

A Dynamic Vote-Rank Based Approach for Effective Sequential Initialization of Information Spreading Processes within Complex Networks

Patryk Pazura¹ Kamil Bortko¹, Jarosław Jankowski¹, and Radosław Michalski²

¹ Faculty of Computer Science and Information Technology, West Pomeranian University of Technology, Szczecin, Poland

{ppazura,kbortko,jjankowski}@wi.zut.edu.pl

² Wrocław University of Science and Technology
Department of Computational Intelligence
Wrocław, Poland

radoslaw.michalski@pwr.edu.pl

Abstract. Seed selection is one of the key factors influencing information spread within networks. Whereas most solutions are based on single-stage seeding at the beginning of the process, performance increases when additional seeds are used. This enables the acquisition of knowledge about ongoing processes and activating new nodes for further influence maximisation. This paper describes an approach based on the Vote-Rank algorithm with dynamic rankings for sequential seed selection. The results prove the increased performance of dynamic rankings compared to the static version and show how the frequency of ranking updates affects both performance and computational costs.

Keywords: Information spreading · Vote-Rank · Influence maximization · Seed selection · Sequential seeding · Viral marketing · Social networks

1 Introduction

Information spreading processes are observed in various aspects of social interaction and commercial activity. They are behind social movements [9], viral marketing [21], political campaigns [4], and spread of misleading information [2]. After gathering additional knowledge about their performance, further actions are often taken to change their dynamics and increase or decrease coverage [25]. Apart from solutions focused on seeding at the beginning of the process without any further actions, other solutions gather knowledge from the process and use additional seeds to improve the process including but not limited to sequential seeding [13], seeding scheduling [34], or adaptive seeding [33]. The approach proposed in this paper is based on the selection of sequential seeds with the use of the Vote-Rank algorithm based on adaptive rankings recomputed before additional seeds are selected. Recomputation uses knowledge gathered about infections within the network and Vote-Rank only considers inactive nodes, not

the nodes already activated. This enables selecting only those nodes with higher potential for spreading as seeds, selected within areas not covered by infections. In this study, we compared the results from static Vote-Rank with the proposed dynamic approach and investigated the influence of recomputation frequency on the final outcome. The remainder of the paper is organized as follows: Section 2 provides a literature review, the conceptual framework is presented within Section 3, followed by experimental results presented in Section 4. The results are summarized and the paper is concluded in Section 5.

2 Literature Review

Information spread within social networks has received attention from researchers from various disciplines. Studies related to information spread have focused on factors affecting their dynamics, the roles of social ties, network topology, and the roles of the links within the network [26]. Models derived from epidemiology research, like SIS, SIR, and their variants, were initially used for prediction and analysis [19]. Later, more dedicated approaches, like the independent cascade model [20] and linear threshold model [5] considered network structures. They were verified with the use of agent-based simulations and the Monte Carlo method [6] or analytical solutions like mean-field models [32] or branching processes [15]. Research was initially performed mainly on single layer static networks, but in recent years, more attention is being focused on multi-layer networks [32] and spreading processes within temporal networks [11][16]. Apart from single processes, multiple processes were analysed with mechanisms related to competition, cooperation, and other forms of interaction [3].

Information spread processes are usually initialised by selected nodes, called seeds, with the use of a dedicated seed selection methods [10]. The influence maximisation problem leads to several challenges and solutions for initial nodes selection [20]. The main goal is to select a set of seeds with high potential to initiate the spread and activate their neighbours. Early approaches were mainly based on heuristics with high degree or other centrality measures like closeness or eigenvector centrality [36]. Apart from simple heuristics, the greedy approach is much more effective, delivering results closer to optimum [20]. Further attempts were made to improve its computational performance with possible applications within larger networks [7]. The number of seeds was analysed to find the minimal effective seed sets [27]. Other solutions considered costs in a form of budgeted solutions [29]. The negative impact of high intensity seeds on users was identified as an over-exposure problem [1]. Another possible goal is limiting overlapping seeds and maximising distance between seeds with the Vote-Rank algorithm, which ranks seed candidates by its votes acquired from direct and indirect neighbours, with higher ranks grouped together with increasing distance from other seeds [38].

