Greedily maximizing social influence

February 23, 2018

In this document we present the results of some experimentation with the algorithm presented in the paper ”Maximizing the Spread of Influence through a Social Network” by Kempe / Kleinberg / Tardos.

1 Implementation

Let $G = (V, E)$ be the graph we want to analyze. Let $A \subseteq V$ be the output of the algorithm, the subset of nodes we will target. Let $k := |A|$ be the size of the desired initial active node set. Let $\sigma(A)$ be the expected number of nodes that are activated after a simulation with the cascade / linear threshold model, given the initial set A.

The algorithm starts with an empty set, and adds at each iteration a node to the set A : it choses greedily the node that, if added to A, increments the most our objective function $\sigma(A)$ in this step. Because we cannot evaluate the expected value analytically, we simulate the process *num_sim* times, and approximate the expected value with the average value.

The pseudocode of the algorithm is the following :

Algorithm 1 Greedy Algorithm

$A := \emptyset$
while $|A| < k$ **do**
  for all nodes $v \in V \setminus A$ **do**
    for num_sim iterations **do**
      Simulate IC / LT with initial nodes $A \cup v$
    end for
    Calculate an average activation score for v
  end for
  Determine the node v with the highest average score
  Set $A := A \cup v$
end while
return A

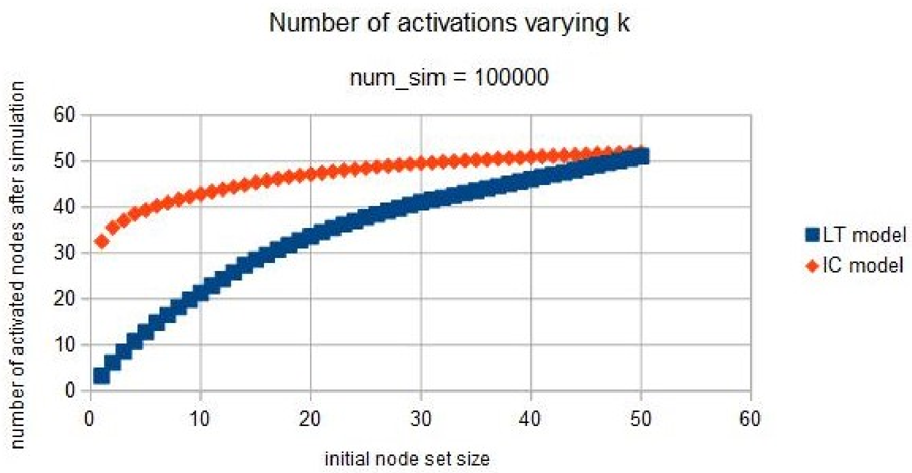
Obviously, the algorithm can produce different results whether used with the linear threshold model or with the independent cascade model.

2 Graph Used

We used an undirected graph of anonymized facebook data, with $|V| = 52$ and $|E| = 292$.

3 Results

We first observe how $\sigma(A)$ increases in both models when we allow for a bigger size k of the initial set. We use $num\_sim = 100000$ as in the paper.



We then want to analyze how imprecise the approximated value of $\sigma(A)$ is when *num_sim* is low. We find that on our graph, even with only 10 simulations, the estimation is quite near the value for 100000 simulations.