hw3: Maximize influence

balilarder

May 2017

1 Description

My method to find the influential seed set is to compute and rank there **score**, and the size of the seed set is 30. By choosing the nodes with highest score as seed set, test the influence spread size.

Because the edges weight in the given network are same, which is the probability of propagation ($p = 0.05$) to infect its neighbors if a node is active in a iteration, we can ignore the features on the edges and focus on the nodes. The intuition of the method is to mix the effect of **degree count and betweenness centrality** with a parameter, w. The following is the formula to compute the score of a node:

$$score_{node} = w \times deg_{node} + (1 - w) \times betweenness_{node}$$

Both of them play an important role in finding seeds, with the intention of degree is obvious, which is that "the node with more edges has the more chances to try to spread out the information through the edges". In terms of betweeness, a centrality based on the shortest path, is to find the node that can reach a node from an another node within the minimum of path length, meaning the "hubs" or the "bridge" in the network, make sure the propagation between different groups.

2 Analysis

The experiment is to set w and observe the final influence spread size. We can see if we only use betweenness as score, the method performs the worst. Except of that, they performs almost equal and not bad. One possible reason is that we should normalize the degree count for the nodes to avoid bias, but we realize the effect of the degree in this method is indispensable. On the other hand, we expect the method will perform much better if we implement the spreading process. By doing so, we can try nodes one by one in iterations and ignore the seed list generated originally if it has been infected already, then try the next one. Or, also compute the marginal gain to the score. The running time is depend on the computation of the betweenness, so compute betweenness in advance can save much time.

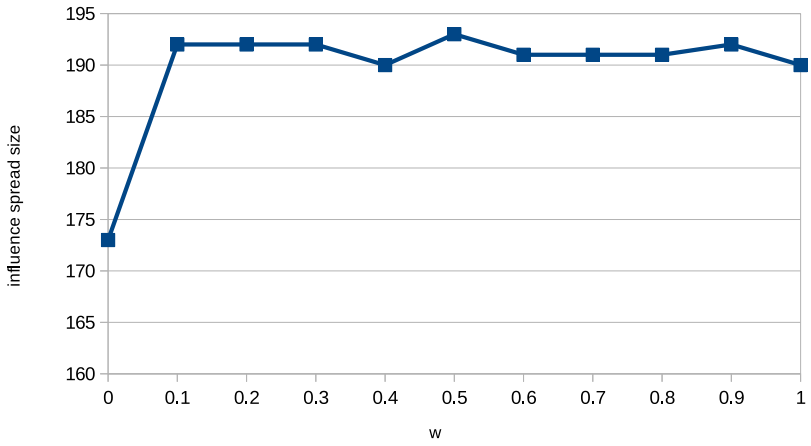

Figure 1: The relation between the influence spread size and the parameter, w