Most of earlier solutions focused on selection of seeds at the beginning of the process, without additional actions during the process. Another possibility is spreading seeding over time, with only a fraction of the seeds used at the

beginning, in the form of sequential seeding [13], seeding scheduling [34], or adaptive seeding [33]. The main mechanics are based on avoiding selecting nodes with a high potential to be naturally activated by their neighbours as seeds. Sequential seeding can be used to revive stopped processes or add seeds when the processes are still ongoing [14]. This approach was proven to never deliver worse results than single-stage seeding [17]. Extensions showed how performance of sequential seeding is affected by the topology of networks [26], entropy-based centrality [30], and effective degree [12].

3 The Conceptual Framework

Static rankings used for seed selection only at the beginning of the process create the threat of selected seeds becoming victims of the natural diffusion process. Sequential seeding uses the sequence of seeds instead of using all in a single step to deliver a better result due to the potential of a natural diffusion process. This paper presents an approach that improves the Vote-Rank algorithm with the recomputations and the use of nodes that are effective from the perspective of diffusion processes, that are not yet activated. Earlier study showed that seeds generated with Vote-Rank used sequentially deliver better results than with a single stage [13], but only static Vote-Rank generated at the beginning of the process was analysed and used. The approach presented here avoids gathering votes from network nodes that are no longer effective for information spreading, meaning they were already activated and used for information spreading.

3.1 Illustrative Example

Although Vote-Rank is effective for seed selection, the static ranking generated only at the beginning of the process creates the possibility that the initially good candidates with a high number of votes are no longer good candidates for seeding in the next stages. The proposed approach assumes creation of new ranks based on Vote-Rank before seeding actions. During rank computation, voting only considers votes from nodes available for activation. Already active nodes are not considered. As a result, not yet activated seed candidates with a higher number of direct and indirect connections are preferred. To demonstrate the potential performance of the proposed approach, a toy example is presented in Figure 1. A small network based on nine nodes was generated with the use of the Watts–Strogatz model [37]. A rewiring probability of 0.1 was used and a value of two was assigned to the neighbourhood within which the vertices of the lattice were connected.

Simulation was performed with the use of the coordinated execution proposed in [17], where agent-based simulations processes are not based on randomly generated values in each run, but on values assigned to network edges $A \rightarrow B$, representing the probability of passing information from node A to B and from B to A. If propagation probability is assigned to whole the process with the value p , then information is passed through all edges with weights $\leq p$. The number

of network versions with weights is equal to the planned number of runs. The main advantage of this process is the ability to compare methods within identical conditions. The same approach was used for all simulations here. Figure 1 shows the network used with weights assigned to the edges and simulated spreading according to the independent cascade model [20]. Figure 1 **(A)** shows the process based on sequential seeding with the use of static Vote-Rank and Figure 1 **(B)** shows the process based on dynamic Vote-Rank. Descriptions are provided within the figure caption together with the mechanics behind the dynamic Vote-Rank approach and the advantage produced by a higher number of activations within the network.

3.2 General Assumptions for Experimental Study

In general, experimental study was planned within two stages, with different experimental plans, goals and assumptions. In the first stage, illustrated in Fig. 2 **(Stage I)**, the main goal was to analyse differences between coverage of processes with the use of static (computed only one at the beginning) and dynamic (recomputed before additional seeding takes place) Vote-Rank. Ranking of nodes is created in step 0 and Vote-Rank $VR(0)$ is created. In n subsequent steps additional seeds are selected from the ranking according to the sequential approach and new activations are performed. Seeding is conducted in revival mode, after process dies out. While nodes ranking based on the Vote-Rank is effective at the beginning of the process, together with ongoing process more and more nodes are activated. Proposed approach is focused on better utilization of the knowledge about activations. Vote-Rank is recomputed before each seeding action. Recomputation is based on reduced network without activated nodes. It is assumed that dynamic raking will deliver better network coverage, what is illustrated in Fig. 2 **(A2)**. At each seeding step i , Vote-Rank is computed and new ranking $VR(i)$ is used.

