
d_v = \begin{cases} 
3 & \text{w.p. } 0.5 \\
1 & \text{w.p. } 0.5
\end{cases}
\]
Messaging App: Wildfire

Wildfire empowers devices by removing central service providers. It also has stronger anonymity properties than Secret, Whisper, and Yik Yak.

Anonymous, distributed, secure implementation.