Think Globally, Act Locally: On the Optimal Seeding for Nonsubmodular Influence Maximization

Grant Schoenebeck, Biaoshuai Tao, Fang-Yi Yu
Contagions, Diffusion, Cascade...

- Ideas, beliefs, behaviors, and technology adoption spread through networks
- Why do we need to study this phenomena?
  - Better Understanding
  - Promoting good behaviors/beliefs
  - Stopping bad behavior
Influence Maximization

Find the best $K$ nodes to maximize adoptions [KKT03]

• Input
  – Social network $G$
  – Model of contagions
  – Total number of seeds $K$, budget
Influence Maximization

Find the best $K$ nodes to maximize adoptions [KKT03]

- **Input**
  - Social network $G$
  - Model of contagions
  - Total number of seeds $K$, budget

- **Output**
  - Seed set $I$, s.t. $|I| = K$
Find the best $K$ nodes to maximize adoptions

- **Input**
  - Stochastic hierarchical blockmodel (SHBM)
  - $r$-complex contagion
  - Total number of seeds $K$, budget

- **Output**
  - Seed set $I$, s.t. $|I| = K$
Motivation

Can we promote good behaviors/beliefs on a social network if we only know the community structure of the network?
Outline

• Stochastic Hierarchical Blockmodel
• $r$-complex contagions
• Main result
Outline

• Stochastic Hierarchical Blockmodel
• $r$-complex contagions
• Main result
Possible information about Networks

• Full information
Possible information about Networks

• Full information
Possible information about Networks

• Full information
• Query
  – Edge query, node query, ...
Possible information about Networks

- Full information
- Query
  - Edge query, node query, ...
Possible information about Networks

• Full information
• Query
  – Edge query, node query, ...
• Coarser information
  – Community structure,
  – Centrality,
  – Betweenness
Community Structure

- Social networks often can be easily divided into communities densely connected internally.
Hierarchical Community Structure

- Social networks often can be easily divided into communities densely connected internally.
- A community can be easily divided into many sub-communities.
Hierarchical Community Structure

USA

- East coast
  - Boston
  - NYC
- West coast
  - Seattle
  - San Francisco
- Great Lake
  - Ann Arbor

West coast

- Seattle
- San Francisco

East coast

- Boston
- NYC

Great Lake

- Ann Arbor
Stochastic Hierarchical Blockmodel \((V_T, E_T, w, \nu)\)

Connectivity matrix \(w\)

- \(w(R) = 0.1\)
- \(w(F) = 0.5\)
- \(w(G) = 0.6\)
- \(w(H) = 0.4\)

Relative population \(\nu\)

- \(\nu(A) = 30\%\)
- \(\nu(B) = 0.9\)
- \(\nu(C) = 0.85\)
- \(\nu(D) = 0.88\)
- \(\nu(E) = 0.87\)
- \(\nu(F) = 0.87\)
- \(\nu(G) = 0.6\)
Outline

• Stochastic Hierarchical Blockmodel (SHBM)
• $r$-complex contagions
• Main result
\(r\)-Complex Contagions [CLR 79; GEG13]

- Given an initial seed set \(I = \{u, v\}\), and a graph \(G\)
Given an initial seed set \( I = \{u, v\} \), and a graph \( G \)

Node becomes infected if it has at least \( r(= 2) \) infected neighbor
\(r\)-Complex Contagions

- Given an initial seed set \(I = \{u, v\}\), and a graph \(G\)
- Node becomes infected if it has at least \(r\) infected neighbor
Given an initial seed set $I = \{u, v\}$, and a graph $G$.

Node becomes infected if it has at least $r$ infected neighbors.

Local activation function $f_v(x) = \mathbb{I}[x \geq r]$
$r$-Complex Contagions

- Given an initial seed set $I = \{u, v\}$, and a graph $G$.
- Node becomes infected if it has at least $r$ infected neighbor.
- The total number of infected vertices $\sigma_{r,G}(I)$

$$f_v(x) = \mathbb{I}[x \geq r]$$
$$\sigma_{r,G}(I)$$
Given an initial seed set $I = \{u, v\}$, and a distribution over graphs, $\mathcal{G}$, e.g., SHBM.

