Network Diffusion — Framework to Simulate Spreading Processes in Complex Networks

Wrocław University of Science and Technology

Michał Czuba [1], Mateusz Nurek [1], Damian Serwata [1], Yu-Xuan Qi [2], Mingshan Jia [2], Katarzyna Musial [2], Radosław Michalski [1], Piotr Bródka [1]

[1]Wrocław University of Science and Technology
[2]University of Technology Sydney

Network Science Lab

Introduction

The problem of selecting an optimal seed set to maximise influence in networks has been a subject of intense research in recent years. However, there is still a missing part to be tackled: **multilayer networks**. Methods robust for one-layer-graphs are not easily applicable to their multilayer counterparts. That narrows their usability in real-case scenarios such as marketing campaigns, misinformation tracking, or epidemiology, where multilayer networks usually express actual conditions better. **In this work, we show the efficiency of various metrics used to determine the initial seed set for the Multilayer Linear Threshold Model** (MLTM).

Extending the LTM to multilayer networks

Linear Threshold Model in its initial form [1] cannot be directly applied to multilayer networks — **actors are the subject of the process, while the nodes are their auxiliary representation**... Therefore, we need to define:

- what does it mean that an actor is (or is not) active,
- how does it relate to diffusion dynamics taking place within layers, where it is represented.

In our research, we used the approach proposed by [2] with amendments so that a homogeneity among actors has been imposed in the sense of an activation threshold (μ) and a protocol (v.i.).

Protocol function in MLTM

According to [2], state of the actor n of a multilayer network $M = (N, L, V, E)$ in the time step t is determined by a following function:

$$x_n(t) = \begin{cases} 1, & \text{if } y_n(t) \geq \delta \text{ or } x_n(t-1) = 1 \\ 0, & \text{otherwise} \end{cases}$$

Where:

δ - a parameter of the model, $\delta \in [\frac{1}{|L|}, 1]$,
$y_n(t)$ - a mean input of actor n (represented in K layers) in time t, $y_n(t) = |K|^{-1} \sum_{k \in K} y_v^k(t)$.
$y_v^k(t)$ - an impulse of node v from layer k in time t, $y_v^k(t) \in \{0, 1\}$

Toy example

We decided to examine two extreme cases: $\delta = 1$ (AND) and delta $= \frac{1}{L}$ (OR). In the former one, an actor gets activated if it receives sufficient influence on all layers where it is represented, and conversely the latter, where sufficient input in at least one layer is enough for activation.

