# Network Influence, Learning and Diffusion

## Introduction

People who live in society interact with one another. Through these interactions they influence others or are influenced themselves. Contrary to viral propagation, people have the choice of whether to be activated (i.e. to change their minds) following the interaction. Recently, several models have been created to study this type of network influence. In this short paper, I will discuss some of the main models and one potential business application.

## Methodology

In decision-based propagation models, there are two types of influence responses—diminishing returns and threshold—that determine the model to be used.

In diminishing returns, the idea is that the more the people around you are exhibiting a certain behavior, the more you will tend to converge to that behavior. The tendency is explained by social proof, coined by Robert Caldini, where people tend to follow the lead of people in their community. This tendency has diminishing returns however, signifying that the first people to talk to you influence you a lot but the influence decreases the more people you talk to.

The model used to model diminishing returns is the independent cascade model. This model starts with a set of active nodes. These active nodes try to influence all their neighbors and have a *p* probability of success. This value of *p* depends on a nodes susceptibility to be activated. This value can be uniform for all nodes or may depend on the individual. In general, it is believed that the value of *p* for a population follows a normal distribution bounded at 0 and 1. During the next period, the newly activated nodes try to activate their neighbors with the same *p* probability and this process continues until no more nodes are activated. The fact that having more neighbors around leads to higher chances of activation is because there are more neighbor nodes that will potentially be activated and that will try to convince the target node to activate at each period. This is reflected by this function: $P(n) = 1 - (1-p)^n$, where *P* is the total probability of being activated over all the periods. It is important to note that each node can activate its neighbors during only one period: the period after it was activated.

With threshold response, the idea is that a node will be activated if at least $\theta$ neighboring nodes are activated. Because the nodes have different thresholds, this process could result in a cascade where nodes with higher thresholds get activated because nodes with lower threshold came before and were activated thus reaching the higher threshold. Unlike the independent cascade model, this activation cascade *depends* on what happens in the network around the target node. If the threshold is less than $1-\rho$, where $\rho$ is the cluster density, the cascade propagation is limited to the cluster of the original activated node. This is because the higher the density is, the tighter the community and thus the more separated that community is from the outside world.

The model used to model the threshold decision is the linear threshold model. In this model, the weight of each of the neighbors on each of the nodes is calculated. If the sum of the weights of the activated neighbors is greater than the target node's threshold, then the target node is activated. This process is repeated with an updated weight matrix until no more activation is possible in the network.

It is important to note that, both models started with a set of already activated nodes. Depending on which nodes are activated, there will be a different cascade propagation. In some fields, such as marketing, professionals look for the smallest subset of people to spread information to the largest proportion of the network. As a result, there exists optimization methods to find the nodes that will have the highest effect on the network. Because finding the best subset is computationally heavy, a greedy approach is used as a type of forward selection method. A simulation is done with each node noting the number of nodes that it was capable of influencing. Then, only the node with the cascade propagation is kept and the process is repeated and added to the seed set until the desired seed size is reached. While this model might not always come up with the best seed subset, it comes up with one that is approaches the optimal solution and is significantly better than random selection. It is interesting to note that creating a subset using the top *n* nodes from centrality measures is also possible. This results in seed subsets that are significantly better than random selection but they are worse than using the greedy optimization approach specified above.

## Business Application

The idea of optimal seed selection for large cascade propagation is very relevant in the field of influencer marketing. Whenever a company markets a product, it looks for the cheapest and most effective ways to spread the information of the product's existence and convince the public to buy the product. As a result, companies are very interested in picking the right public figures that will be best able to influence their target market. It would thus be very interesting if a store such as Sephora could create a network of all their customers. Then, they could use the customer's shopping behavior to estimate each customer's probability to be convinced to buy a product. Finally, Sephora could use the greedy optimization method to see which customers are the most likely to reach the maximum amount of their customer base. To those people Sephora could send samples of a new product in order to spread the information and to generate some positive word-of-mouth advertising. This promotional method rests on Robert Caldini's concept of reciprocity; by getting a sample, the customer feels the need to do something good for the company in return, such as share positive opinions about it and to try to convince others to buy the new product.

## References

Stefano Nassini. Social Network Analysis. IESEG School of Business. Lille, 2019.

Leonid Zhukov. Structural Analysis and Visualization of Networks. National Research University Higher School of Economics. Moscow, 2015. Retrieved from http://www.leonidzhukov.net/hse/2015/networks/#module3