Competition in VM – Completing the Circle

Based on [Lu et al. KDD 2013].
Previous work in Competitive VM

• Mainly follower’s perspective: given state (say of seed selection) of previous companies (agents/players):
  – what’s the best strategy for the “follower” to maximize its spread in the face of the competition?
  – What’s the best strategy for the follower to maximize the blocked influence of the opponent?

• Most competitive VM algorithms not scalable or assume unfettered access to the n/w for all players.
  – How realistic is that?
Campaign runners don’t necessarily have unfettered access to the network! There is an owner of the network. Campaigns need owner’s permission. May need to pay the owner.
A New Business Model – Introducing …

Network owner
Provides VM service.

I need 100 seeds

I need 250 seeds

Competition starts after host selects/allocates seeds.

How should the host select/allocate seeds?
Business Model (contd.)

- Host would like to maximize its expected revenue or its proxy, spread, over all companies.

\[ \sigma_{all}(S_1, \ldots, S_K) := \sum_i \sigma^i(S), \text{ where } S = (S_1, \ldots, S_K). \]

- Host should be fair to all companies: the “ROI” or “bang for the buck” for all companies should be nearly the same.

\[ \text{Bang for the buck for company } i = \frac{\text{spread for } i}{\text{budget of } i}. \]
Why is fairness important?

Possible scenario:

<table>
<thead>
<tr>
<th>iPhone</th>
<th>Pixel 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 seeds</td>
<td>30 seeds</td>
</tr>
<tr>
<td>Spread 10000</td>
<td>Spread 2400</td>
</tr>
</tbody>
</table>

For comparable products, if the b4b is substantially different, dissatisfied company(ies) may take their VM business elsewhere!
Propagation Model

• Which model should we use?
• (C)LT – more natural for product adoption.
• Unfortunately, previous CLT proposals have some disadvantages.
  – The CLT discussed last class (abstracted from Chen et al. and Borodin et al.) – not submodular.
  – Spells computational difficulties.
• Let’s review Borodin et al.’s model next.
WPCLT Model

Seeds influence out-neighbors (followers).
Active nodes influence followers.
Once influenced, relative weights of different (campaign) influences induce a random decision making trial.
WPCLT Model

• Core – *LT Model.*

**Nodes activated by time \( t \).**

\[
\begin{align*}
  w_1 & \\
  w_4 & \\
  w_7 & \\
\end{align*}
\]

My threshold is \( \theta \in [0,1] \).

\[
w_2 + w_3 + w_4 + w_5 + w_6 \geq \theta?
\]

Possible outcome at time \( t + 1 \).

\[
p = \frac{w_2 + w_3 + w_4}{w_2 + \cdots + w_6}.
\]

\[
q = \frac{w_5 + w_6}{w_2 + \cdots + w_6}.
\]
WPCLT Analysis

• A node that became active at any time $t' < t + 1$ gets several attempts to throw its weight around its follower (out-neighbor).

• Spread $\sigma_i(S_i, S_{-i}) := \sigma_i(S_1, ..., S_K)$ in WPCLT is non-monotone (!) and non-submodular. — Details in Ben’s talk next week (stay tuned!).

$\theta_v \in (0.5,1)$  $\theta_w \in (0.5,1)$
WPCLT Analysis

• A node that became active at any time $t' < t + 1$ gets potentially unbounded attempts to throw its weight around its follower (out-neighbor).

• Spread $\sigma_i(S_i, S_{-i}) := \sigma_i(S_1, ..., S_K)$ in WPCLT is non-monotone (!) and non-submodular. – Details in Ben’s talk next week (stay tuned!).

$$\theta_v \in (0.5,1) \quad \theta_w \in (0.5,1)$$

Red seed happens to help blue’s cause!
Introducing the K-LT Model

• Analogous to WPCLP Model – 2 phases.
• Phase 1 influence is joint.
• Phase 2 decision making is controlled by nodes that became active at time \( t \).
• This simple change makes the model (i.e., spread function) monotone and submodular!
• How does K-LT propagation work on the previous example?
Proof of both properties

• Establish equivalence to good old live-edge model!
• Leverage M and SM of reachability.
• Live-edge model construction:
  – Every node picks a favorite in-neighbor w/ proportionate prob and possibly no favorite in-neighbor.
  – All arcs not picked and all i/c arcs to seeds are blocked.
  – A node gets color $C_i$ iff it is reachable from a $C_i$-colored seed via a live-edge path.

• **Key proof idea:** establish $\equiv$ between distribution of sets of nodes with different colors at various timepoints in K-LT model and that of sets of nodes reachable by live-edge paths of different length from various colored seeds.
Fine, what do we want to optimize?

- **Obvious candidate:** $\sigma_{all}(S_1, ..., S_K) := \sum_i \sigma^i(S)$, where $S = (S_1, ..., S_K)$.

