Sequential seeding in multilayer networks

Piotr Bródka, Jarosław Jankowski, and Radosław Michalski
1) Wrocław University of Science and Technology, Department of Computational Intelligence, Wybrzeże Wyspiańskiego 27, Wrocław, 50-370, Poland
2) West Pomeranian University of Technology, Faculty of Computer Science and Information Technology, Zólneńska 49, Szczecin, 71-210, Poland

(Dated: 14 September 2020)

Complex networks are the underlying structures of multiple real-world systems: social, biological, computer, or communication, to name only a few. In many cases, they are helpful in modelling processes that happen on top of them, which leads to gaining more knowledge about these phenomena. One example of such a process is the spread of influence. Here, the members of a social system spread the influence across the network by contacting each other, sharing opinions or ideas, or - explicitly - by persuasion. Due to the importance of this process, researchers investigate which members of a social network should be chosen as initiators of influence spread in order to maximise the effect. In this work, we follow this direction, develop and evaluate sequential seeding technique for multilayer networks. The results show that sequential seeding outperforms traditional approach by increasing the coverage and allowing to save seeding budget. However, it also extends the duration of the spreading process.

Sequential seeding is a node activations strategy for influence spreading which distributes activations over time instead of performing all of them at once. It proved its superiority in a majority of seeding scenarios. However, the research until now was limited to simple one layer static networks reflecting only one type of relation.

In this paper, we address that gap by extending sequential seeding to multilayer network scenario. We have performed extensive evaluation using four real networks and six synthetic (three random networks and three multilayer networks) of various sizes and with a various number of layers.

Our results show that sequential seeding outperforms the traditional approach by increasing the coverage and increasing the duration of the spread, confirming findings from previous research for one layer networks. What is more, we have evaluated additional aspect of sequential seeding, namely the savings in the seeding budget. This aspect has not been evaluated in previous research.

The findings presented in this work allow redesigning influence spreading strategies to increase the coverage with limited seeding budget allowing for more effective spreading in multilayer networks.

I. INTRODUCTION

The influence maximisation problem challenges the researchers for more than fifteen years. In its basic formulation, one needs to find a set of nodes in a complex network that activates the maximum number of nodes for a given influence spreading model. These nodes, usually called a seed set, are activated at the beginning of the process, and throughout iterations, they spread the influence across the network. Unfortunately, the discovery of the best seed set equals a complete and in-depth evaluation of spreading capabilities of nodes which is extremely hard and time-consuming task. Thus, multiple heuristics have been proposed which allow to find maybe not the best but good enough seed set. In a simplistic scenario, two primary factors contribute to the problem: the network topology and the influence model. However, when considering more realistic situations, additional factors should be taken into account, such as varying cost of acquiring/activating seed nodes or more complex network structures like temporal or multilayer networks.

Apart from that, recent research demonstrated that for static networks, the concurrent seed nodes activation is superseded by sequential activation of nodes from the seed set. The roots of this approach, called sequential seeding, come from decision making, where information about the consequences of prior decisions should be gathered before making the next one. In sequential seeding, that information relates to the observation of how influence cascades spread in the network, and before selecting seed nodes for next activations, the knowledge on the current state of the process is taken into account. This method contrasts single stage seeding with respect to how the activations are distributed, but what is worth underlining, sequential seeding is actually a meta-method, since it relates to the way how to activate the nodes. Still, the ordering of nodes to be activated can be generated by any heuristics, depending on the time required to generate the seed set. For the independent cascades model and static networks, sequential seeding demonstrated its superiority over single stage activation.

In this work, our goal is to extend the knowledge about the performance of the sequential seeding method by adding another degree of complexity: we are investigating sequential seeding in multilayer network. Multilayer networks are often typical abstraction of the way how we interact with each other or how other complex systems work. For instance, in the case of human interactions, each layer in these networks can represent different means of communication like face meetings, phone calls, text messages, e-mails or WhatsApp communication. Each of these layers can have different properties,
II. METHODOLOGY

In this section we will briefly introduce all methods and techniques we have used in our research.

A. Multilayer network

In the literature, many different definitions of multilayer networks exist\cite{1,2,3,4,5,6,7,8,9,10,11,12,13}, however, in this work, we use the definition of the multilayer network similar to proposed in\cite{13}.

