Programming Assignment 5

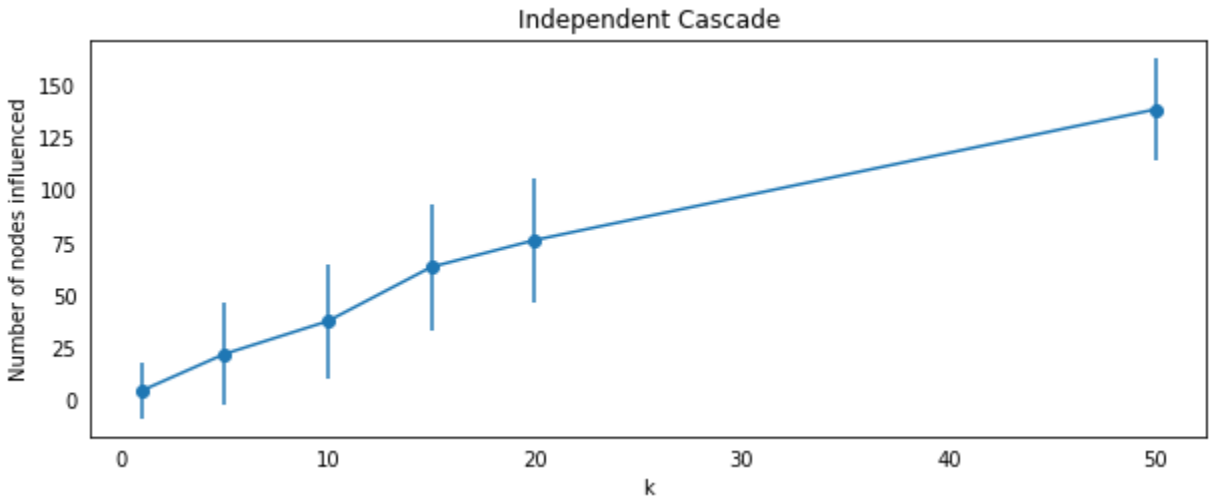
Vinit Ramesh Gore (M20CS064)

Graph

A Barabasi-Albert DiGraph with 1000 nodes and 9950 edges is created and used throughout the assignment.

Independent Cascade

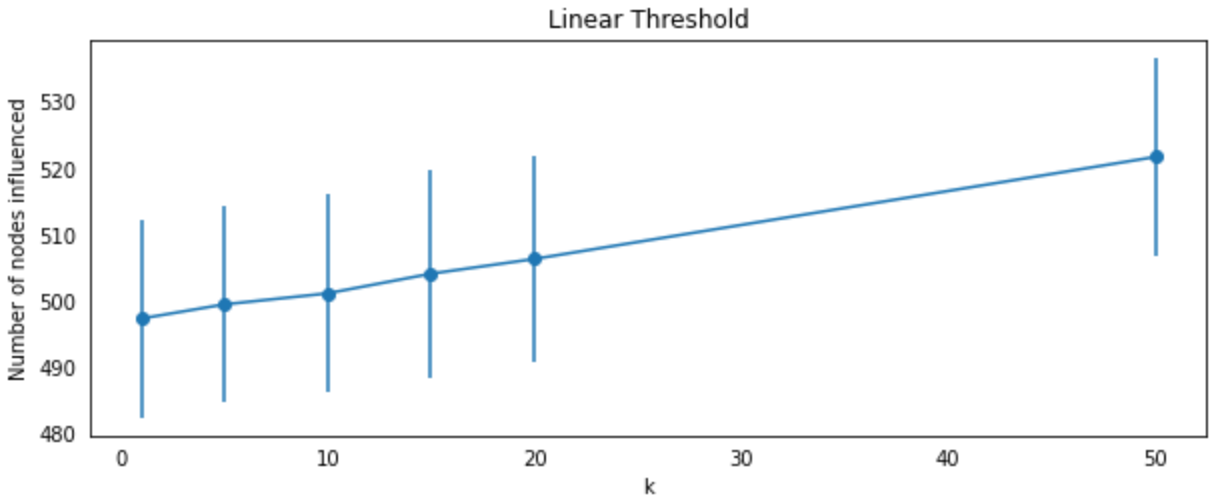
The Independent Cascade model is a stochastic model wherein a seeded node receives one chance to activate its neighbours with a certain probability. Here, we consider probability from the range [0.05, 0.15].
For different values of k and 1000 Monte Carlo simulations each, the number of activated nodes were recorded.



In the above plot, the points denote the average number over all the simulations and the vertical line denotes the standard deviation.

Linear Threshold

The Linear Threshold model is a more realistic model for information diffusion wherein every edge has some probability of activating the target node and every node has a threshold value for activation. A node is activated when the sum of probabilities coming from its seeded neighbours crosses the threshold. Here, we consider random values for edge probabilities and node threshold values.
For different values of k and 1000 Monte Carlo simulations each, the number of activated nodes were recorded.



In the above plot, the points denote the average number over all the simulations and the vertical line denotes the standard deviation.

