Lecture Notes:Maximizing the Spread of Influence through a Social Network
Lecture 1: March 20, 2012
Scribes: Kiana Rod and Yi Wang

Influence maximization problem(IMP)

A social network is the graph of relationships and interactions within a group of individuals that plays a fundamental role as a medium for the spread of information, ideas, and influence among its members. The Influence maximization problem asks, for a parameter k, to nd a k-node set of maximum Influence. This problem has applications in viral marketing, where a company may wish to spread the rumor of a new product via the most influential individuals in popular social networks[1]. Given a network G=(V,E) and a model M of a diffusion process that take place on G, the goal is to find K initial adopters who will lead to most number of adoptions.
For any subset $A \subseteq V$ let $\delta(A)$ denote the expected number of individuals who have adopted the "innovations" given that A is the initial set of adopters.

Input/Output: Input is a network G=(V,E)and a positive integer K and Output is a subset $A \subseteq V$, $\mid A \mid \leq K$ such that $\delta(A)$ is maximized. To understand the hardness of the problem and to come up with good algorithms for the problem, let us visit Maximum Coverage.

Maximum Coverage: Input: A ground set $u = \{$ 1,2,...,n$\}$ and a collection of subsets of $u : C_1, C_2, ..., C_m$

$$\bigcup_{i=1}^{m} C_i = u$$

Additionally a positive integer K.
Output:A sub collection $\{i_1, i_2, ..., i_k\}$ such that $\mid \bigcup_{j=1}^{k} C_{ij} \mid \ is\ maximum.$
For any subset $A \subseteq \{1, 2, ..., m\}$, let

$$Coverage(A) = \mid \bigcup_{i \in A} C_i \mid$$

Algorithm 1 Greedy Algorithm

1: $A \leftarrow 0$
2: $For\ i = 1\ to\ k\ do$
3: $Pick\ j \in \{1, 2, ..., m\}\ coverage\ (A \cup \{j\}) - coverage(A)\ is\ maximum$
4: $A \leftarrow A \cup \{j\}$
5: $return\ A$

Let Opt $\subseteq \{12, .., m\}$ denote an optimal solution to maximum coverage.

Theorem: Let A be the solution returned by Greedy.

$$Coverage(A) \geq (1 - \frac{1}{e})Coverage(Opt)$$

Proof: Let x_i denote the number of new elements covered by the choice in iteration i.

$$(x_i = Coverage(A \cup \{j\}) - Coverage(A)$$

Let y_i denote the number of elements covered by the choice in iterations 1,2,...,i.
Let $z_i = coverage(Opt) - y_i$

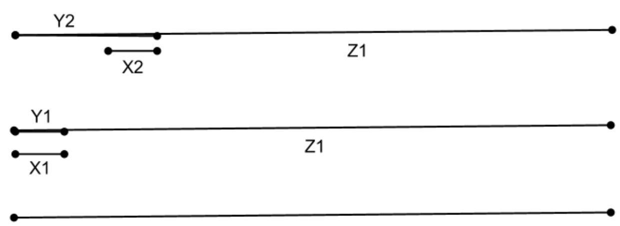

Figure 1: Coverage(Opt), $x_0 = 0$ $y_0 = 0$ $z_0 = Coverage(opt)$

Claim:

$$z_i \leq (1 - \frac{1}{k})^i Coverage(Opt)$$

suppose this were true then :

$$y_k = Coverage(Opt) - z_k \geq Coverage(Opt) - (1 - \frac{1}{k})^k Coverage(Opt)$$

Recall that $e^x \geq 1 + x$ for all real x.

$$y_k \geq Coverage(Opt) - ((e^{\frac{-1}{k}})^k).Coverage(Opt)$$

$$y_k \geq (1 - \frac{1}{e}).Coverage(Opt)$$

Proof of Claim:

Induction Hypothesis:

$$z_{i-1} \leq (1 - \frac{1}{k})^{i-1} Coverage(Opt)$$

This is clearly true for $i - 1 = 0$ (Base Case). Now observe that $x_i \geq \frac{z_{i-1}}{k}$ due to greedy choice.