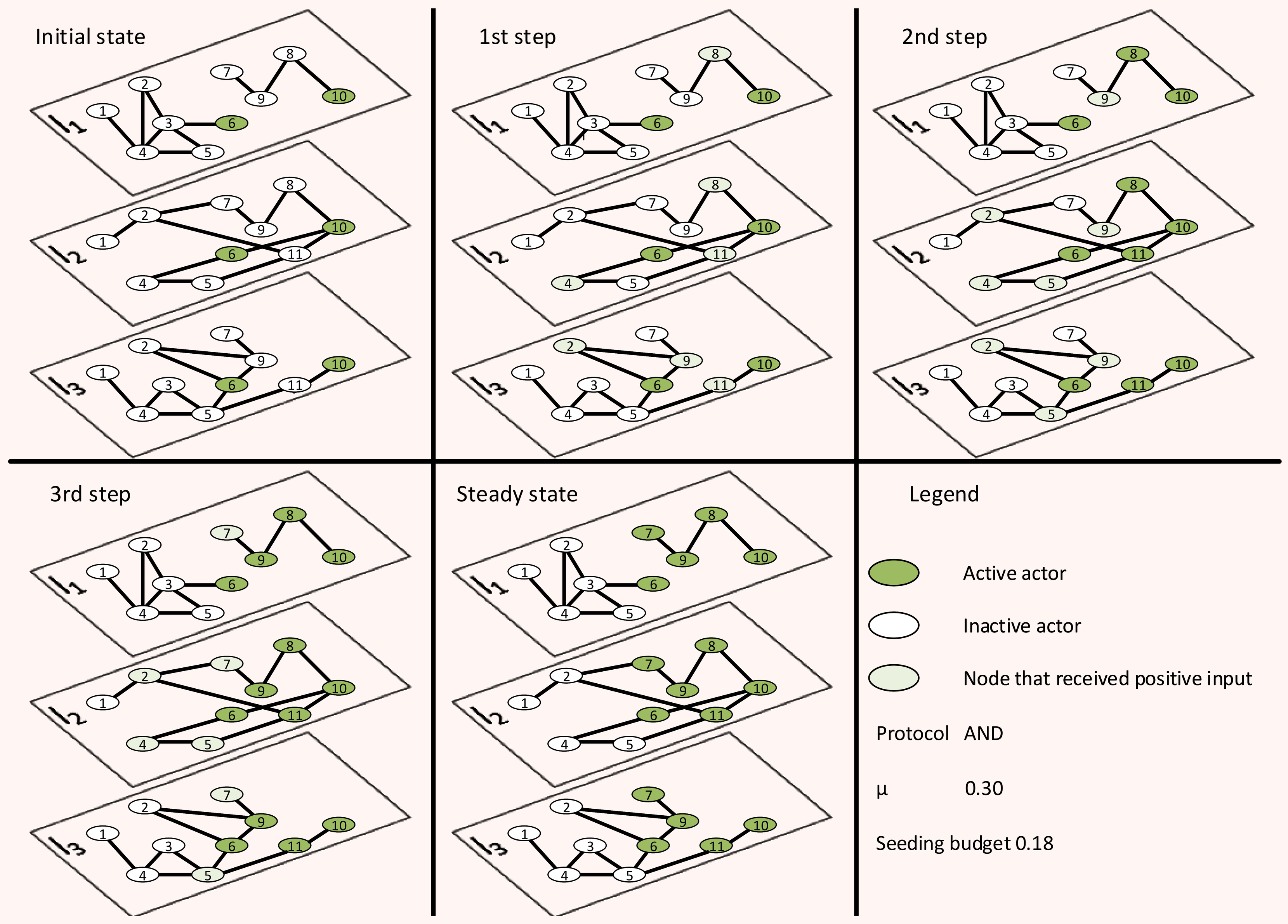

Figure 1. Example of spreading of MLTM in toy network with protocol AND.

A problem we tackled

Budget constrained influence maximisation

Let $\mathcal{S}$ be a family of sets of cardinality s over actors of the multilayer network M in the sense that $\mathcal{S} \subseteq powerset(N)$. Let $\sigma : \mathcal{S} \rightarrow \mathbb{R}^+$ b an arbitrary function that maps a set of actors used as a seed set to a number denoting the expected size of activated actors in a binary discrete system. An influence maximisation problem for seeding budget of size s is an issue of finding $S_0 : \arg\max(\sigma) = S_0 \wedge |S_0| \leq s$.

Measuring an efficiency of the diffusion

We used **Gain** metric to assess a performance of the spreading model that bases on a number of seeds, a number of actors that could be activated, and a number of active actors when diffusion faded down:

$$G = 100 \cdot \frac{|S_D - S_0|}{|N - S_0|}$$

Seed selection methods

During the study, we evaluated the following methods to select seed set for MLTM:

- degree (*deg-c*),
- neighbourhood size (*nghb-1s*, *nghb-2s*),
- PageRank (*p-rnk*, *p-rnk-m*),
- VoteRank (*v-rnk*, *v-rnk-m*),
- k-shell decomposition (*k-sh*, *k-sh-m*),
- random choice (*random*),
- greedy (*greedy*).

We used two ideas to adapt PageRank, VoteRank, and k-shell to multilayer networks: compute the

Used data and parameter space

(a)

Name	Layers	Actors	Nodes	Edges	Note
aucs	5	61	224	620	AUCS CS-AARHUS
ckmp	3	241	674	1370	Coleman, Katz, Menzel Innovation Among Physicians
eutr-A	37	417	2034	3588	The European Air Transport Network
lazega	3	71	212	1659	Lazega Law Firm
er-2	2	1000	2000	5459	Erdős–Rényi networks generated with multinet library
er-3	3	1000	3000	7136	
er-5	5	1000	5000	15109	
sf-2	2	1000	2000	4223	Scale-free networks generated with multinet library
sf-3	3	1000	3000	5010	
sf-5	5	1000	5000	10181	

(b)

Protocol	OR	AND
s range	$[1, 2, ..., 10, 15, 20, 25, 30]$	$[15, 20, 25, 30, 31, ..., 40]$
μ range	$[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]$	
networks	real (aucs, ckmp, lazega, eutr-A), Erdős–Rényi (er-2, er-3, er-5), Scale-free (sf-2, sf-3, sf-5)	
s. s. method	deg-c, greedy, k-sh, k-sh-m, nghb-1s, nghb-2s, p-rnk, p-rnk-m, random, v-rnk, v-rnk-m	

Table 1. Networks used in experiments with their basic parameters shortlisted (tab. 1a), values of each evaluated parameter - in total 27,720 executed experiments (tab. 1b).