Node becomes infected if it has at least $r$ infected neighbor

The total number of infected vertices $\sigma_{r,G}(I) = \mathbb{E}_G[\sigma_{r,G}(I)]$
Nonsubmodular vs Submodular InfMax

Submodular InfMax

- linear threshold, independent cascade
- Complexity:
  - $(1 - 1/e)$-approximation

For all $A \subset B \subseteq V$, and $x \in V$

$$f_v(A \cup \{x\}) - f_v(A) \geq f_v(B \cup \{x\}) - f_v(B)$$
Nonsubmodular vs Submodular InfMax

Submodular InfMax

- **linear threshold**, independent cascade
- **Complexity:**
  - $(1 - 1/e)$-approximation

For all $A \subset B \subseteq V$, and $x \in V$

$$f_v(A \cup \{x\}) - f_v(A) \geq f_v(B \cup \{x\}) - f_v(B)$$
## Nonsubmodular vs Submodular InfMax

### Submodular InfMax
- **linear threshold, independent cascade**
- Complexity:
  - $(1 - 1/e)$-approximation

### Nonsubmodular InfMax
- **$r$-complex contagions, general threshold model**
- Complexity:
  - NP-hard to approximate within $n^{1-\epsilon}$
    - [KKT03]
  - NP-hard to approximate within $n^{1-\epsilon}$ on SHBM if nodes can have different thresholds $r$ [ST17]
Outline

• Stochastic Hierarchical Blockmodel
• $r$-complex contagions
• Main result
Find the best $K$ nodes to maximize adoptions

- **Input**
  - stochastic hierarchical blockmodel, $\mathcal{G} = (V_T, E_T, w, v)$
  - $r$-complex contagion
  - Total number of seeds $K$, budget

- **Output**
  - Seed set $I$ to maximize $\sigma_{r,G}(I)$ s.t. $|I| = K$. 
Research Question

Find the best $K$ nodes to maximize adoptions

- **Input**
  - stochastic hierarchical blockmodel, $\mathcal{G} = (V_T, E_T, w, v)$
  - $r$-complex contagion
  - Total number of seeds $K$, budget

- **Output**
  - Seed set $I$ to maximize $\sigma_{r,\mathcal{G}}(I) \approx \max_{|I'|\leq K} \sigma_{r,\mathcal{G}}(I')$.

- Parameters:
  - $w(R) = 0.1$
  - $w(F) = 0.5$
  - $w(G) = 0.6$
  - $w(H) = 0.4$
  - $w(A) = 0.8$, $v(A) = 30%$
  - $w(B) = 0.9$, $v(B) = 10%$
  - $w(C) = 0.85$, $v(C) = 20%$
  - $w(D) = 0.88$, $v(D) = 20%$
  - $w(E) = 0.87$, $v(E) = 10%$
Main result

Given $r$, budget $K$, and a SHBM $(V_T, E_T, w, v)$ with $n \to \infty$, we should put all seed into a community

$$t^* = \arg \max v(t)n \cdot w(t)^r$$

– Large communities
– Proper separation
– Dense tree

<table>
<thead>
<tr>
<th>Community</th>
<th>$w$</th>
<th>$v$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.8</td>
<td>30%</td>
</tr>
<tr>
<td>B</td>
<td>0.9</td>
<td>10%</td>
</tr>
<tr>
<td>C</td>
<td>0.85</td>
<td>20%</td>
</tr>
<tr>
<td>D</td>
<td>0.88</td>
<td>20%</td>
</tr>
<tr>
<td>E</td>
<td>0.87</td>
<td>10%</td>
</tr>
</tbody>
</table>

$w(R) = 0.1$
Observation 1

- Does $r$-complex contagion spread with constant number on Erdős-Rényi Graph $G(n, p)$? [JLT'12]

<table>
<thead>
<tr>
<th>Condition</th>
<th>$p$ equality</th>
<th>Spread Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subcritical</td>
<td>$p = o\left(\frac{1}{n^r}\right)$</td>
<td>doesn’t spread</td>
</tr>
<tr>
<td>Critical</td>
<td>$p = \frac{c}{n^r}$</td>
<td>spread with constant probability</td>
</tr>
<tr>
<td>Supercritical</td>
<td>$p = \omega\left(\frac{1}{n^r}\right)$</td>
<td>spread with high probability</td>
</tr>
</tbody>
</table>
Main Result

