Modelling and Analysis of Complex Networks

Exercise 8

Due: 13:00 on Nov. 29, 2024

The maximum score of this assignment is: **16 points**. Please submit the assignment in any readable data format (.txt, .doc, .pdf, .md ...) and submit the assignment before the deadline. If you have additional information concerning your answers, please also upload the document to the Moodle, or include the link to the document in your answers (e.g., link to your Github repository). Please indicate your team number in your submission.

The goal of this lab is to examine and experiment with several techniques related to influence analysis in complex networks. Influence is defined between two users and represents how possible is for a user to adapt the behavior or copy the action of another user.

We will first introduce the diffusion model which can be used to simulate a spreading process that takes place over the network. Subsequently, we will utilize the structure of the network to identify influencers based on a heuristic manner. Finally, we will examine the algorithmic solution to the problem of influence maximization, which is choosing the optimum nodes to maximize the spread of information, and compare it with the heuristic approach based on the simulations.

We will again rely on the network science dataset for our use case. The following questions can be answered with the help of networkx and NDlib. You may also use other packages to deal with the problem. Please answer the following questions and submit your executable code.

(a) Create a method to evaluate our chosen seeds, based on epidemic simulation. As a means for evaluation, we use the Independent Cascade model, to compute the number of influenced nodes during an influence spread over the network. This model assumes that a node v has only one chance to influence each of its neighbors u based on the probability ‘p v,u‘. We define the model's parameters (threshold=0.01) based on common approaches in the literature, and run the epidemic through 10 steps in order to get an approximation and retreive fast results (in normal circumstances it is 10 thousand).

(b) Compute K-core score: Given our undirected network G, C k is defined as the k-core subgraph of G if it is a maximal connected subgraph in which all nodes have degree at least k. Then, each node $v \in V$ has a core number c(v) = k, if it belongs to a k-core but not to a (k + 1)-core. The cohesion of subgraphs increases as k increases. Let us denote as C the set of nodes with the maximum core number k_max. Compute the top 20 nodes in terms of the k-core they belong to and simulate their spreading.

(c) Implement Greedy IM: Influence Maximization is the problem that lies at the heart of influence analysis and addresses how to find a set of nodes, such that if they start a diffusion, the number of infected nodes in the network (influenced spread) will be maximized. It has a broad range of applications, from viral marketing, which was the initial motivation for the problem, to epidemiological containment and political campaign management. The problem can be formulated as follows: given a social network, a diffusion model with some parameters and a number k, find a seed set S $\subset$ V of size k such that the influence spread is maximized. We will use the well-known method from Kempe et al. that is based on the fact that the function of the influence spread under the IC and LT models is monotone non-decreasing and submodular, which gives a $(1-1/e)$ approximation ratio to the optimal.

(d) Plot the spreading of k-core and greedy IM for a seed set of 20 to compare them.