- **Proposition:** $\sigma_{all}(S) = \sigma_{LT}(\bigcup_i S_i)$.

Proof: Phase 1 logic $\equiv$ influence propagation under classic LT model. ■

Observe, -- this prop. holds for WPCLT too!

-- seed selection (regardless of who gets which seed) can be done analogously to doing business in LT model.

-- decouple seed selection from allocation.

-- Next focus on allocation.
Optimal Seed Selection

• We focus fair seed selection.
• Fairest SS ensures minimal “divergence” in the b4b ratios (amplification factors) accorded to different companies.
• How should we model “divergence”?
• ∃ many options.
  – Min-max: the happiest guy shouldn’t be too happy!
  – Max-min: the saddest guy shouldn’t be too sad!
  – Directly minimize the $\frac{\max b4b}{\min b4b}$ ratio (or difference).
Fair Seed Allocation

• More formally, given the global seed set $S$, partition it into $K$ subsets $S_1, S_2, \ldots, S_K$, such that:
  • $|S_i| = b_i$ (budget)
  • $S_i \cap S_j = \emptyset, \forall i \neq j$ and $\bigcup_{i=1}^{K} S_i = S$
  • max. bang for the buck is minimized (min-max).

Recall, bang for the buck for company

$$i = \frac{\sigma_i (S)}{b_i}.$$
Calculating Spread/Marginal gain

• Can completely avoid MC simulation: derive a closed-form expression for $\sigma_i(S_i, S_{-i})$.

• Theorem: $\sigma_i(S_i, S_{-i}) = \sum_{v \in V} Pr[\exists S_i \rightarrow v \text{ in the subgraph of } G \text{ induced by } V - S_{-i}] = \sum_{v \in S_i} \sigma_{LT}^{V - S + u} (\{u\})$.

\[
\sum_{u \in S_i} \sigma_{LT}^{V - S + u} (\{u\}).
\]

Adjusted Marginal Gain

\[
\sigma_1(\{u\}, \{v\}) = 1 + 0.3 \\
\sigma_2(\{v\}, \{u\}) = 1 + 0.5. \\
\sigma_{all} = \sigma_1 + \sigma_2 = \sigma_{LT}(\{u, v\}) = 2.8.
\]
How hard is FSA?

- **Theorem:** Strongly NP-complete for $K \geq 3$.

Proof: Reduction from 3PARTITION: Given $A = \{a_1, \ldots, a_{3m}\}, s(a_i) \in \mathbb{Z}^+$,

  $\exists A_1, \ldots, A_m \subseteq A, A_i$ pairwise disjoint $|A_i| = 3$:

  $\sum_{a \in A_1} s(a) = \ldots = \sum_{a \in A_m} s(a)$?

3PARTITION is strongly NP-complete, i.e., is NP-complete even when $\sum_{a \in A} s(a)$ is bounded by a polynomial in $m$. 
Illustrating the Reduction

• E.g. instance of 3PARTITION: \( m = 3; A = \{3, 2, 1, 4, 1, 1, 2, 2, 2\} \).

• Corresp. Instance of FSA: 9 seeds each w/ AMG = size. \( m \) companies each with seed budget 3.

Suppose \( \exists \) a PTIME algorithm for FSA. Run it on the reduced instance above. The max \( b4b = \frac{18}{3 \times 3} \) iff the input instance is a YES-instance.
When $K = 2$

- It resembles PARTITION in this case.
- Can be solved in pseudo-poly time using dynamic programming.
- Efficient heuristic for general case: *Needy Greedy!*
Needy Greedy Algorithm

Algorithm NEEDY-GREEDY

Input: $S$ (with $\delta_u, \forall u \in S$) and $b_i, \forall i \in \{1, \ldots, K\}$.

Output: A $K$-partition of $S$, with $|S_i| = b_i, \forall i$.

1. Initialize $S_i = \emptyset, \forall i$;
2. for each $u \in S$ do
3. $T \leftarrow \{i \mid 1 \leq i \leq K, |S_i| < b_i\}$;
4. $j \leftarrow \text{argmin}_{i \in T} \left\{ \frac{\sigma_i(S_i, S_{-i})}{b_i} \right\}$;
5. $S_j \leftarrow S_j \cup \{u\}$;

Can use a max-heap to run efficiently – in $O(|S| \log K)$ time.
Experimental Evaluation & Conclusions

- EE: See paper.
- Role of host.
- Fair seed allocation.
- Complexity (NP-complete) and efficient heuristic solution.
- **Questions**: What if the influence weights and node thresholds depend on the company?
- Other sort orders for seeds in *Needy Greedy*?
- What if the products significantly differ in popularity: e.g., fairness achievable b/w Nokia and Galaxy is limited. How do we handle this?
- Adaptive seed selection/allocation (game theory)?
- Sometimes propagation phenomenon may proceed like an “S-curve” – need “critical mass” of seeds to reach tipoff point. Blind pursuit of FSA here may be problematic.