Initial results

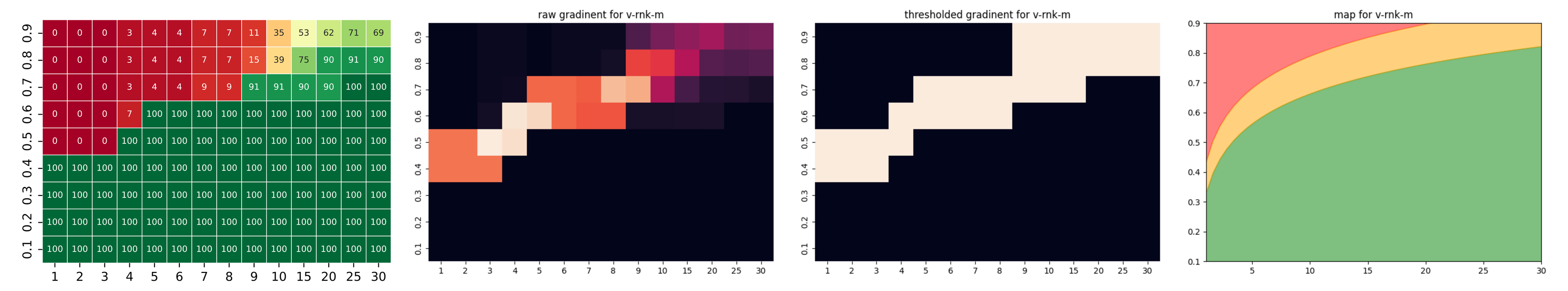

Figure 2. Example of three zones on G heatmap for the *v-rnk-m* method on *aucs* network with protocol *OR*. (1) a heatmap of the raw gain, (2) computed gradient magnitude, (3) thresholded gradient that divides the area into three sections, and (4) the final division of the $\mu \times s \times G$ space.

We observed:

- existence of three areas on the heatmaps ("robust", "transitional", and "unrobust"),
- that length of the diffusion is highest in the "transitional" zone,
- for Erdős-Rényi networks the transitional zone is the narrowest,
- for Scale-Free networks transitional zone is typically present,
- for real networks the most significant differences between the $\mu \times s \times G$ heatmaps.

