Social and Information Networks

Tutorial #8: Influence Spread

University of Toronto CSC303
Winter/Spring 2021
Ian Berlot-Attwell

Week 9: Mar 15-19 (2021)

Today's agenda

In lecture we've covered Influence maximization under the linear threshold and independent cascade influence models

Today:

- Questions from Lecture
- A more general model of influence spread
- Non-progressive influence maximization
- Quercus Quiz

Questions?

Influence Models: Linear Threshold

- Each node $v \in V$ has a random threshold $t_v \sim \mathrm{Unif}([0, 1])$
- Each directed edge $(u, v) \in E$ has some fixed weight $w_{uv} \in [0, 1]$ such that:

$$\forall v \in V : \sum_{u \in V : u \to v} w_{uv} \leq 1$$

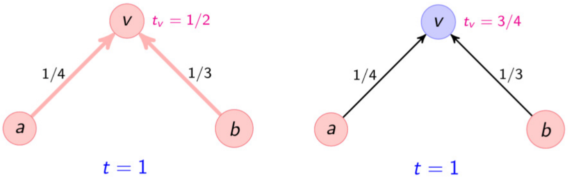
- At time step t, node v is infected if the sum of incident edges from infected nodes exceeds the threshold

Influence Models: Linear Threshold

- Each node $v \in V$ has a random threshold $t_v \sim \text{Unif}([0,1])$
- Each directed edge $(u,v) \in E$ has some fixed weight $w_{uv} \in [0,1]$ such that:

$$\forall v \in V : \sum_{u \in V : u \to v} w_{uv} \leq 1$$

- At time step t, node v is infected if the sum of incident edges from infected nodes exceeds the threshold



- Example where a and b are infected at $t = 0$, and v is or is not infected depending on the random variable t_v

The General Threshold Model

- Question: What are the key parts of the linear threshold model? How may we generalize them?

The General Threshold Model

- Question: What are the key parts of the linear threshold model? How may we generalize them?

- We retain our random threshold $t_v \sim \text{Unif}([0,1])$
- Instead of weighted edges, for each node v we defined a *threshold function* $f_v : \mathcal{P}(V) \to [0,1]$
- Let $\mathcal{I}_t(v) : V \to \mathcal{P}(V)$ is the function that maps v to v's infected neighbours at time time
- An uninfected node v now becomes infected if

$$f_v(\mathcal{I}_t(v)) > t_v$$

The General Threshold Model

- Question: What are the key parts of the linear threshold model? How may we generalize them?

- We retain our random threshold $t_v \sim \text{Unif}([0,1])$
- Instead of weighted edges, for each node v we defined a *threshold function* $f_v : \mathcal{P}(V) \to [0,1]$
- Let $\mathcal{I}_t(v) : V \to \mathcal{P}(V)$ is the function that maps v to v's infected neighbours at time time
- An uninfected node v now becomes infected if

$$f_v(\mathcal{I}_t(v)) > t_v$$

- Question: How do we represent Linear Threshold model as a General Threshold Model

The General Threshold Model

- Question: What are the key parts of the linear threshold model? How may we generalize them?

- We retain our random threshold $t_v \sim \text{Unif}([0,1])$
- Instead of weighted edges, for each node v we defined a *threshold function* $f_v : \mathcal{P}(V) \to [0,1]$
- Let $\mathcal{I}_t(v) : V \to \mathcal{P}(V)$ is the function that maps v to v's infected neighbours at time time
- An uninfected node v now becomes infected if

$$f_v(\mathcal{I}_t(v)) > t_v$$

- Question: How do we represent Linear Threshold model as a General Threshold Model
  - $f_v(S) := \sum_{u \in S} w_{uv}$
- Question: Is the expected number of eventual adopters, $f(S)$, submodular? Is it monotone?

The General Threshold Model

- Question: What are the key parts of the linear threshold model? How may we generalize them?