In the Stage II (Fig. 2) study is focused on effects of frequency of rank recomputation on final coverage and computational costs related to the time of rank calculation. Fig. 2 **(B1)** shows approach with recomputations taken in every seeding step. It is expected to be the most effective in terms of coverage, but with highest computational time needed for new rankings creation. Another possibility is to reduce computational time with medium intervals between computations and medium coverage increase Fig. 2 **(B2)**. Together with growing intervals between recomputations, the performance of selected seeds will be dropping. It is considered as lowest performance at Fig. 2 **(B3)**, but still better than for processes based on static rankings.

3.3 Plan of Experiments

An experimental setup is based on two types of Vote-Rank rankings (static and dynamic) and 10 real networks $N1-N10$ from [28] [37][23][22][24][22][8][31][35][18] respectively, containing from 1133 to 16264 nodes and from 5451 to 146160 edges. All values used for all parameters are presented in Table 1. Simulations were

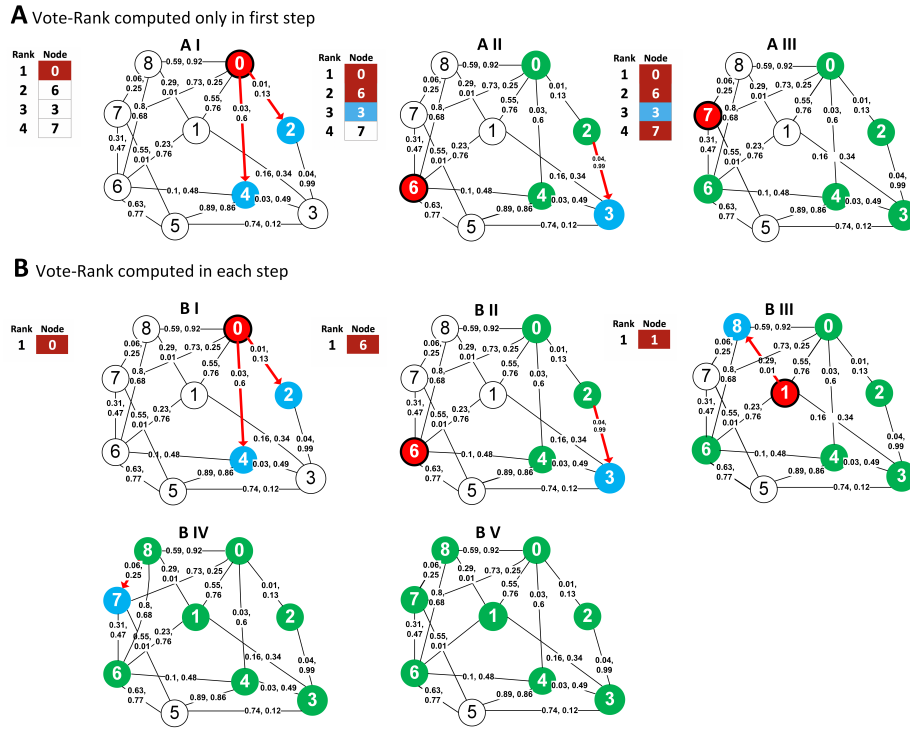
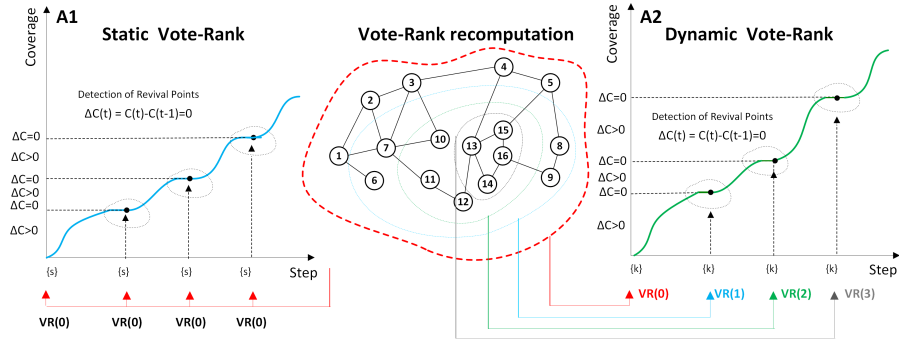