Greedy Algorithm

The Greedy algorithm finds the set of most influential nodes with cardinality constraints. It provides best approximation guarantee of $1 - \frac{1}{e}$. However, the approach is very straightforward:

1. Start with empty set of seed nodes
2. Loop over 3, till S contains k nodes
3. For every node in the network, check if its marginal improvement is the maximum. If yes, add to S

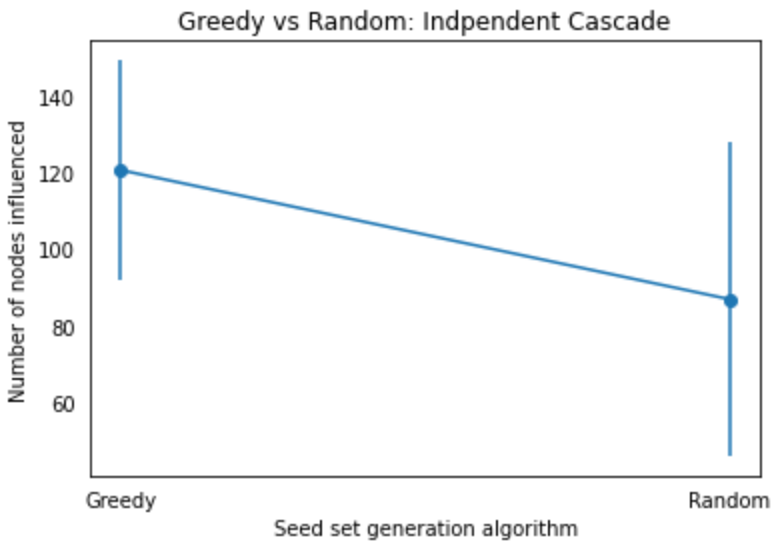
The time complexity of the Greedy algorithm is O(nk).

CELF++ Algorithm

The CELF++ algorithm overcomes the slowness of Greedy algorithm over real-life large networks by exploiting the submodular property of the function. It runs through all the nodes in the beginning to initialize node properties and store them in heap according to their marginal spread. Then a single loop is run to add the best nodes into S till it contains k nodes.

Greedy over ICM

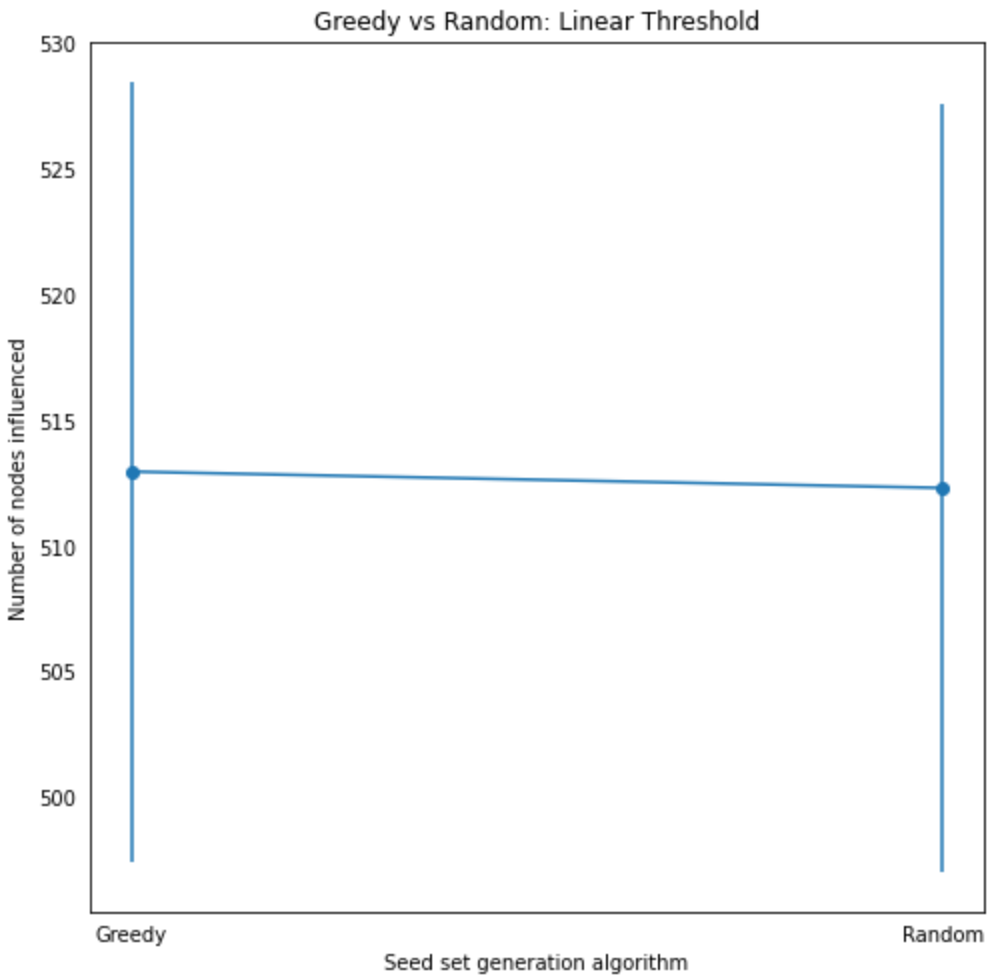
The CELF++ algorithm is used to run Greedy over ICM for 10 Monte Carlo (MC) simulations of IC diffusion. The seed set outputted by CELF++ is used to run on 1000 MC simulations for the ICM model. To compare with random seed sets, 10 random seed sets are run for the same setting on the ICM model and their mean and standard deviation is recorded.



The above plot shows the performance of the seed set generated by the Greedy (CELF++) algorithm over the aggregate of random seed sets over the ICM model.

Greedy over LTM

The CELF++ algorithm is used to run Greedy over LTM for 10 Monte Carlo (MC) simulations of LT diffusion. The seed set outputted by CELF++ is used to run on 1000 MC simulations for the LTM model. To compare with random seed sets, 10 random seed sets are run for the same setting on the LTM model and their mean and standard deviation is recorded.



The above plot shows the performance of the seed set generated by the Greedy (CELF++) algorithm over the aggregate of random seed sets over the LTM model.