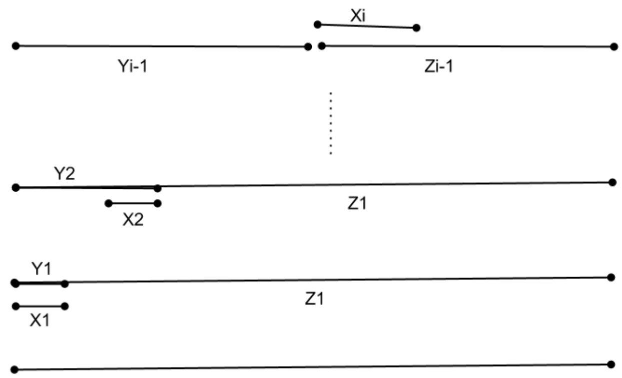

Figure 2: Coverage(Opt)

$$z_i = z_{i-1} - x_i$$

$$z_i \leq z_{i-1} - \frac{z_{i-1}}{k} = (1 - \frac{1}{k})z_{i-1} \leq ((1 - \frac{1}{k})^i)Coverage(Opt)$$

what might a similar greedy algorithm for IMP be?

Algorithm 2 Greedy Algorithm

1: $A \leftarrow 0$
2: $For\ i = 1\ to\ k\ do$
3: $Pick\ v : \delta(A \cup \{v\}) - \delta(A)\ is\ maximized$
4: $A \leftarrow A \cup \{v\}$
5: $return\ A$

Models: Linear Threshold and Independent Cascade Model

The Linear Threshold and Independent Cascade Models [2] are two of the most basic and widely studied diffusion models. Both the Linear Threshold and Independent Cascade Models involve an initial set of active nodes A_0 that start the diffusion process. We define the influence of a set of nodes A, denoted $\delta(A)$, to be the expected number of active nodes at the end of the process, given that A is this initial active set A_0. While in general it is computationally hard to find an optimal set of initial adopters, the linear threshold and independent cascade models satisfy the monotonicity and sub modular properties. The greedy strategy is to iteratively add (to whatever nodes have already been selected) one new initial adopter to maximize the expected marginal gain. The linear threshold model originally proposed by Granovetter [3]. this model is defined over a graph representing a social network of potential adopters. There exists a subset of individuals who have already adopted the innovation. Each member is assumed to adopt the innovation if the fraction of her neighbors that have adopted is above a certain threshold. A node v is influenced by each neighbor w according to a weight $b_{v,w}$ such that:

$$\sum_{w \in neighbors(v)} b_{v,w} \leq 1$$

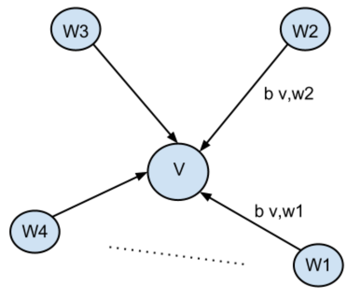

Figure 3: v is influenced by each neighbor w

For each node v there is a threshold $\theta_v \in [0, 1]$ let A be the initial adopters. These nodes are active, the rest are initially inactive. The diffusion process unfolds deterministically discrete steps. In time step t all nodes that were active in time step t-1 remains active.In addition we activate any inactive node v for which:

$$\sum_{w \in ActiveNeighbors(v)} b_{v,w} \geq \theta_v$$

where $\delta(A)$ is the number of nodes that become active.
A social network is represented as a directed graph, with each person (customer) as a node. Nodes start either active or inactive. An active node may trigger activation of neighboring nodes. According to monotonicity assumption active nodes never deactivate. The proof of approximation depends on:

i) Coverage is monotonic . i.e.

$$Coverage(A \cup \{j\}) \geq Coverage(A)$$

ii) Coverage is sub modular for any A,B : $A \subseteq B$

$$Coverage(B \cup \{j\}) - Coverage(B) \leq Coverage(A \cup \{j\}) - Coverage(A)$$

For the Linear Threshold Model and Independent Cascade Model , δ is both monotonic and sub modular.

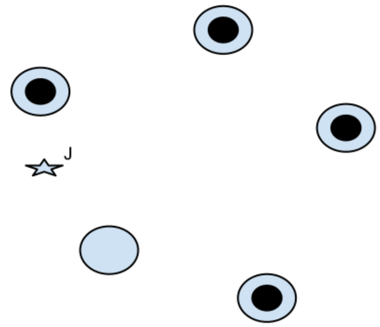

Figure 4: Coverage is sub modular for any A,B : $A \subseteq B$

References

[1] D. Kempe, J. Kleinberg, E. Tardos. Maximizing the Spread of Influence through a Social Network. Proc. 9th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, 2003.

[2] J. Goldenberg, B. Libai, E. Muller. Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth. Marketing Letters 12:3(2001), 211-223.

[3] M. Granovetter, Threshold models of collective behavior, the american Journal of sociology, vol. 83, no. 6, pp.1420-1443, May 1978