CSC200: Lecture 35

Allan Borodin

Annoucements and todays agenda

- Announcements
  1. First 3 questions for A4 have been posted.
  2. Some students are trying to survey how much interest there is in the course CSC375 (the enriched version of CSC373) in the hope that it will be offered again. If interested in particpating in this survey (even if you are not currently interested in taking CSC375) please look at: https://csc.cdf.toronto.edu/mybb/showthread.php?tid=12601

- Todays agenda
  1. Introduce the stochastic linear threshold and independent cascade influence spread models.
  2. A "natural" greedy algorithm for maximizing the value of a monotone submodular set function subject to a cardinality constraint.
  3. Comparing the greedy algorithm with other simple algorithms.
  4. We will conclude this discussion with some further comments and also mention models for competitive influence spread.
  5. We begin the small world phenomenon in Chapter 20.

Computationally manageable influence maximization models; monotone submodular set functions

- Some spread models have the following nice properties.

Let $f(S)$ be size (or more generally a real value benefit since some nodes may be more valuable) of the final set S of adopters satisfying:

1. Monotonicity: $f(S) \leq f(T)$ if S is a subset of T
2. Submodularity: $f(S+v) - f(S) \geq f(T+v) - f(T)$ if S is a subset of T

- Where have we seen such functions before?
- The simple threshold examples considered thus far are monotone processes but are not submodular in general. Are these contrived worst case network examples?
- But some variants of the threshold model and related models do satisfy these properties. We consider two such stochastic models.

Linear threshold model

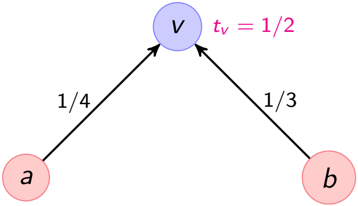
- We have an edge weighted (undirected or directed) network where weight $w(u, v)$ represents the relative influence (e.g. quantitative version of weak and strong ties) of node u on node v.
- Now each nodes threshold $q(v)$ is chosen randomly in $[0, 1]$ to model lack of knowledge as to how easy it is to influence a given individual.
- A node v adopts A if the sum of all edge weights into v exceeds the randomly chosen $q(v)$.
- **Goal:** find an initial set of k adopters so as to maximize the expected number (or benefit) of eventual adopters. (This is a stochasitic process so that we are trying to optimize the expected value of the process.)
- **Aside:** We often use the language of disease spread and say "infected nodes" rather than "already influenced nodes".

The linear threshold model

- Each node v chooses a threshold t_v randomly from $[0, 1]$.
- Each edge (u, v) has assigned weight w_{uv} from $[0, 1]$ such that

$$\sum_{u \to v} w_{uv} \leq 1.$$

- In each step t, a node v is infected if the weighted sum of incident edges coming from infected neighbors exceeds threshold.



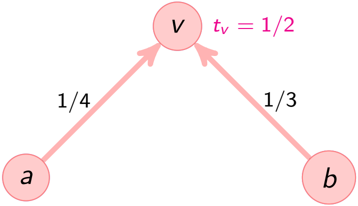
$t = 0$

The linear threshold model

- Each node v chooses a threshold t_v randomly from $[0, 1]$.
- Each edge (u, v) has assigned weight w_{uv} from $[0, 1]$ such that

$$\sum_{u \to v} w_{uv} \leq 1.$$

- In each step t, a node v is infected if the weighted sum of incident edges coming from infected neighbors exceeds threshold.



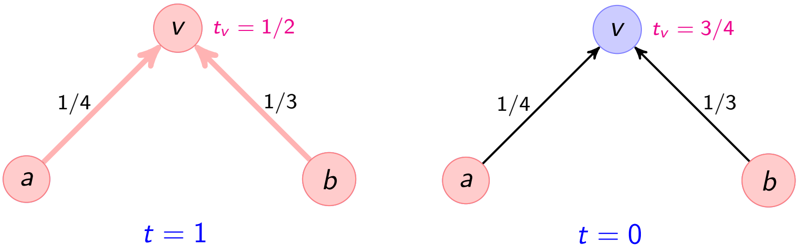
$t = 1$

The linear threshold model

- Each node v chooses a threshold t_v randomly from $[0, 1]$.
- Each edge (u, v) has assigned weight w_{uv} from $[0, 1]$ such that

$$\sum_{u \to v} w_{uv} \leq 1.$$

- In each step t, a node v is infected if the weighted sum of incident edges coming from infected neighbors exceeds threshold.