Given $r$, budget $K$, and a SHBM $(V_T, E_T, w, v)$ with $n \to \infty$, we should put all seed into a community

$$t^* = \arg \max v(t)n \cdot w(t)^r$$
Main Result

Given $r$, budget $K$, and a SHBM $(V_T, E_T, w, v)$ with $n \to \infty$, we should put all seed into a community

$$t^* = \arg \max v(t)n \cdot w(t)^r$$
Observation 2

Given two isolated $G(n, p)$ with $p = cn^{-1/r}$, and budget $K$, to maximize the infection you should:

1) Go all in $(K, 0)$
2) Hedge your bet: $(K/2, K/2)$
Take-Home Messages

• For nonsubmodular influence maximization (e.g., $r$-complex contagion), putting seeds together to create synergy is more beneficial.

• In sharp contrast to submodular influence maximization (e.g., Linear Threshold, Independent Cascade) where we should spread the seeds to avoid waste of seeds’ power.
Open Problems on Influence Maximization

• Information about graphs
  – Community structure, Centrality, Betweenness
  – Node query, Edge query

• Beyond submodular contagions models
  – $r$-complex contagions
  – general threshold [GGSY16]
  – 2-quasi-submodular [ST17]
Technical Lemma

Let $E_k^n$: the event that $k$ seeds do not infected the graph $G(n, p)$ with $p = cn^{-1/r}$. For all $k \geq r - 1$

\[
\Pr(E_{k+2}^n) \Pr(E_k^n) < \Pr(E_{k+1}^n) \Pr(E_{k+1}^n)
\]

as $n \to \infty$. 

Both graphs are not infected

$k + 2$  $k$  $<$  $k + 1$  $k + 1$
Erdős-Rényi Graphs $\mathcal{G}(n, p)$ with $p = cn^{-1/r}$

Equivalent (when $n \to \infty$) inhomogeneous random walk on $\mathbb{R}$:

• Start at $x = k$;
• In each iteration $i$:
  – move to the left by 1 unit;
  – sample $\xi_i \sim \text{Poisson}\left(\left(\frac{i-1}{r-1}\right) \cdot cn^{-1/r}\right)$, move to right by $\xi_i$ units;
• Terminate if hits $x = 0$;

Two cases:
• Hit $x = 0$: not infected, $E_k^n$
• Go to infinity: infected
Let $E^n_k$: the event that $k$ seeds do not infected the graph $G(n, p)$ with $p = cn^{-1/r}$. For all $k \geq r - 1$
\begin{align*}
\Pr(E^n_{k+2}) \Pr(E^n_k) &< \Pr(E^n_{k+1}) \Pr(E^n_{k+1})
\end{align*}
Both graphs are not infected as $n \to \infty$. 

![Diagram](attachment://diagram.png)
A coupling argument

We couple the two walks $A, B$ in the same way until...

$A, B$ are symmetric, $\mathcal{E}_{symm}$

$A$ hits the $x$-axis, $\mathcal{E}_{skew}$
When $A, B$ are symmetric to $y = x$, $\mathcal{E}_{symm}$
When $A$ hits the $x$-axis, $\mathcal{E}_{skew}$

- Both needs to move $S + 1$ units to reach $(0, 0)$.
  - $A$: $S + 1$ steps *sequentially*.
  - $B$: $S$ steps in $x$-direction and 1 step in $y$-direction *in parallel*.
- $B$ is easier to reach $(0, 0)$, as the Poisson mean is increasing.

$$\xi_i \sim \text{Poisson}\left(\frac{(i-1) \cdot c^r}{r-1}\right)$$
When \( A \) hits the \( x \)-axis, \( \mathcal{E}_{\text{skew}} \)

- Both needs to move \( S + 1 \) units to reach \((0, 0)\).
  - \( A \): \( S + 1 \) steps \textit{sequentially}.
  - \( B \): \( S \) steps in \( x \)-direction and 1 step in \( y \)-direction \textit{in parallel}.
- \( B \) is easier to reach \((0, 0)\), as the Poisson mean is increasing.
Beyond Dense Tree

• Find the densest community
Beyond Dense Tree

- Decompose into dense subtree
  - Find the densest community
  - Dynamic programming