Fig. 1. Results from (A) Spreading process based on static Vote-Rank computed only once, at the beginning of the process. Initially computed ranking is used during whole process. (B) Information spreading process based on sequential seeding and dynamic Vote-Rank computed in every step when seeding takes place. Both sequential seeding processes use three seeds with single seeds used per simulation step. Contagion process is based on Independent Cascade Model with propagation probability $PP = 0.1$. Every edge has assigned two values of weights, therefore to activate neighbour of node A - suppose that is node B, first value on the edge between A and B must be smaller or equal to propagation probability value. (A) Information spreading is initiated by three seeds used in sequence in a form of sequential seeding. Nodes in the network are ranked by Vote-Rank algorithm, with top four nodes presented in the table. First seed is used at the beginning (A I), second seed is used when the process dies out (A II), and the third seed in the same way (A III). In step A III we select last node as seed. While node 3 from Vote-Rank was activated in a natural way in stage A III, node 7 is selected as a seed. This process ends with the total of 6 activated nodes. (B) During this process we compute Vote-Rank at each seeding step. While only one seed is used in each step, only one node is needed with the highest value of Vote-Rank. (B I) In first step according to the ranking node 0 is selected as seed. It tries to activate its neighbors with $PP = 0.1$ and nodes 2 and 4 become active, because of appropriate weighs assigned to the edges allowing transmission. (B II) In the next step, previous activated nodes infecting further, node number 3 becomes active, and node 6 is activated as a seed. In the last stage, (B III), node 1 is selected as a seed and it is activating node number 8. As a result 7 nodes within the network are activated.

Stage I Static vs dynamic Vote-Rank for sequential seeding



Stage II Vote-Rank's recalculations frequency - coverage increase vs computational costs

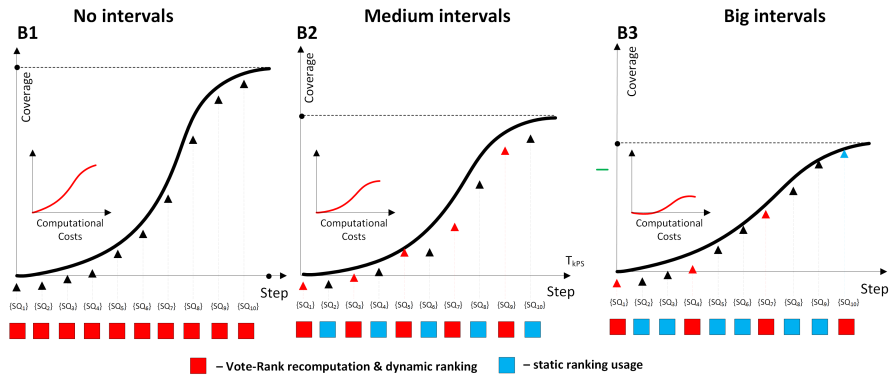


Fig. 2. Two parts of empirical study based on **Stage I**: Comparison of performance of sequential seeding based on static (**A1**) and dynamic ranking of nodes with the use of Vote-Rank algorithm (**A2**) **Stage II**: Analysis of impact of recomputations frequency on coverage and computational costs with (**B1**) computations in each seeding step (**B2**) medium intervals between computations and (**B3**) big intervals with low frequency and costs of computations (lowest coverage).

performed ten times per configuration and results were averaged. For Stage I, together with other diffusion parameters we obtained $R \times N \times PP \times SF \times SP$ for simulations when we were seeding each time when diffusion dies out. Propagation probability (PP) represents propagation probability according to Independent Cascade Model [20]. Each activated node is contacting all not active neighbours and with given probability activates them with only single possible attempt. Seeds fraction (SF) represents the percentage of nodes selected as seeds. Number of seeds per step (SP) represents the number of seeds used in each step of sequential seeding. They are resulting in 4,500 configurations.

Table 1. Networks and parameters of diffusion used in simulations

Symbol	Parameter	No. of variants	Values
R	Ranking type	2	Static Vote-Rank, Dynamic Vote-Rank
N	Network	10	Real networks from various areas
PP	Propagation probability	9	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9
SF	Seeds fraction	5	1%, 2%, 3%, 4%, 5%
SP	Seeds per step	5	1, 2, 4, 8, 16
RI	Recomputation interval	5	1, 2, 4, 8, 16

While the goal of the Stage I was to compare performance of Vote-Rank based on the static and dynamic rankings, in the Stage II main goal was to analyse the impact of recomputations frequency on final coverage with the network. Five recomputation intervals were used. It creates experimental space with $R \times N \times PP \times SF \times SP \times RI$ combinations for simulations when we were seeding with fixed intervals what makes total 22,500 combinations.