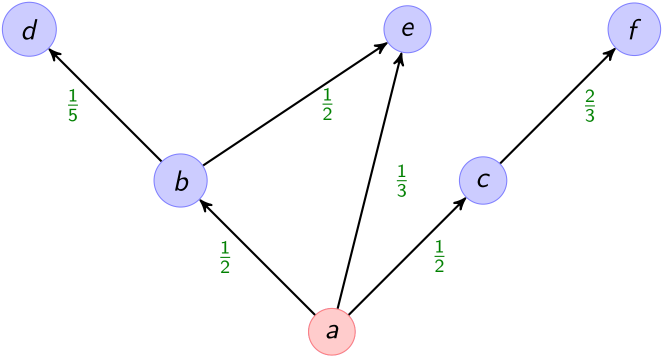
- We retain our random threshold $t_v \sim \text{Unif}([0,1])$
- Instead of weighted edges, for each node v we defined a *threshold function* $f_v : \mathcal{P}(V) \to [0,1]$
- Let $\mathcal{I}_t(v) : V \to \mathcal{P}(V)$ is the function that maps v to v's infected neighbours at time time
- An uninfected node v now becomes infected if

$$f_v(\mathcal{I}_t(v)) > t_v$$

- Question: How do we represent Linear Threshold model as a General Threshold Model
  - $f_v(S) := \sum_{u \in S} w_{uv}$
- Question: Is the expected number of eventual adopters, $f(S)$, submodular? Is it monotone?
  - No, consider that on a clique we could define f_v so that all nodes are infected for a specific initial set $S \subset V$, and otherwise no new nodes are infected

The Independent Cascade Model

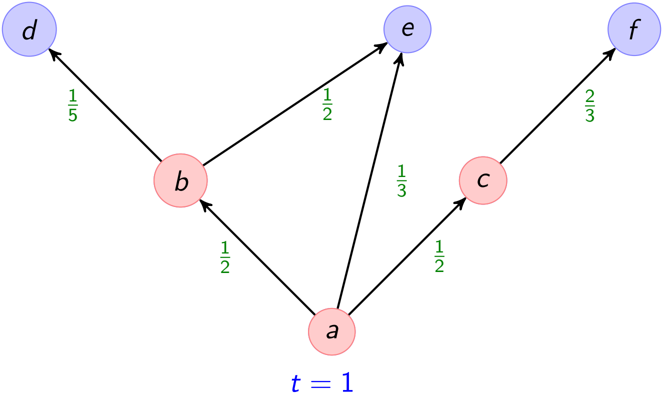
- Each edge (u, v) has an associated probability p_{uv}.
- In each step t, nodes that adopted technology at step $t - 1$ "infect" each of their uninfected neighbors independently with probability p_{uv}.



$t = 0$

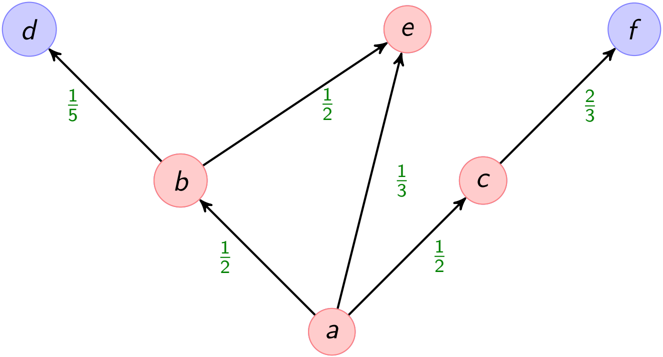
The Independent Cascade Model

- Each edge (u, v) has an associated probability p_{uv}.
- In each step t, nodes that adopted technology at step $t-1$ "infect" each of their uninfected neighbors independently with probability p_{uv}.



The Independent Cascade Model

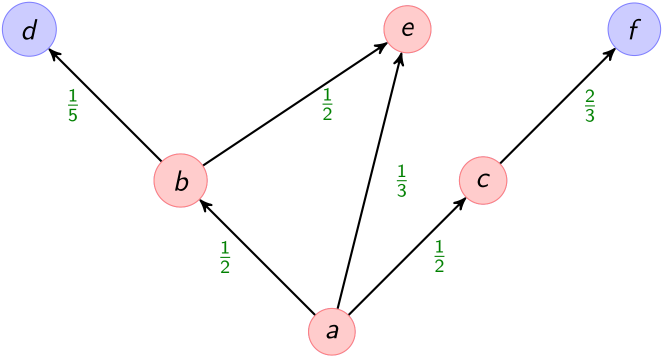
- Each edge (u, v) has an associated probability p_{uv}.
- In each step t, nodes that adopted technology at step $t-1$ "infect" each of their uninfected neighbors independently with probability p_{uv}.



