Final Project

John Carlsson Lukas Runt

***Abstract*—The aim of the project was to solve two tasks. The first to identify how a certain network was generated and the other to maximize the cumulative revenue obtained auctions over a marketing campaign. The results showed that the network was generated using a configuration model with the degrees beeing normaly distributed around 23 with a standard deviation of 5. To maximize the cumulative revenue bla bla and bla was used**

I. Introduction

This section introduces the tasks at hand.

- Task1:
  Consider the network N represented in the file net_x, where x is your group number. This network has been generated with one of the network models seen during the course. You have to analyze the network N with the network mining tools (the ones shortlisted in the midterm project) and guess which model has been used for creating it. Your guess has to be supported by an appropriate set of experiments to confirm that networks generated with the proposed model have characteristics similar to N (note that you have to guess also the parameters of the model).

- Task2:
  Your company is specialized in producing applications for social networks. You have been assigned the task of selling these applications with the goal of maximizing the revenue for the company. The contract between your company and the social network allows you to access to your preferred nodes over the network, and to spread whichever information you want from that node. However, the contract also constraints you to access to only one node every day. Then, your strategy is every day to select a node and spread from this node information about a network auction for selling k items (i.e., each day you start a different auction). Clearly, the information spreading process is not deterministic: in fact, even if a node invites a neighbor into the auction, it is not for sure that the latter will participate in the auction. We assume that the probability that a neighbor accepts to participate depends on the strength of the relationship among the two communicating nodes. Unfortunately, while you know the topology of the network, you do not know the strength of these relationships.

  Hence, you have to choose every day the node from which to start a social network auction in order to maximize the cumulative revenue obtained by all the auctions over the entire campaign. To this aim you are required at each day to both be able to reach agents with large valuation about the items you are selling and to selects a kind of auction in order to maximize the amount of money that you are able to collect from them. Notice that you can use different types of auctions in different days.

II. Task1

This section describes the methodology used in task 1.

A. Analysis

First an analysis was made of the given network to see the overall structure. This was made with the NetworkX library for python, where we quickly and easily can get some information about the graph.

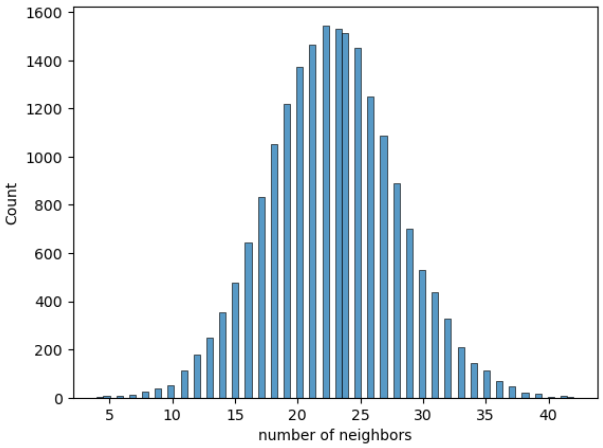
The network has the following characteristics:

Number of nodes	20 000
Number of edges	227739
Directed?	False
Connected?	True

We see that the graph is one giant component that is undirected.

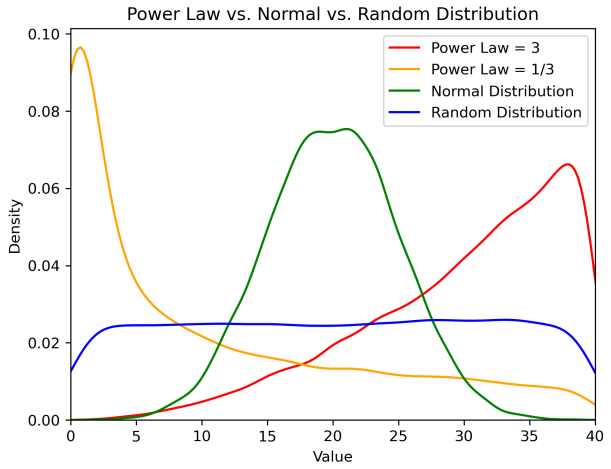
By looking at the distribution of neighbors we get the following graph.

Fig. 1. Number of neighbors



The number of neighbors that each node has gives us clues as to how the graph was constructed, if the degrees follow a distribution, we know that a configuration model was used to construct the graph. So we plot the given graph and compare it to some common distributions.

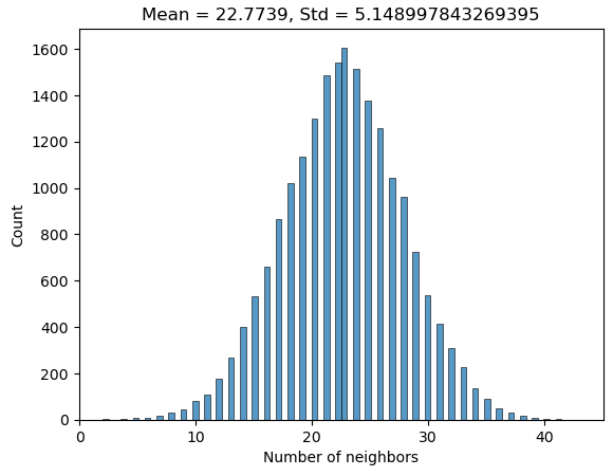
Fig. 2. Comparison of distributions



B. Experiments

Now that we know how the graph looks, can we replicate it? We from the prvious chapter that the graph strongly ressembles a normal distribution and we can then use a configuration model to create a similar graph. We can get the precise distribution and mean from the model which are 22.77 and 5.15. By creating one hundred lists of degrees based these parameters and taking the average and creating a graph with a configuration model, we get the following graph of neighbour distribution.

Fig. 3. Number of neighbors expirimental graph



They are evidently very similar.

Now looking at connectivity and if the graph is directed. We get the same results as for the original graph.

by comparing different measures of centrality, here the min and max values are shown, we can see that the graphs are indeed very similar.

Centrality	Original network	Expirimental network
Degree	0.2323	0.2323
Closeness	0.2323	0.2323
Betweenness	0.2323	0.2323

III. Task2

This section describes the methodology used in task 1.

A. Analysis

First an analysis was

B. Experiments

words
words

IV. Results Conclusion

This section presents the results of the project.

A. Task 1

Based on the previous chapter, we believe that the network was created by using a configuration model. The parameters for the model was the following:

Parameter	Value
Nodes	20000
Degree	$\sim \mathcal{N}(\mu, \sigma^2)$
μ	23
σ	5

Degree is a list containing the number of neighbors for each node.