4 Empirical Study

4.1 Performance of dynamic Vote-Rank for sequential seeding

Overall analysis compared the dynamic Vote-Rank based approach with static Vote-Rank for sequential seeding with results presented in Fig. 3 (A). All simulation cases of dynamic approach results with not worse, and in most cases higher coverage than static approach. The finest obtained improvement is at the level of 40%. Comparing the value of the Wilcoxon test there is a statistically valid ($p < 0.05$) difference between results from spreading coverage for static and dynamic rankings. The value of Hodges-Lehmann estimator at the level 27.564 indicates a significant improvement in result continuity.

Fig. 4 (A) and (B) show all simulation cases in terms of (A) seeds per step and (B) propagation probabilities. For propagation probability the most noticeable difference is for $PP = 0.1$. Further tendency is visible - the higher propagation probability, the lower increase of performance.

In the next stage the role of propagation probability, number of seeds and network was analysed. Fig. 5 (A) shows a systematic decline of performance as

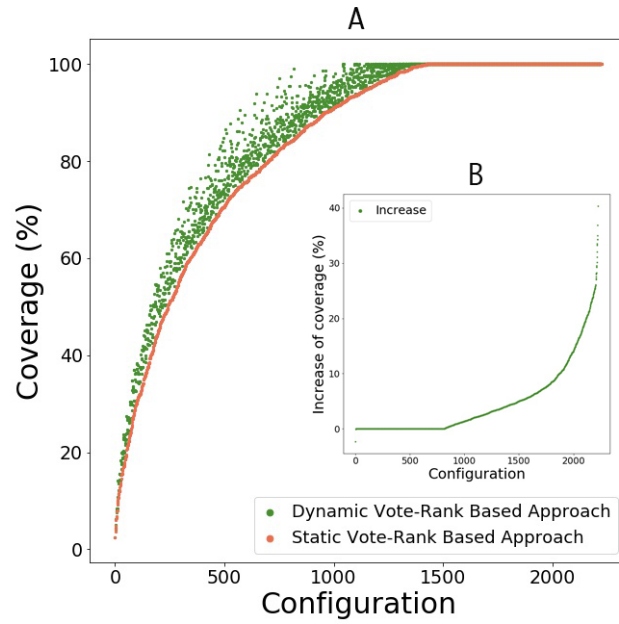


Fig. 3. (A) The comparison of coverage for dynamic Vote-Rank Based Approach and Static Vote-Rank based approach. (B) The increase of the coverage obtained by Dynamic Vote-Rank Based Approach.

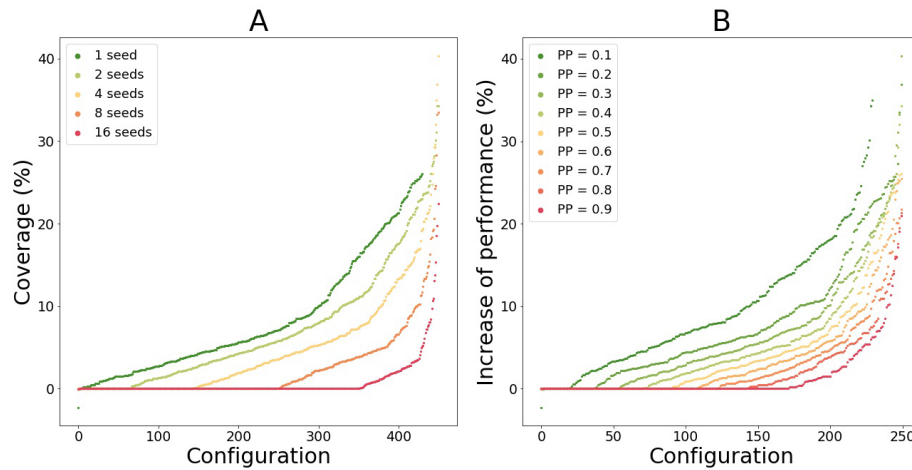


Fig. 4. Results from all simulation cases for (A) The number seeds per step with sorted by performance value (B) Propagation probabilities sorted by performance.

