9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Seed Selection for Linear Threshold Model in Multilayer Networks

Keywords: multilayer networks, linear threshold model, seeding, influence maximisation

Motivation

The problem of selecting an optimal seed set to maximise influence in networks has been a subject of intense research in recent years. However, while one can list many works concerning this issue, there is still a missing part to be tackled: multilayer networks. As has been proved in the literature, methods robust for networks with one layer are not easily applicable to their multilayer counterparts [1]. That narrows the usability of state-of-the-art research in real-case scenarios such as marketing campaigns, misinformation tracking, or epidemiology, where multilayer networks usually express real conditions better. In this work, we show the efficiency of various metrics, used to determine the initial seed set for multilayer Linear Threshold Model (LTM) [2].

Research subject and methodology

During the study, we evaluated seven different methods: random choice, greedy algorithm, degree centrality, neighbourhood size, as well as page-rank, vote-rank, and k-shell decomposition. Three latter were implemented in two, parallel ways: classical, i.e. from the perspective of node, and with the actor-wise approach so that, we were able to preserve consistency in results, regardless of the lack of widely recognisable extensions of their to multilayer networks.

The aforementioned metrics were investigated on LTM extended to multilayer networks by introducing so-called ”protocols”. These are aggregation functions determining whether to activate an actor by inputs from layers where it exists. We used two border cases: *OR*, where an actor gets activated if it receives sufficient influence on at least one layer and *AND*, where activation takes place only if the influence was sufficient on all layers. In addition, we took into account various ranges of node activation thresholds ($\mu = [0.1, 0.9]$) distributed linearly by 0.1. Seeding budgets ($s = [1\%,30\%]$) expressed as percentages of initially active actors, were used entirely in the 0th epoch of the experiment. We also decided to distribute activation thresholds homogeneously to limit the degrees of freedom in the experimental setup. As a result, for each seed selection method, about 3000 parameter combinations were obtained and evaluated.

In order to provide a sufficient diversity of propagation mediums, we executed experiments on three types of networks, which can be characterised by different properties: artificial scale-free, artificial Erdős–Rényi and real ones - in total ten graphs up to ca. 5000 edges and 10000 nodes. Finally, the robustness of methods was tracked by two main metrics - length of diffusion and gain achieved by the model. In tab. 1 we introduce definitions of these indicators.

Initial findings

At the moment of submitting this application, the works are still ongoing, however, some conclusions can be drawn. Fig. 1 shows the mean G obtained by LTM using given seed selection methods for two protocols.

Table 1: Indicators of diffusion efficiency taken into account.

Symbol	Indicator	Description
D	Length of Diffusion	Number of iterations passed in simulation until a steady state is reached
A	Active actors	Number of active actors in the network at the end of simulation as a percentage of all actors in the network
S	Seed actors	Number of actors selected to be seeds as a percentage of all actors in the network given a seeding budget value
G	Gain	Performance of the model given number of seed actors and a number of actors that could be activated: $G = \frac{A-S}{100-S}$

One can see that protocol makes a big difference in the model's performance - obviously, *OR* gives better results than *AND*. We can also note, that protocols are affected by the number of layers - for *AND* the performance of the model is better for shallower networks, while for *OR* this relation is the opposite.

Another conclusion is that there is no leading selection method which always provides the best results. Except for random and greedy algorithms (that stand out from other results), the mean performance of evaluated methods is levelled for protocol *OR*. What is interesting, the same phenomena cannot be observed for *AND* - EU Air Transportation network, which is the deepest network used in the evaluation (37 layers in total), disturbs the results. For this graph, the performance of metrics that have been implemented in two ways, i.e. node- and actor-wise (vote-rank, page-rank, and k-shell decomposition) differs in favour of the latter.

At the moment of submitting this abstract, we can also report a tab. 2. It shows an average place of seed selection methods in the ranking of obtained G from runs of the model with parameters from evaluated ranges with seeds selected by a given method. Currently, these ranking have been computed only for protocol *OR*, however, we consider them significant.

Here, the most important finding is that depending on the network type, the methods with the highest place in the ranking are different. For instance, selecting seeds with Degree Centrality in real networks results in high efficiency of spreading, and at the same time, applying this method to Erdős–Rényi graphs gives performance below average. Nonetheless, PageRank and VoteRank are usually the best choice and guarantee good results (which is consistent with fig. 1). What is more, when given methods are ranked relatively close to each other, one can consider a criterium of time complexity to choose the proper algorithm, so that fast yet akin in results routine is preferred (e.g. light Degree Centrality vs heavy PageRank).