Multilayer network is defined as quadruple $M = (N, L, V, E)$, where

- $N$ is a set of actors,
- $L$ is a set of layers,
- $V$ and a set of nodes, $V \subseteq N \times L$,
- $E$ is a set of edges $(v_1, v_2) : v_1, v_2 \in V$, and if $v_1 = (n_1, l_1)$ and $v_2 = (n_2, l_2) \in E$ then $l_1 = l_2$.

An example of a multilayer network is presented in Fig. 1, where $L = \{l_1, l_2, l_3\}$, $N = \{1, 2, \ldots, 11\}$, and $((1, l_2), (2, l_2))$ is an example of an edge in $E$.

B. Independent Cascade model

To simulate influence spreading in network, we have used Independent Cascade model (ICM)\cite{13}. In this model, each activated node has a one chance to activate its neighbours with a defined propagation probability (PP). If a node activates its neighbours, they will have the chance to do the same in the next iteration of the process. Since the initial version of this model has been proposed for simple one layer networks, we had to adjust it to multilayer scenario, similarly to\cite{11}.

In a multilayer network, each newly activated actor will attempt to activate all its neighbours on each layer independently. This reflects the situation where for example we work (first layer) and play football (second layer) with someone. In consequence we have more chances to influence that person due to multiple channels of interaction.

Additionally, if the actor is activated on one layer it becomes active on all of them, i.e. it does not matter if someone convinced us at work or during football practice, we will have the same opinion everywhere. Apart from these two changes, the multilayer ICM works the same as the original, one layer, model.

C. Seed selection strategy

As mentioned in the previous section, there is a number of seed selection strategies, and since evaluating all of them is not the aim of our research, we have decided to select three simple and most commonly used seed selection strategies to observe if they have any impact on our results.

The first one is a multilayer degree centrality i.e. the number of edges adjacent to each actor on all layers. The second one is the multilayer neighbourhood size, i.e. the number of all distinct actors each actor is linked to in all layers\cite{13}.

---

**FIG. 1. Toy example of a multilayer network**

| Actor | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|-------|---|---|---|---|---|---|---|---|---|----|----|
| Degree Centrality | 3 | 7 | 6 | 9 | 8 | 6 | 4 | 7 | 5 | 5 |
| Neighbourhood Size | 2 | 7 | 4 | 5 | 4 | 6 | 2 | 2 | 4 | 3 | 3 |

**TABLE I. The values of degree centrality and neighbourhood size measures for each actor for exemplary multilayer network (Fig. 1)**
Both measures may seem to be similar. However, they both can yield very different results in terms of actors importance ranking. For example, for our toy network presented in Fig. 1 we can see that the most important actors according to degree centrality are actors 4 and 5, while if we use the neighbourhood size, actors 7 and 6 are the most important (Table I). The third seed selection strategy was random seed selection.

Using each strategy for each network, we have created the ranking of actors with the most important on top and least important on the bottom. If two actors had the same value of measure they were ordered according to actor id, e.g., for our toy example actors 3 and 6 have the same value of degree centrality, but in the final ranking actor 3 would be higher than actor 6. Next, the rankings were saved, and during simulations, the same rankings have been used regardless of other simulation parameters. This was especially important in case of random seed selection.

D. Single stage seeding

Single stage seeding (SS) is a traditional approach to spreading initiation in one layer and multilayer network.[11, 12]

In this technique, before we initiate the process we create a seed set consisting of one or more actors (or simply nodes in one layer networks) selected using some heuristic e.g., actors with the highest degree centrality, k-shell centrality, page rank, neighbourhood size, VoteRank or by simply selecting them at random.

When having a seed set defined, in single stage seeding approach, we activate all actors in the set at once (in a single stage) and allow them to influence other actors in the network. At this point, we do not have any additional control over the spreading and simply wait until it ends i.e., there are no more actors that can be activated. The toy example of this process is presented in Fig. 2.

E. Sequential seeding

Most of earlier influence maximisation research was based on the selection of all seeds at the beginning of the process, without additional actions taken after the process is launched. It was a different assumption than in real information spreading processes where various actions are taken during the process to improve its performance. For instance, additional marketing budget is used, or new content revealed.

The example of such a process would be a The Grand Tour show, where, Amazon, instead of realising all episodes at once (traditional approach for VOD platforms like Netflix or Amazon Prime Video) decided to release episodes weekly in order to increase the number of sold Amazon Prime memberships.

Recently several attempts were taken to model additional seeding actions during spreading processes in the form of sequential seeding, adaptive seeding and seeding scheduling. Sequential seeding research used the highest decomposition of the problem, starting from single seed per stage (i.e., on seed per spreading model iteration) to evaluate the effect of spreading the seeding process over time.