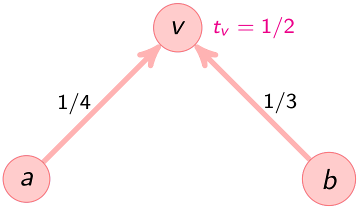


The linear threshold model

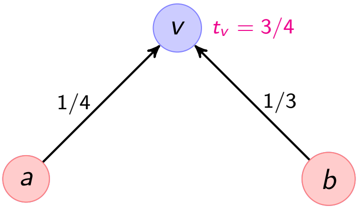
- Each node v chooses a threshold t_v randomly from $[0, 1]$.
- Each edge (u, v) has assigned weight w_{uv} from $[0, 1]$ such that

$$\sum_{u \to v} w_{uv} \leq 1.$$

- In each step t, a node v is infected if the weighted sum of incident edges coming from infected neighbors exceeds threshold.



$t = 1$



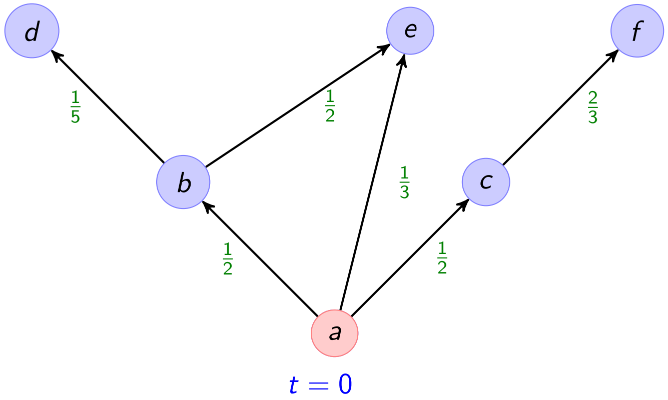
$t = 1$

Independent cascade influence model

- We again have an edge weighted network (as in threshold model) but now the weights $p(u,v) \leq 1$ represent the probability that node u will influence node v given one and only one chance to do so.
- That is, if node u adopts A at time t, then with probability $p(u,v)$, node v will adopt v at time $t+1$.
- After this node u will *not* have another opportunity to influence v.
- **Goal:** is to find initial set of adopters to maximize the expected number of eventual adopters.
- Threshold and (especially) cascade processes are motivated by models for the contagious spread of disease. Should disease spread and influence spread should be governed by similar processes?
  - See `http://www.economist.com/blogs/babbage/2012/04/social-contagion`

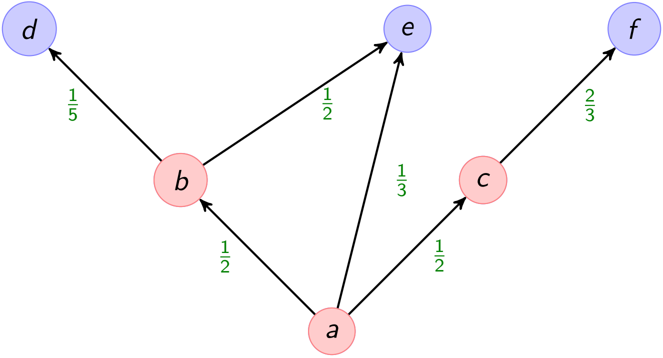
The Independent Cascade Process

- 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 with probability p_{uv}.



The Independent Cascade Process

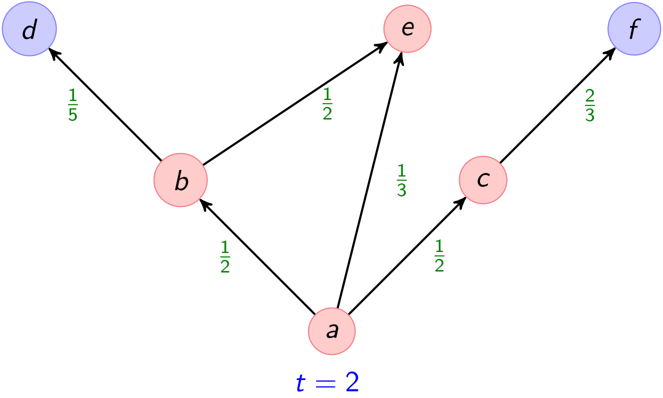
- Each edge (u, v) has an associated probability p_{uv}.
- In each step t, nodes that adopted technology at step $t-1$ "infect" each of their uninfected neighbors with probability p_{uv}.