Coarse assessment of seed selection methods

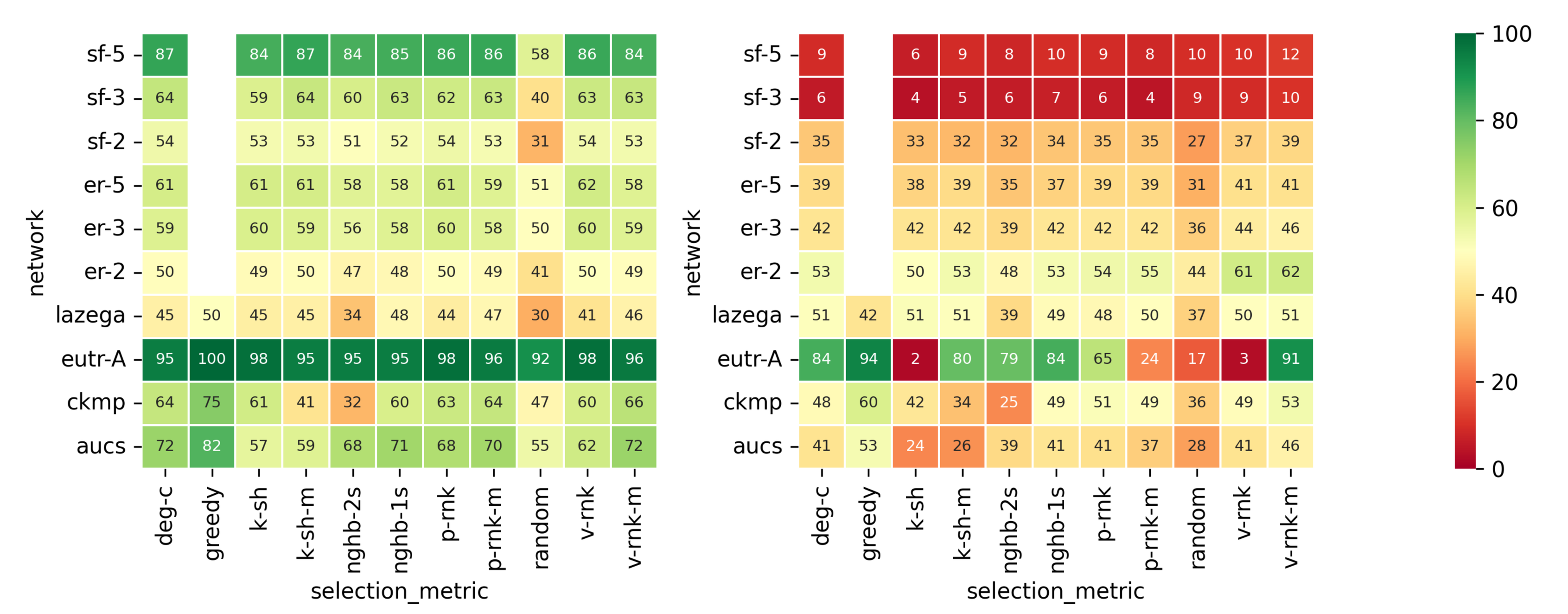

Figure 3. Heatmaps of mean G achieved by evaluated seed selection methods on multilayer networks. **Left:** protocol OR. **Right:** protocol AND.

We observed:

- network type impacts the G, surpassing the influence of seed selection methods,
- $random$ and $greedy$ stand out of the another methods.
- network topology strongly affects efficiency of MLTM diffusion (more layers lead to better G for OR, whereas worse G for AND),
- most of the evaluated methods behave similarly regarding the network for which they are applied.

Rankings of seed selection methods

(a)

Seed Selection Method	Networks			
	Real	Erdős–Rényi	Scale-free	All
deg-c	3.12	4.57	5.55	4.28
greedy	**1.33**	not computed		
k-sh	5.46	4.82	7.95	6.02
k-sh-m	4.58	4.63	6.31	5.11
nghb-2s	4.14	5.44	6.53	5.25
nghb-1s	2.91	4.22	5.18	3.99
p-rnk	3.05	3.60	5.02	3.81
p-rnk-m	3.87	3.94	6.51	4.68
random	5.23	5.42	6.74	5.74
v-rnk	3.93	3.08	4.36	3.80
v-rnk-m	2.11	**3.04**	**2.85**	**2.61**

(b)

Seed Selection Method	Networks			
	Real	Erdős–Rényi	Scale-free	All
deg-c	2.57	3.20	**2.90**	2.86
greedy	**1.03**	not computed		
k-sh	2.67	3.20	3.69	3.13
k-sh-m	3.03	3.23	3.18	3.13
nghb-2s	3.31	3.93	4.26	3.78
nghb-1s	2.72	3.43	3.67	3.22
p-rnk	2.31	2.83	3.04	2.69
p-rnk-m	2.57	2.98	3.54	2.99
random	3.44	4.55	4.54	4.10
v-rnk	2.62	**2.34**	2.92	**2.62**
v-rnk-m	2.35	3.53	3.49	3.05

Table 2. Rankings of seed selection methods with respect to the achieved G for protocols OR (tab. 2a) and AND tab. 2b) grouped by graph type (the higher number the better).

Conclusions

- we confirmed that efficiency of centrality based heuristics is located between *random* and *greedy*,
- we have proposed an extension of VoteRank, PageRank and k-shell to multilayer networks,
- for both protocols VoteRank is the most efficient method,
- in real case scenarios, it could be better to make the μ smaller (i.e. invest in a mass marketing) than to put efforts in selecting the optimal S_0 (i.e. conduct the campaign with influencers).

References

[1] David Kempe, Jon Kleinberg, and Éva Tardos. "Maximizing the Spread of Influence Through a