PP increases. The value of increase of coverage from over 8% for PP = 0.1 to nearly 2% for PP = 0.9 is observed. Wilcoxon test was used to analyse results for propagation probability (PP). For PP = 0.1 Hodges-Lehmann estimator was obtained at the level 11.69 while for PP = 0.9 it was at the lower level 5.37. Statistical significance of results was confirmed with $p < 0.05$. Here it is also visible a two-fold drop in the difference. The values oscillate mostly in the range from 8.4 to 11.69.

Similar tendency is visible in Fig. 5 (B), i.e. for number of seeds per step. At the value of 1 it is 8%, then at 2 about 7% and almost twice to the level of about 4% for the number of seeds 4. A decrease to the level of 1% occurs for 16 number of seeds. Based on the Wilcoxon test, taking into account the number of seeds per step, differences for dynamic and static rankings are visible. A downward trend in the statistical difference is observed, which is statistically significant for each number of seeds per step. We see a two-fold difference of Hodges-Lehmann estimator from 17.11 to 8.8 (for SP = 2: 17.11; for SP = 4: 15.33; for SP = 8: 12.46; for SP = 16: 8.8).

Fig. 5 (C) shows differences in results for used network. With networks N1 and N2, it maintains the increase of coverage level above 9%, followed by a decrease to about 1% for network N3. Network 4 maintains a level close to 10%. Starting from network N5 to N10, we see a clear decline to 2-4%. Here we can divide the fall into two groups: a large decrease in network from N5 to N10 and a increase in networks N1, N2, N4. Wilcoxon test was also used for comparing results for pairs of dynamic and static rankings for all used networks N1 - N10. Statistical significance ($p < 0.05$) was obtained. The range of results is from 5.37 for network N3 and up to 10.01 for network N2. Other results reach values close to 9.00, so they are closer to the network with the best result.

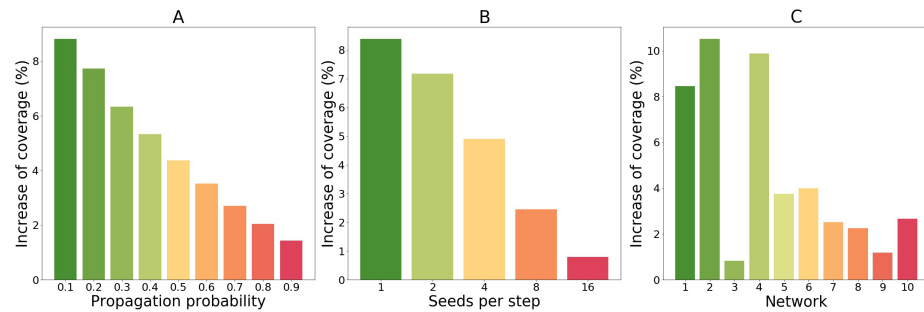


Fig. 5. Difference in averaged increase of coverage for every case of (A) Propagation probability, (B) Seeds per step and (C) Network

4.2 Adjusting computations frequency - trade off between computational costs and coverage increase

In this stage of analysis we considered pros and cons of dynamic Vote-Rank based approach in terms of seeding percentage and recomputation interval to figure out how it affect on coverage performance and computational time. We assumed as seeds per step (SP) number of seeds for every stage of sequential seeding process. In terms of number of seeds (SP), in Fig. 6 (A) is shown how the number of seeds per step (SP) affected coverage performance for each recomputation interval. The lowest coverage performance was observed for 16 seeds per step and interval with value 16, while for 1 seed per step with recomputation in each step the highest coverage performance was obtained. Mean value of coverage performance for 1 seed group is 77.79, for 2 seeds is 76.29, for 4 seeds is 75.75, for 8 seeds is 74.96, for 16 seeds is 70.85. It means that coverage performance decreases along with the number of seeds per step and growing interval between recomputations. Regarding how number of seeds per step and interval affects computational time, we show it in Fig. 6 (B), the longest calculation time was observed for 16 seeds with the highest interval, and was equal to 75.25 seconds. The shortest calculation time was observed for 1 seed, and was 10.99 seconds. For 2 seeds was 11.72 seconds, for 4 seeds was 14.76 seconds, for 8 seeds was 22.09 seconds. As we can infer, adding more seeds isn't profitable. Both computational time and coverage performance fare worse than with lower number of seeds.