Several strategies were analysed. The first one was unconditional seeding which used a small number of seeds at the beginning of the process and then in each stage of simulation, additional seed or seeds were used. Another approach within the same study was based on revival mode. In this extension, instead of adding a new seed during each iteration we add new seed only when activation cascade stops, thus we revive the spreading process. Experiments showed that using all seeds at the beginning is not the best strategy. A large fraction of nodes activated during the seeding stage can be activated in the natural spreading progress by their neighbours. Saved seeding resources can be used to activate nodes difficult to reach with natural spreading process, for example in isolated network segments like in our toy examples (Fig. 2 and Fig. 3).

The performance of sequential seeding was analysed for most often used heuristics like degree-based selection or greedy approach with better results for all seed selection strategies. Following study[3] also verified the performance of seed selection with the use of VoteRank method and top-k strategy for influence maximisation for networks with community structure. Another studies analysed the potential of seed selection methods based on entropy centralities and the role of network typologies.

Unfortunately, all previous research was limited to simple one layer networks. Thus, we have decided to extend previous work on sequential seeding to multilayer scenario.

Two approaches have been adapted. The first one is a classic sequential seeding (SQ). The toy example of this process is presented in Fig. 3. For SQ, first we select seeds the same way as for single stage seeding (SS) process, but instead activating all of them at once we add one seed in each stage (one seed per iteration of ICM process) taking as a seed the node which is the highest-ranked not activated node on our ranking list. We add seeds until we consume whole seeding budget or there is no one else to activate (i.e., all nodes in the network are already active). As it can be seen in presented toy examples for single stage (Fig. 2) and sequential (Fig. 3) seeding, using the second approach allowed us to activate additional section of the network, increasing the final coverage of the process.

The second approach was sequential seeding with revival (SQr) where instead of adding one seed in every stage we wait until the spreading stops and only then add new seed to revive the process.

F. Coordinated execution

Coordinated execution principle was initially designed for one layer networks and introduced in [11]. In our paper we adjust it to the multilayer scenario. It allows us to evaluate and compare different seed activation strategies using Independent Cascade model despite the fact that ICM is not a deterministic model.

In coordinated execution approach instead of running ICM and drawing for each active actor if it can activate its neighbour or not, we preselect the edges which can transmit the
FIG. 2. A toy example of single stage seeding in multilayer network (toy network from Fig. 1). Before we start spreading we calculate degree centrality, rank actors according to this measure and select two actors with the highest degree centrality, i.e., actor 4 and 5, as seeds (see Table I). Next we activate both of them and start spreading process simulated using Independent Cascade model. It finishes after three iterations with 7 actors activated (63.6% of network).

influence. To be more specific, for each network we create a number of instances of this network where for each edge, based on ICM propagation probability, we assign a binary choice independently for \( A \) to \( B \) and \( B \) to \( A \) telling us if \( A \) can activate \( B \) and vice-versa.

Using this approach for each network instance, we can easily compare the results for single stage seeding and sequential seeding since they are not influenced by drawing results during ICM spreading. In other words, spreading path will always be the same, e.g., if an actor \( A \) is activated, it will always
SEQ

ACT

FAC

FIG. 3. A toy example of sequential seeding in multilayer network (toy network from Fig. 1). Before we start spreading we calculate degree centrality, rank actors according to this measure and select actor with the highest degree centrality (Table I), i.e., actor 4 as the first seed. Next, we activate actor 4 and start ICM spreading process. The pattern of activations is exactly the same as in single stage seeding example (Fig. 2). However, due to the fact that we have used only one seed (half of our seeding budget) to start the spreading we still can add one more seed in the second stage (second iteration). We select the first not active actor on degree ranking (i.e. the actor with the highest degree from not active actors) as our second seed actor (Table I). Since actors 4, 5, and 2 are already active, we select actor 9. Thanks to that we are able to activate additional part of network resulting in 11 activated actors (100% of network) after four iterations.

activate B and never C regardless of seed activation strategy.