$t = 1$

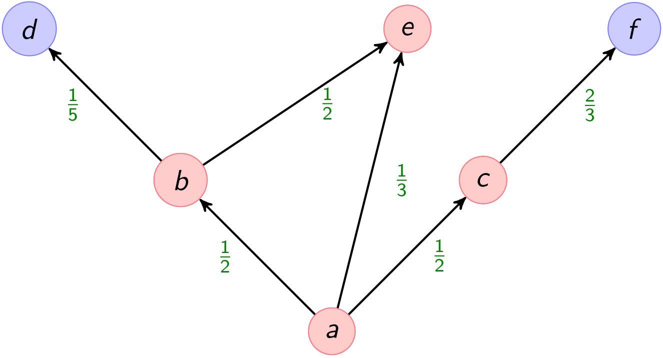
The Independent Cascade Process

- 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 with probability p_{uv}.



The Independent Cascade Process

- 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 with probability p_{uv}.



$t = 3$

How to select a good set of initial adopters

- While in general it is computationally hard to find an optimal set of initial adopters, the stochastic linear threshold and independent cascade models satisfy the monotonicity and submodular properties (for $f(S)$ being the expected number of eventual adopters).

- This allows for a very simple "greedy" algorithm that (provably) selects a set S such that $f(S)$ is within a factor $(1 - \frac{1}{e}) \sim .63$ of optimality. What does "decreasing marginal gains" suggest?

How to select a good set of initial adopters

- While in general it is computationally hard to find an optimal set of initial adopters, the stochastic linear threshold and independent cascade models satisfy the monotonicity and submodular properties (for $f(S)$ being the expected number of eventual adopters).
- This allows for a very simple "greedy" algorithm that (provably) selects a set S such that $f(S)$ is within a factor $(1 - \frac{1}{e}) \sim .63$ of optimality. What does "decreasing marginal gains" suggest?
- The greedy strategy is to iteratively add (to whatever nodes have already been selected) one new initial adopter so as to maximize the expected marginal gain.
- We need to simulate the stochastic process for sufficiently many trials to determine the next node to add. (When different nodes to have different values, accurate simulation requires that the ratio of such values is reasonably bounded.)

An experimental study comparing methods: Kempe, Kleinberg, Tardos

- To test the usefulness of the models being studied, Kempe et al. compare the best marginal gain greedy algorithm with three other simple methods that do not require simulating the process.
- Namely, they compare against:
  - Greedy by highest degree first
  - Greedy by centrality, i.e. by best average path
  - Random choice of adopters
- The experimental data set is an undirected multi-graph based on jointly authored papers by physicists.
- Here we have r edges between u and v if they have been co-authors on r papers.
  - In the threshold model, weights $w(u, v)$ are chosen proportional to the multiplicity of edges between u and v.
  - In the independent cascade model, each edge is given the same probability of success.
  - In the weighted cascade model, probabilities are set proportional to the degree.