$t = 2$

The Independent Cascade Model

- Each edge (u, v) has an associated probability p_{uv}.
- In each step t, nodes that adopted technology at step $t-1$ "infect" each of their uninfected neighbors independently with probability p_{uv}.



$t = 3$

The General Cascade Model

- Question: What are the key parts of the general cascade model? How may we generalize them?

The General Cascade Model

- Question: What are the key parts of the general cascade model? How may we generalize them?

- We retain the idea that nodes infected at time $t-1$ attempt to infect their neighbours with some probability at time t
- We let the probability that some node v is infected by a node u as $p_v(u, F)$ where $F \subset V$ is the set of nodes that have already tried and failed to infect v
- $p_v : V \times \mathcal{P}(V) \to [0, 1]$
- Question Is there a problem with this model?
  - As written thusfar, it could depend on the order in which nodes attempt to infect v. For this reason, p_v is restricted to be order independent
  - For any set of infected neighbours $u_1, u_2, \ldots u_l$ the order in which they infect v the overall probability of infection must be the same

The General Cascade Model

- Question: What are the key parts of the general cascade model? How may we generalize them?

- We retain the idea that nodes infected at time $t-1$ attempt to infect their neighbours with some probability at time t
- We let the probability that some node v is infected by a node u as $p_v(u, F)$ where $F \subset V$ is the set of nodes that have already tried and failed to infect v
- $p_v : V \times \mathcal{P}(V) \to [0, 1]$
- Question Is there a problem with this model?
  - As written thusfar, it could depend on the order in which nodes attempt to infect v. For this reason, p_v is restricted to be order independent
  - For any set of infected neighbours $u_1, u_2, \ldots u_l$ the order in which they infect v the overall probability of infection must be the same
- Question: How do we represent Independent Cascade model as a General Cascade Model

The General Cascade Model

- Question: What are the key parts of the general cascade model? How may we generalize them?
- We retain the idea that nodes infected at time $t-1$ attempt to infect their neighbours with some probability at time t
- We let the probability that some node v is infected by a node u as $p_v(u, F)$ where $F \subset V$ is the set of nodes that have already tried and failed to infect v
- $p_v : V \times \mathcal{P}(V) \to [0,1]$
- Question Is there a problem with this model?
  - As written thusfar, it could depend on the order in which nodes attempt to infect v. For this reason, p_v is restricted to be order independent
  - For any set of infected neighbours $u_1, u_2, \ldots u_l$ the order in which they infect v the overall probability of infection must be the same
- Question: How do we represent Independent Cascade model as a General Cascade Model
  - $p_v(u, F) := p(u, v)$

The General Threshold Model & The General Cascade Model

- General Threshold Model: Node v is infected at time $t+1$ if $f_v(\mathcal{I}_t(v)) > t_v$
- General Cascade Model: Node u, infected at time t, infects node v with probability $p(u, S)$ where S is the set of nodes that have failed to infect u thusfar
- Question: Can we represent a general threshold model as a general cascade model?

The General Threshold Model & The General Cascade Model

- General Threshold Model: Node v is infected at time $t+1$ if $f_v(\mathcal{I}_t(v)) > t_v$
- General Cascade Model: Node u, infected at time t, infects node v with probability $p(u, S)$ where S is the set of nodes that have failed to infect u thusfar
- Question: Can we represent a general threshold model as a general cascade model?