III. EXPERIMENT SETUP

The experiments have been performed using the multinet library and ten multilayer networks (table II). Following the coordinated execution principle one hundred instances of each network for each propagation probability (PP) were generated by assigning binary choices of propagation or not for each edge, independently for A to B and B to A activation. This resulted in 9,000 (10 × 9 × 100) network instances. For each network instance we have run 36 (4 × 3 × 3) simulations for all possible parameters combinations with four different seed counts, three seed selection strategies and finally, three seed activation strategies (all parameters are described in table III). In total there were 324,000 different simulation cases.
Sequential seeding in multilayer networks

| Name | Layers | Actors | Nodes | Edges | Description |
|------|--------|--------|-------|-------|-------------|
| N1   | 5      | 61     | 224   | 620   | AUCS CS-AARHUSS |
| N2   | 3      | 241    | 674   | 1370  | Ckm Physicians Innovatort |
| N3   | 37     | 417    | 2034  | 3588  | EU Air Transportationt |
| N4   | 7      | 3    | 212   | 1659  | Lazega Law Firmt |

TABLE II. Ten networks used in experiments, their parameters and short description.

| Parameter | Values | Description |
|-----------|--------|-------------|
| N - Network | N1 - N10 | Four real networks (N1-N4) and six artificial (N4-N10). For details please see table III |
| PP - Propagation probability | 0.01, 0.02, 0.03, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50 | Nine different values of propagation probability for Independent Cascade model |
| SC - Seed count | 0.02, 0.05, 0.10 | The percentage of the actors in the network selected as seeds |
| 3S - Seed selection strategy | Degree centrality, neighbourhood size, random | Three seed selection strategies |
| SA - Seed activation strategy | SS, SQ, SQr | Three seed distribution strategies: SS - single stage seeding, SQ - sequential seeding, SQr - sequential seeding with revival |

TABLE III. Values for each parameter evaluated during experiments.

IV. RESULTS

In the figure [4] the comparison between 108,000 cases of single stage (SS) and sequential (SQ) seeding strategies is presented. The results have been sorted by the SS percentage of activated actors. Blue colour indicates the increase in the coverage in the percentage of all actors in the network i.e. how much more actors have been activated by SQ in comparison to SS. Orange colour indicates how big percentage of seeds have not been used after activating all actors in the network (i.e. the budget saved by SQ strategy).

Sequential seeding always achieves at least the same results as SS, and in 74% of cases, SQ performs better than SS. What is more in 31.75% of cases (including 43.63% of cases where SQ yields the same results as SS) SQ do not have to use all seeds to activate all actors in the network. This shows the new benefit of the sequential seeding approach, i.e. by observing the process as it progresses, wiser decisions we can save some portion of the budget needed to activate seeds. This feature might be crucial if one is doing an advertisement campaign, and each seed is an actual cost for the company. What is more, previous papers investigating sequential seeding in simple one layer networks looked only at the increase in the coverage or interplay between coverage in time, but none of them evaluated seeds savings.

On average SQ was able to activate 1.074 times more nodes than SS, with minimum one (i.e. the same number) and maximum nine times more. Regarding the saved seeds, on average SQ saved 21% of seeds with 0% as minimum and 95.5% as maximum.

Similar but not so overwhelming results can be noticed when we compare SQ with sequential seed with revival (SQr - see section II E). It is always at least as good as SQ, in 44.21% of cases it is better than SQ, and in 1.32% of cases SQr was able to activate all nodes in the network with lower number of seeds than SQ (Fig. 5).

However, the higher coverage and savings of SQ and SQr comes with the price. Similarly to the observations presented in [3], also in case of multilayer networks using SQ and SQr seed activation strategy results in much longer spreading process. On average, for SQ it takes nine times and for SQr 12.9 times more iterations of the ICM to finish the spreading process (Fig. 5).

In the following sections, we will take a more in-depth look at each of those three main elements, i.e. spread coverage, spread duration and saved seeds, in the context of parameters we have used in the experiments.

| Parameter | Spread coverage | Saved seeds | Spread duration |
|-----------|----------------|-------------|-----------------|
| Network   | SS SQ SQr       | SS SQ SQr   | SS SQ SQr       |
| N1        | 71% 73% 75%     | 13% 25% 15% | 151% 216%       |
| N2        | 58% 62% 65%     | 15% 20% 400%| 74% 72%         |
| N3        | 88% 89% 90%     | 27% 40% 131%| 172%           |
| N4        | 61% 67% 68%     | 6% 6% 751%  | 973%           |
| N5        | 63% 65% 65%     | 32% 36% 1048%| 1673%         |
| N6        | 67% 70% 70%     | 37% 41% 882%  | 1446%         |
| N7        | 81% 84% 84%     | 51% 56% 570%  | 953%           |
| N8        | 51% 57% 57%     | 0% 0% 2108%  | 2719%         |
| N9        | 56% 62% 63%     | 3% 3% 1903%  | 2503%         |
| N10       | 73% 77% 77%     | 27% 29% 1070%| 1526%         |