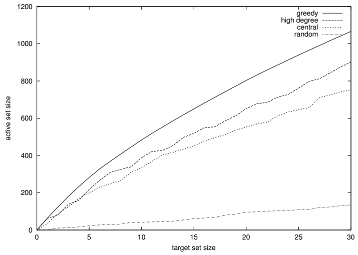


Figure 1: Results for the linear threshold model

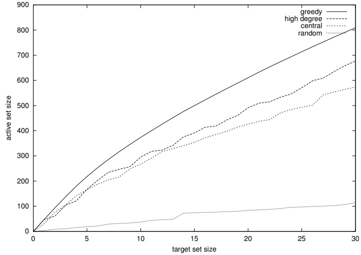


Figure 2: Results for the weighted cascade model

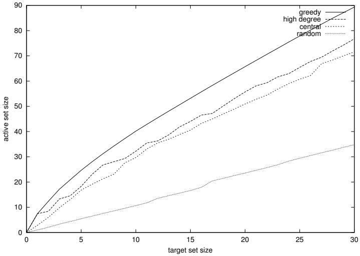


Figure 3: Independent cascade model with probability 1%

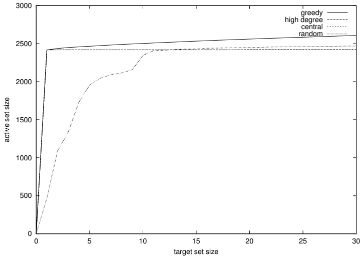


Figure 4: Independent cascade model with probability 10%

Experimental Results from Kempe, Kleinberg, Tardos (2003): "Maximizing the spread of influence through a social network," KDD-03.

What do these experiments suggest

- Cearly the main suggestion is that the marginal gain greedy not only gaurantees a good (expected) “social welfare”, it outperforms the other three simple methods for this specific social network. For example, why should marginal gain greedy outperform the highest degree method?
- The naive random method has the worst performance.
- The improved performance of the marginal gain greedy algorithm comes at a cost, namely having to simulate the process many times to achieve a good candidate for the next node to influence. This can be prohibitive for a large network. There is recent work that addresses the issue of efficiency for large networks.

Some lessons to be learned about influence in a social network (Chapter 19)

- In population-level effects, it can be relatively difficult for a new technology/product/idea to get past a tipping point
- In contrast in social networks, new products/ideas (rumours) can spread extensively and quickly.
- But tightly knit communities (clusters) can stall the spread.
- We saw in the early part of the course that weak ties are often bridges or local bridges between different communities.
- Hence such weak ties may convey some degree of awareness to another community but not likely to change behaviour especially if that change has risks as in political movements and high stakes economic decisions.

Further considerations (collective action)

- Section 19.6 seems to have been (but was not) written after events during the last few years in the mideast, Hong Kong and even Toronto on a smaller scale.

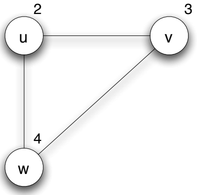
- The discussion here begins to combine aspects of social network interaction (e.g. transmitting information) with direct benefit population effects (being part of a large demonstration).

- In particular, the organization for demonstrations against a regime can begin with discussions within a community but for someone to participate, it usually takes some knowledge that there will be a sufficiently large population wide participation.

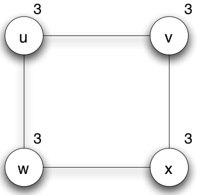
- On a smaller scale, when challenging a mayor or a CEO, the same phenomena may be operating.

Knowledge and common knowledge

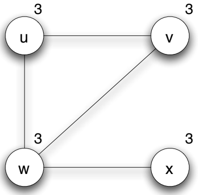
- Our first example of a tightly knit community blocking a complete cascade occurred even when everyone knew the common threshold q.
- A uniform threshold is not that realistic in any reasonable size social network.
  - We might have a sense of the thresholds for our friends but not of all their friends (and their friends friends, etc.)
- The example in Figure 19.14 illustrates the impact of limited knowledge even when everyone knows the entire network but only knows their friends and their own absolute (i.e. not fractional in this example) thresholds.
- Here threshold k means that the node (being me) will participate if at least k people (including myself) will do so.



(a) *An uprising will not occur*



(b) *An uprising will not occur*



(c) *An uprising can occur*

[Fig 19.14, E&K]

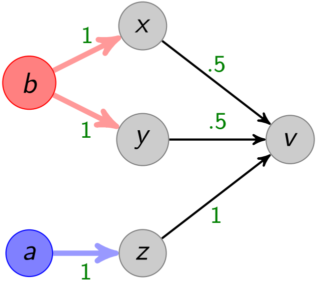
Further considerations: competitive influence spread

- In many economic, social, and political settings the spread of influence is a competitive process.
- It may be that both technologies (political factions, etc.) A and B are competing for new adopters in a social network by promotion via an initial set of adopters (people with vested interests, etc.).
- There are many models for how such competition is resolved.
- One possibility is to use the stochastic independent cascade model and then the first technology (political faction, etc.) to have a "path of adoption" succeeds (breaking ties in some manner).
- That is, after the edge probabilities are instantiated, we consider the shortest paths to a node (if any exist) from the initial adopters (party faithful, etc.) to the initially uncommitted.

