Graph-based analysis of Influence Spread

Viktor Homolya

Abstract: The influence maximization is a well-known problem in network science. This problem is to target k nodes as seeds in a network G and maximize the spread of influence in this network. Lots of models have been created for this problem and to find relatively good results is easy. From the neighborhood of graph's nodes we can define local optima in the influence maximization. We aim to find connections between the structure of graph and the local optima for classifying the problem's difficulty for a given input graphs or to find properties that make easier the searching of nearly global optimal results in huge networks.

Keywords: influence spread, social network, local optima network

Introduction

The influence maximization is often used in viral marketing. In economy the problem can be interpreted as we have a limit for promoting (count of seeds) and we try to maximize the spread of information about our product. The graph represent the community. As a social network, the graph does not have multiple edges or loops. In our work the graph is directed (both way) for the different influences between people.

We can defined local optima by neighborhood from the graph. From local optima we build a network, the Local Optima Network (LON). With LON we map the codomain of original graph. We use some network science node properties (e.g. betweenness centrality, closeness centrality) and basic graph properties in special ways for the analysis.

This work made for find connections between the original graph and the created LON, which can be help in the influence maximization problem. It's not goal to optimize the problem by LON.

Influence maximization and models

The problem defined by Kempe et al. [1]. They showed the problem in the basic models (cascade models and threshold models) which they considered is NP-hard. We worked with two simple models.

Independent Cascade model (IC): every edge has a parameter p ($0 \leq p \leq 1$) representing the probability of spreading over the edge. In an iteration a freshly influenced node tries to spread to the (out)neighbors according to the edge parameter. In this model if a node v could not influence its neighbor w in first time it can't be anymore. If every influenced node can't make new influenced node(s) the model will stop.

Linear Threshold model (LT): every node and edge have a parameter t and w ($0 \leq t, w \leq 1$). The t parameter is the node's threshold value, w is the force of the influence via the edge. The sum of the incoming edges' w has to be less or equal than 1 for each node. A node v will be influenced in the next iteration if the sum of influenced incoming edges' w is bigger or equal than t value of v. The stop condition is the same as IC's.

In the work of Kempe et al. [1] the equivalence of IC models and LT models is proven. Hence, we can use only one of these models. We made tests in both models and the LT proved to be slower. So the bigger part of testing was made with IC models. From the stochastic nature of parameters a value of seed set ('influence value') came from average of influenced nodes' quantity with R times repeated evaluation.

Local Optima Network

We defined local optima from the neighborhoods of the nodes. A local optimum is a k sized subset of nodes like every solution in this problem. From these local optima created a graph, the Local Optima Network (LON). The LON approach was investigated for continuous optimization in [2], and for combinatorial optimization problems in [3]. Influence maximization belongs to the latter case, although we need to introduce some changes and new definitions. Every detected local optimum is a node in the

LON. Two nodes are connected in LON if the optimizer algorithm found them one after another. The LON is a directed, weighted graph depends on the optimizer. This optimizer has to take consideration the neighborhoods of nodes. We created a hill climbing algorithm for this task. This algorithm can traverse the edges of LON multiple times during the building phase, and this is the reason why LON is a weighted graph.

The stop condition of building is arbitrary. If we build LON for a long time we can perceive that some nodes' indegree is higher. We selected stop condition to find sufficient quantity of points with high sum of indegree. The indegree in LON characterizes the basins of a local optimum. We dropped the low weighted edges and the isolated nodes. With this filtering we created a connected graph. For analyzing the spreading in the original graph G via the LON and the filtered LON (H) we defined some measures like average of betweenness centrality of nodes from a seed set and neighbors of them in G.

In order to gain knowledge about the basins of the original problem we evaluated the closeness centrality (CC) values of nodes of H graph. The reciprocals of weights of edges were the distances in CC. With these CC values and the influence values we could classify the solvability a given problem to be difficult or easy. The high influence valued points represent nearly global optimal results. The low CC value means the point was hardly available for the hill climbing related algorithm. If we saw the most of the high influence valued points have high CC values we graded this problem to easy. Every created measure compare with influence value for find graph properties to make easier the optimization in huge networks.

Acknowledgements

I want to say gratitude to supervisor Tamás Vinkó for his helping. The project has been supported by the European Union, co-financed by the European Social Fund (EFOP-3.6.3-VEKOP-16-2017-00002).

References

[1] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2003.

[2] Tamás Vinkó and Kitti Gelle. Basin Hopping Networks of Continuous Global Optimization Problems. Central European Journal of Operations Research 25(2017) 985-1006

[3] Fabio Daolio, Marco Tomassini, Sébastien Vérel, Gabriela Ochoa Communities of Minima in Local Optima Networks of Combinatorial Spaces Physica A: Statistical Mechanics and its Applications, Elsevier, 2011, 390 (9), pp.1684 - 1694