| Seed Selection Strategy | Degree | Neighbourhood | Random |
|-------------------------|--------|---------------|--------|
| 67% 71% 72%             | 21% 26% | 956% 1403%    |
| 67% 71% 72%             | 21% 26% | 952% 1396%    |
| 67% 70% 70%             | 21% 26% | 796% 1077%    |

| Propagation Probability | 0.01 | 0.02 | 0.03 | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
|-------------------------|------|------|------|------|-----|-----|-----|-----|-----|
| Spread coverage          | 16% 18% 18% | 0% 0% 0% 1666% 2402% |
| Saved seeds              | 29% 33% 34% | 0% 0% 0% 1205% 2075% |
| Spread duration          | 41% 46% 47% | 0% 1% 1027% 1830% |
| 59% 65% 67%              | 4% 7% 873% 1454% |
| 81% 87% 88%              | 14% 19% 794% 1023% |
| 91% 95% 96%              | 32% 40% 670% 770% |
| 94% 97% 97%              | 42% 50% 620% 693% |
| 95% 97% 98%              | 47% 55% 621% 686% |
| 95% 98% 98%              | 49% 58% 633% 696% |

| Seed Count | 0.02 | 0.05 | 0.1 | 0.2 |
|------------|------|------|-----|-----|
| Spread coverage | 61% 63% 64% | 8% 12% 23% 234% 391% |
| Saved seeds   | 65% 68% 68% | 16% 23% 489% 773% |
| Spread duration | 68% 73% 73% | 25% 30% 952% 1394% |
|              | 72% 80% 80% | 35% 38% 1930% 2611% |

TABLE IV. Influence of experiments parameters on coverage, saved seeds and time lost. The results for each parameter have been averaged for all other parameters.
Sequential seeding in multilayer networks

FIG. 4. The comparison between SS and SQ strategy for all 108,000 cases. The results have been sorted by the SS percentage of activated actors. Blue colour indicates the gain i.e. how much more actors have been activated by SS in comparison to SQ. Orange colour indicates how big percentage of seeds have not been used (i.e. the budget saved by SS strategy) after activating all actors in the network.

FIG. 5. The comparison between SQ and SQr strategy for all 108,000 cases. The results have been sorted by the SS percentage of activated actors. Blue colour indicates the increase in the coverage in the percentage of all actors in the network i.e. how much more actors have been activated by SQ in comparison to SS. Orange colour indicates how big percentage of seeds have not been used (i.e. the budget saved by SQ strategy) after activating all actors in the network.

FIG. 6. The comparison between 108,000 cases of SS and SQ strategy. In the background and on the right Y-ax we have the same information as on figure 4 while on the front and left Y-ax we have the information on how many times more iterations SQ needed to finish in comparison to SS.
A. Network type

When we look at the results for different networks (Table IV) we can see the number of layers affects the final coverage, what is in line with previous research on that topic. However, there is no evidence that SQ or SQr improvement in terms of coverage depends on the network source (real vs synthetic) or type (random vs scale-free).

However, the number of layers seems to affect saved seeds. Networks with a higher number of layers (N3, N7 and N10) have on average the highest number of saved seeds in their groups (real networks, synthetic-random, synthetic-scale-free respectively).

At the same time the increase in number of iterations is smaller for networks with higher number of layers. What is also interesting, the increased spread duration does not affect real networks as much as synthetic networks which is very good news for real applications.

B. Seed selection strategy

There is no evidence that the seed selection strategy affects the influence of SQ or SQr over the spreading process (Table IV). All three approaches yield similar results.

It needs to be emphasised that this is an expected outcome, as sequential seeding was intended to work on top of, and be independent on, any seed selection strategy, and our results confirm that.

C. Propagation probability

SQ and SQr improve the total coverage regardless of propagation probability (PP) used in ICM (Table IV). They are of course affected by it in a similar way as single stage seeding, i.e., smaller the PP is the smaller coverage we have.

For higher PP (PP ≥ 0.1) where SQ and SQr cannot improve the coverage too much, since usually the entire network is activated, they allow to save seeding budget. As mentioned before, this is a powerful advantage of sequential seeding where by observing the process we can decide to stop acquiring additional seeds and save our budget.

In terms of process duration, SQ and SQr affect more the spreads where we have smaller propagation probability. However, this could also be caused by a simple fact that SQ and SQr ended “earlier” because there was no one else to activate.