The Wave Propagation Process

- Two technologies A and B with their sets of initial adopters I_A and I_B.
- Technology spreads according to the Independent Cascade process.
- If a node is successfully infected at the same step t by both
  - set of nodes V_A that adopt technology A
  - set of nodes V_B that adopt technology B

  it will adopt technology A with probability $\dfrac{|V_A|}{|V_A| + |V_B|}$

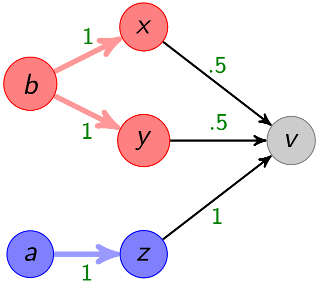


Example

The Wave Propagation Process

- Two technologies A and B with their sets of initial adopters I_A and I_B.
- Technology spreads according to the Independent Cascade process.
- If a node is successfully infected at the same step t by both
  - set of nodes V_A that adopt technology A
  - set of nodes V_B that adopt technology B

  it will adopt technology A with probability $\dfrac{|V_A|}{|V_A| + |V_B|}$

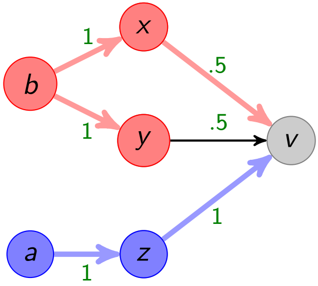


Example

The Wave Propagation Process

- Two technologies A and B with their sets of initial adopters I_A and I_B.
- Technology spreads according to the Independent Cascade process.
- If a node is successfully infected at the same step t by both
  - set of nodes V_A that adopt technology A
  - set of nodes V_B that adopt technology B

  it will adopt technology A with probability $\dfrac{|V_A|}{|V_A| + |V_B|}$



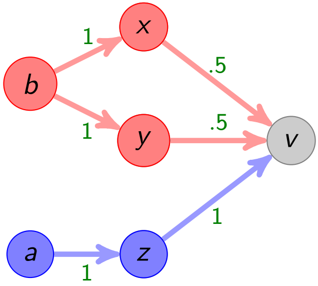
Example

- $Pr[v \text{ adopts } A \,|\, x, z \text{ reached } v] = \frac{1}{2}$

The Wave Propagation Process

- Two technologies A and B with their sets of initial adopters I_A and I_B.
- Technology spreads according to the Independent Cascade process.
- If a node is successfully infected at the same step t by both
  - set of nodes V_A that adopt technology A
  - set of nodes V_B that adopt technology B

  it will adopt technology A with probability $\dfrac{|V_A|}{|V_A| + |V_B|}$



Example

- $Pr[v \text{ adopts } A \,|\, x, z \text{ reached } v] = \frac{1}{2}$
- $Pr[v \text{ adopts } A \,|\, x, y, z \text{ reached } v] = \frac{1}{3}$

Further considerations: the "bilingual option"

- In the advanced material (Section 19.7C), the possibility of a third option is considered.
- Here the model allows an individual to maintain both technologies (languages, ideologies, cultural practices) but at a cost c.
- Every individual now can choose to be unilingual (adopting just A or just B) or to be bilingual adopting both (denoted AB).
- Ignoring the cost, the coordination benefit (for each edge) is represented in Figure 19.18.

		w: A	w: B	w: AB
v	A	a, a	$0, 0$	a, a
	B	$0, 0$	b, b	b, b
	AB	a, a	b, b	$(a, b)^+, (a, b)^+$

Figure : A Coordination Game with a bilingual option. Here the notation $(a, b)^+$ denotes the larger of a and b. [Fig 19.18, E&K]

A concluding comment for chapter 19

- The last sentence of the chapter makes the final comment:

  > *Even small extensions such as the one considered here (the bilingual option) can introduce significant new sources of complexity, and the development of even richer extensions is an open area of research.*

- Indeed such analytic studies of influence spread in more complex networks is an emerging field of significant research interest impacting computer science, sociology, economics, and political science.