$$
\begin{aligned}
p_v(u, S) &= P(u \text{ infects } v | S \text{ didn't infect } v) \\
&= \frac{P(u \text{ infects } v \wedge S \text{ didn't infect } v)}{P(S \text{ didn't infect } v)} \\
&= \frac{P(f_v(S \cup \{u\}) > t_v \geq f_v(S))}{P(t_v \geq f_v(S))} \\
&= \frac{f_v(S \cup \{u\}) - f_v(S)}{P(1 - f_v(S))}
\end{aligned}
$$

The General Threshold Model & The General Cascade Model

- General Threshold Model: Node v is infected at time $t+1$ if $f_v(\mathcal{I}_t(v)) > t_v$
- General Cascade Model: Node u, infected at time t, infects node v with probability $p(u, S)$ where S is the set of nodes that have failed to infect u thusfar
- Question: Can we represent a general cascade model as a general threshold model?

The General Threshold Model & The General Cascade Model

- General Threshold Model: Node v is infected at time $t+1$ if $f_v(\mathcal{I}_t(v)) > t_v$
- General Cascade Model: Node u, infected at time t, infects node v with probability $p(u, S)$ where S is the set of nodes that have failed to infect u thusfar
- Question: Can we represent a general cascade model as a general threshold model?
- Let $S = \{s_1, s_2, \ldots s_k\}$, and $S_i := \{s_1 \ldots s_i\}$

$$\begin{aligned}
f_v(S) &= P(S \text{ infects } v) \\
&= 1 - P(S \text{ doesn't infect } v) \\
&= 1 - \prod_{i=1}^{k} P(u_i \text{ doesn't infect } v | S_{i-1} \text{ doesn't infect } v) \\
&= 1 - \prod_{i=1}^{k} (1 - p(u_i, S_{i-1}))
\end{aligned}$$

Non-Progressive Influence

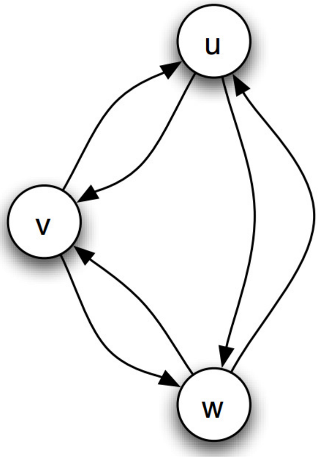
- Thusfar, all the influence models we've seen are *progressive*, nodes that become infected never cease being infected
- Suppose we're modeling something like the use of a subscription service
  - Users can start or stop any any time
  - We assume users are more likely to subscribe if people they know are also subscribed
  - We want to maximize our revenue, or rather the sum of the number of people subscribed at each timestep
  - We can create an initial set of adopters, but these initial adopters can be at different points in time
- How can we model this? How can we pick our initial adopters?

Reducing Non-Progressive Influence to Progressive Influence

- We can model non-progressive influence as progressive influence using a layered graph
- For our original graph $G = (V, E)$, and a time horizon of τ timesteps, we create G^τ by creating τ duplicates of the nodes and edges of G (e.g. v becomes v_t for $t = 1, 2, \ldots \tau$)
- We add directed edges from u_t to v_{t+1} for all u_t such that $(u, v) \in E$
- This is the same approach as we saw in class that allowed us to model a special case of SIS as SIR
- We can now analyze this problem or choose initial adopters on G^τ as if it were a progressive influence problem

Reducing Non-Progressive Influence to Progressive Influence

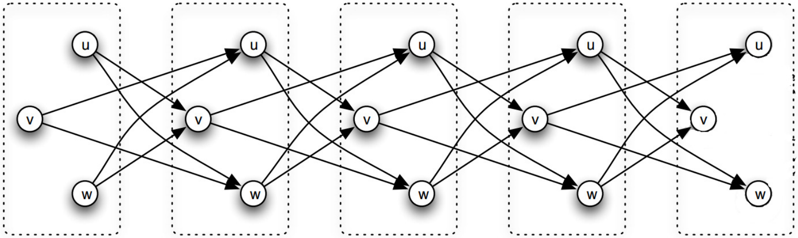
G



[Modified from E&K Fig 21.5]

Reducing Non-Progressive Influence to Progressive Influence

G^5



[Modified from E&K Fig 21.6a]

Quercus Quiz