D. Seed count

The last parameter we have investigated is seed count (SC). In this case the results are very intuitive (Table IV). The higher number of seeds results in higher coverage for SS, SQ and SQr. Bigger seeding budget we have, the more effective SQ and SQr are, and allow for a higher increase in the final coverage.

For all values of SC, we can observe saved seeds, with a higher number of saved seeds for higher values of SC. However, this might be caused by a simple fact that the bigger budget we have at the beginning the more we can save at the end.

Finally, the more seeds we have the budget for, the longer lasts the spreading process. Because we acquire only one seed per stage, our seeding is usually much longer than the entire spreading in case of SS. For example for network N9, SS ended on average after 5-6 iterations regardless of SC, while just seeding process for SQ took on average 20, 50, 100 and 178 iterations for SC 0.02, 0.05, 0.1, 0.2 respectively. Please note that lower number of iterations than number of seeds in case of SC = 0.2 is a result of process ending before seeding ended due to activation of all nodes in the network.

E. Statistical analysis

The last part of our analysis was the statistical significance evaluation of the results. To do so, Wilcoxon signed-rank test was used as a measure of the difference between sequential and single stage seeding. The results presented in Table IV show higher coverage in terms of Hodges–Lehmann estimator Δ for sequential seeding when compared to single stage seeding with different values of used parameters. Overall results from all simulations showed Δ = 3.45 with p-value < 2.2e-16. Values of Δ > 0 confirm significantly higher values for coverage of sequential seeding when compared to single stage approach. For more detailed results, Hodges-Lehmann estimator was computed for each propagation probability PP used in simulations. It confirms the higher performance of the proposed approach (p-value < 0.05) for all used propagation probabilities. The highest performance (Δ = 6.39) was observed for the propagation probability at the medium level PP = 0.05 while the lowest performance (Δ = 1.14) for low PP = 0.01. In general lower performance if observed for low and high propagation probability values. For high propagation probability, any used strategy can bring good results while for low PP any strategy can fail.

The analysis performed in terms of seed count shows the best performance of sequential seeding for the highest used seed count (0.2); Δ = 7.80. The lowest performance (Δ = 1.16) was observed for the lowest seed count, i.e., 0.02. The main reason is the fact that together with the high number of seeds, there is a higher chance that we will select as seeds nodes that will be activated by their neighbours anyway. Sequential seeding reduces this problem by the activation of additional seeds when the process stops.

Another dimension of analysis took into the account used seed selection strategy. Degree and Neighbourhood based seed selection delivered similar results in terms of Δ with values 4.20 and 4.07, respectively. The lowest performance (Δ = 2.34) was observed for random seed selection.

The results were dependent on used networks. For real networks, the highest differences with Δ = 5.40 was obtained for N3 (EUAir) network. The lowest performance with Δ = 1.93 sequential seeding achieved for N1 (AUCS) network.
| Parameter | Value | Hodges – Lehmann Δ | p-value |
|-----------|-------|---------------------|---------|
| All       | All   | 3.45                | <2.2e-16|
| PP        | 0.01  | 1.14                | <2.2e-16|
|           | 0.02  | 3.11                | <2.2e-16|
|           | 0.03  | 4.80                | <2.2e-16|
|           | 0.05  | 6.39                | <2.2e-16|
|           | 0.1   | 5.63                | <2.2e-16|
|           | 0.2   | 3.05                | <2.2e-16|
|           | 0.3   | 2.47                | <2.2e-16|
|           | 0.4   | 2.37                | <2.2e-16|
|           | 0.5   | 2.29                | <2.2e-16|
| SC        | 0.02  | 1.16                | <2.2e-16|
|           | 0.05  | 2.47                | <2.2e-16|
|           | 0.1   | 4.85                | <2.2e-16|
|           | 0.2   | 7.80                | <2.2e-16|
| 3S        | degree | 4.20              | <2.2e-16|
|           | neighbourhood | 4.07            | <2.2e-16|
|           | random  | 2.34              | <2.2e-16|
| Real nets | N1    | 1.93                | <3.66e-14|
|           | N2    | 3.17                | <2.2e-16|
|           | N3    | 5.40                | <2.2e-16|
|           | N4    | 2.28                | 1.702e-08|
| Synthetic nets | N5           | 1.58              | <2.2e-16|
|           | N6    | 1.35                | <2.2e-16|
|           | N7    | 3.28                | 1.654e-13|
|           | N8    | 5.35                | <2.2e-16|
|           | N9    | 5.65                | <2.2e-16|
|           | N10   | 3.45                | <2.2e-16|

TABLE V. The differences between results for single stage and sequential seeding represented by Hodges–Lehmann estimator Δ with positive values for better results for sequential seeding.