Apart from the presented conclusions, works are still ongoing. One of awaiting tasks is obtaining rankings for protocol *AND*. We also need to process data regarding diffusion length and perform statistical analysis. These issues will be addressed in the final work.

References

[1] F. Erlandsson, P. Bródka, and A. Borg. Seed selection for information cascade in multilayer networks. In C. Cherifi, H. Cherifi, M. Karsai, and M. Musolesi, editors, *Complex Networks & Their Applications VI*, pages 426–436, Cham, 2018. Springer International Publishing.

[2] Y. D. Zhong, V. Srivastava, and N. E. Leonard. Influence spread in the heterogeneous multiplex linear threshold model. *IEEE Transactions on Control of Network Systems*, 9(3):1080–1091, 2022.

Figures

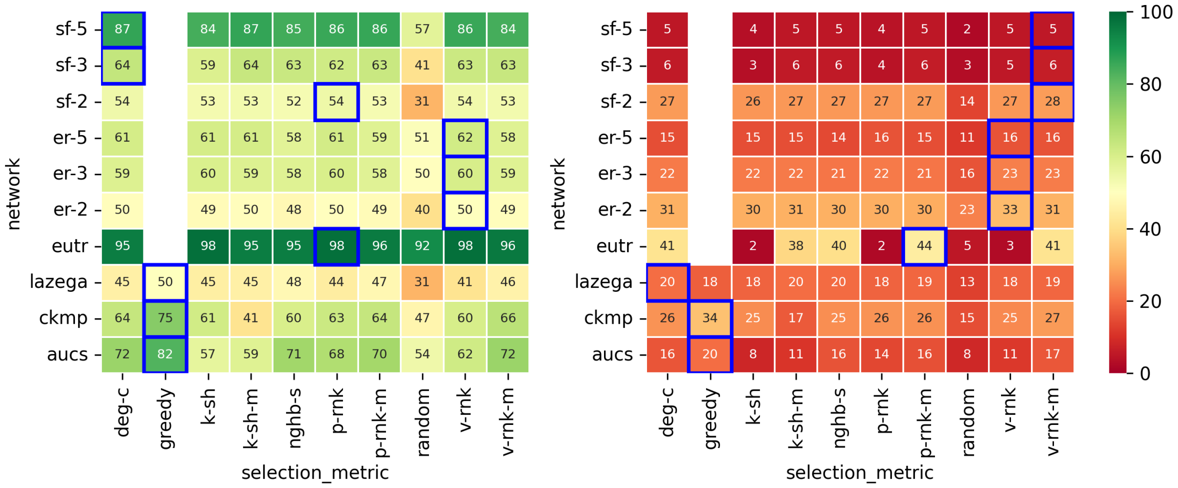

Figure 1: Heatmaps of mean G for evaluated seed selection methods. **Left:** protocol *OR*. **Right:** protocol *AND*. **Horizontal axes notation**: degree centrality (*deg-c*), greedy algorithm (*greedy*), K-shell decomposition (classic *k-sh*, actorwise *k-sh-m*), neighbourhood size (*nghb-s*), PageRank (classic *p-rnk*, actorwise *p-rnk-m*), random choice (*random*), VoteRank (classic *v-rnk*, actorwise *v-rnk-m*). **Vertical axes notation:** AUCS CS-AARHUS (*aucs*), Ckm Physicians Innovation network (*ckmp*), Lazega Law Firm (*lazega*), EU Air Transportation network (*eutr*), Erdős–Rényi network with n layers (*er-n*), scale-free network with n layers (*sf-n*).

Table 2: Mean positions of seed selection methods in the rankings of achieved G for respective runs of LTM with protocol *OR* and μ, s from evaluated ranges, grouped by network type.

Seed Selection Method	Networks			
	All	Erdős–Rényi	Scale Free	Real
Degree Centrality (deg-c)	2.93	3.20	**2.90**	**2.74**
Greedy	**1.03**	not computed		**1.03**
K-shell (k-sh)	3.19	3.19	3.69	2.83
K-shell (k-sh-m)	3.20	3.23	3.19	3.19
Neighbourhood Size (nghb-s)	3.29	3.43	3.67	2.89
PageRank (p-rnk)	**2.75**	**2.84**	**3.04**	**2.46**
PageRank (p-rnk-m)	3.06	**2.99**	3.55	**2.74**
Random	4.18	4.56	4.54	3.62
VoteRank (v-rnk)	**2.69**	**2.34**	**2.92**	2.79
VoteRank (v-rnk-m)	3.11	3.52	3.49	2.52