Even though the overall performance is at different levels for all networks, sequential seeding always outperforms single stage seeding with p-value < 0.05. Differences were also observed for synthetic networks. For random networks (N5-N7) a growing number of layers resulted in increased performance of sequential seeding with Δ up to 3.28 for the network with five layers, while for scale-free networks (N8-N10) lower number of layers delivered better results at the levels of 5.35 and 5.65 for two and three layers.

V. CONCLUSIONS

Multilayer networks are usually considered as a better approximation of real interactions between nodes in the network (especially in the case of social networks) since they allow to model multiple relations between nodes. Unfortunately, most of the existing research is focused on simple one layer networks, and the same problem was with the sequential seeding approach. In this paper, we have addressed this issue by extending sequential seeding to multilayer scenario and evaluating it on four real and six synthetic networks. The main results are following: (i) sequential seeding is always at least as good as single stage and in many cases (74%) is better; (ii) sequential seeding very often (31.75%) allows to save seeding budget since it does not need so many seeds to activate all nodes in the network (this is especially important in cases where sequential and single stage seeding produce the same results); however (iii) better coverage and saved seeding budget comes with the price of extended duration of the spreading campaign (on average 9 times longer process). Thus, sequential seeding should be used with caution in campaigns where we have a fixed deadline for influencing people e.g. during a presidential campaign. On the other hand, this approach is useful in application where we do not care so much about time, have limited resources but want to influence as many people as possible.

When looking at the results in detail, we can observe that the number of layers is bound with the savings of seeds: the more layers the network has, the more budget is saved. Another phenomenon observed is that sequential seeding performs the best for rather small propagation probability, but for higher it can also contribute to saving budget, so multiple real-life scenarios can demonstrate the superiority of this method, but the outcome can be of different kind. Lastly, sequential seeding method also proved that it is independent from the underlying seed selection strategy.

When considering future work directions, one of the most interesting ones is to observe how sequential seeding performs when the types of layers are of different kind. One can also think of varying propagation probability for each layer reflecting different intensity of interactions, but this requires additional experiments as well as interpretation coming from social science field.

ACKNOWLEDGMENTS

This work has been supported by National Science Center, Poland, grant number 2016/21/B/HS4/01562.

DATA AND CODE AVAILABILITY STATEMENT

All real networks are available at CoMuNe lab repository [https://comunelab.fbk.eu/data.php](https://comunelab.fbk.eu/data.php). What is more, both real and synthetic networks are published at GitHub repository [https://github.com/pbrodka/SQ4MLN](https://github.com/pbrodka/SQ4MLN) (FullNet folder).

Additionally, for the sake of reproducibility, all networks generated by coordinated execution procedure (900 networks for each evaluated network), results of all experiments as well as the R code, were published at GitHub repository as well.

REFERENCES

1. D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the spread of influence through a social network,” in Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (2003) pp. 137–146.
2. J. Jankowski, P. Bródka, P. Kazienko, B. K. Szymanski, R. Michalski, and T. Kajdanowicz, “Balancing speed and coverage by sequential seeding in complex networks,” Scientific reports 7, 1–11 (2017).
Sequential seeding in multilayer networks

10

3 J. Jankowski, B. K. Szymanski, P. Kazienko, R. Michalski, and P. Bródka, “Probing limits of information spread with sequential seeding,” Scientific reports 8, 1–9 (2018).

4 M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, “Multilayer networks,” Journal of complex networks 2, 203–271 (2014).

5 R. Michalski, P. Kazienko, and J. Jankowski, “Convince a dozen more and succeed—the influence in multi-layered social networks,” in 2013 International Conference on Signal-Image Technology & Internet-Based Systems (IEEE, 2013) pp. 499–505.

6 S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin, “The structure and dynamics of multilayer networks,” Physics Reports 544, 1–122 (2014).

7 M. De Domenico, C. Granell, M. A. Porter, and A. Arenas, “The physics of spreading processes in multilayer networks,” Nature Physics 12, 901–906 (2016).

8 P. Bródka, K. Musial, and J. Jankowski, “Interacting spreading processes in multilayer networks: a systematic review,” IEEE Access 8, 10316–10341 (2020).

9 P. Basaras, G. Iosifidis, D. Katsaros, and L. Tassiulas, “Identifying influential spreaders in complex multilayer networks: A centrality perspective,” IEEE Transactions on Network Science and Engineering 6, 31–45 (2017).

10 X. Huang, D. Chen, D. Wang, and T. Ren, “Identifying influencers in social networks,” Entropy 22, 450 (2020).

11 F. Erlandsson, P. Bródka, and A. Borg, “Seed selection for information cascade in multilayer networks,” in International Conference on Complex Networks and their Applications (Springer, 2017) pp. 426–436.

12 M. Magnani and L. Rossi, “The ml-model for multi-layer social networks,” in 2011 International Conference on Advances in Social Networks Analysis and Mining (IEEE, 2011) pp. 5–12.

13 P. Bródka and P. Kazienko, “Multilayer social networks,” in Encyclopedia of Social Network Analysis and Mining edited by R. Alhajj and J. Rokne (Springer New York, New York, NY, 2018) pp. 1408–1422.

14 J. Goldenberg, B. Libai, and E. Muller, “Talk of the network: A complex systems look at the underlying process of word-of-mouth,” Marketing letters 12, 211–223 (2001).

15 M. Berlingerio, M. Coscia, F. Giannotti, A. Monreale, and D. Pedreschi, “Foundations of multidimensional network analysis,” in 2011 international conference on advances in social networks analysis and mining (IEEE, 2011) pp. 485–489.

16 M. Salehi, R. Sharma, M. Marzolla, M. Magnani, P. Siyari, and D. Montesi, “Spreading processes in multilayer networks,” IEEE Transactions on Network Science and Engineering 2, 65–83 (2015).

17 K. Musial, P. Kazienko, and P. Bródka, “User position measures in social networks,” in Proceedings of the 3rd Workshop on Social Network Mining and Analysis SNA-KDD ’09 (Association for Computing Machinery, New York, NY, USA, 2009).

18 Q. Hu, Y. Gao, P. Ma, Y. Yin, Y. Zhang, and C. Xing, “A new approach to identify influential spreaders in complex networks,” in International Conference on Web-Age Information Management (Springer, 2013) pp. 99–104.

19 L. Page, S. Brin, R. Motwani, and T. Winograd, “The pagerank citation ranking: Bringing order to the web,” Tech. Rep. (Stanford InfoLab, 1999).

20 J.-X. Zhang, D.-B. Chen, Q. Dong, and Z.-D. Zhao, “Identifying a set of influential spreaders in complex networks,” Scientific reports 6, 27823 (2016).

21 L. Seeman and Y. Singer, “Adaptive seeding in social networks,” in 2013 IEEE 5th Annual Symposium on Foundations of Computer Science (IEEE, 2013) pp. 459–468.

22 A. Sela, I. Ben-Gal, A. S. Pentland, and E. Shmueli, “Improving information spread through a scheduled seeding approach,” in Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 (2015) pp. 629–632.

23 J.-L. He, Y. Fu, and D.-B. Chen, “A novel top-k strategy for influence maximization in complex networks with community structure,” PloS one 10, e0145283 (2015).

24 C. Ni, J. Yang, and D. Kong, “Sequential seeding strategy for social influence diffusion with improved entropy-based centrality,” Physica A: Statistical Mechanics and its Applications, 123659 (2019).

25 Q. Liu and T. Hong, “Sequential seeding for spreading in complex networks: Influence of the network topology,” Physica A: Statistical Mechanics and its Applications 508, 10–17 (2018).

26 L. Rossi and M. Magnani, “Towards effective visual analytics on multiplex and multilayer networks,” Chaos, Solitons & Fractals 72, 68–76 (2015).

27 J. Coleman, E. Katz, and H. Menzel, “The diffusion of an innovation among physicians,” Sociometry 20, 253–270 (1957).

28 A. Cardillo, J. Gómez-Gardenes, M. Zanin, M. Romance, D. Papo, F. Del Pozo, and S. Boccaletti, “Emergence of network features from multiplexity,” Scientific reports 3, 1344 (2013).

29 T. A. Snijders, P. E. Pattison, G. L. Robins, and M. S. Handcock, “New specifications for exponential random graph models,” Sociological methodology 36, 99–153 (2006).

30 M. Magnani and L. Rossi, “Formation of multiple networks,” in International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction (Springer, 2013) pp. 257–264.

31 M. Matteo, V. Davide, and D. Mikael, Multinet: Analysis and Mining of Multilayer Social Networks (2020).