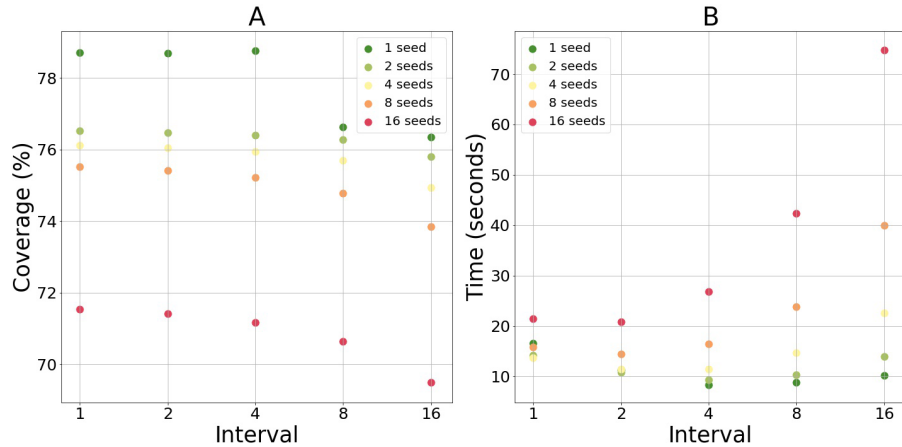


Fig. 6. Results from all simulation cases with showed how the number of seeds affects (A) the coverage and (B) computational time.

In terms of recomputation interval (RI) we carried out analysis to find out how recomputation interval effects on the efficiency. In Fig. 7 (A) is showed coverage performance with each of the colors representing results for different

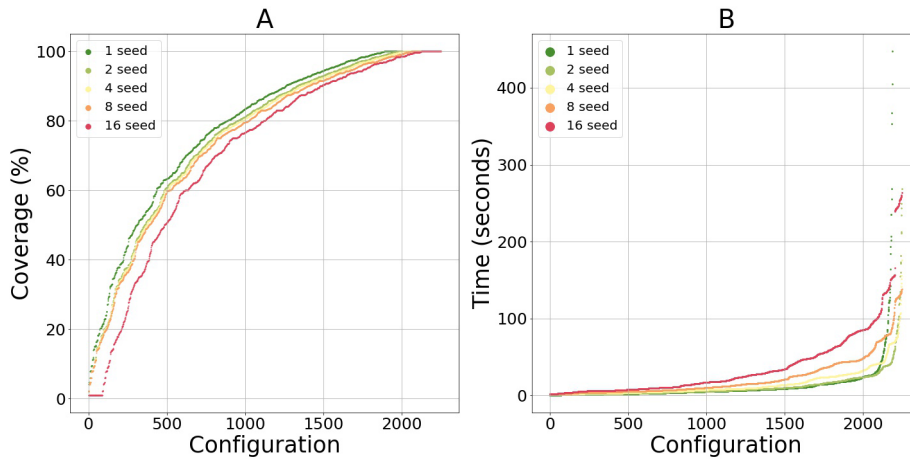


Fig. 7. (A) Results from all configurations for different recombination intervals and the number of seeds per stage (B) Relation between the interval steps of seeding process computational time for all configurations. Results were grouped by value of interval step and averaged.

number of seeds per step, while in Fig. 7 (B) computational time is showed. In Fig. 7 (A) as we can see tendency that the greater we set interval, the smaller coverage is obtained. We can also notice relationships concerning seeds per step similar to those in Fig. 7 (A) and (B). The most effective combinations values of recombination interval and seeds per step is small value of these both. When it comes to Fig. 7 (B) computational time, for smaller recombination interval, there is no need to calculate a rank for steps forward. Consequently we calculate smaller rank, which turned out to have a positive impact on computational time. Analyzing the intergroup comparison using Wilcoxon tests we can see that the smallest differences between the intervals are showed when interval 1 is compared to 4, and interval 2 to 4. They are at the level of 14.61. The biggest differences are more pronounced when comparing intervals 8 to 16, where the differences reach about 35.24, i.e. over two and a half times, than the values from the top. On average, the range of results is in the range of 14.61 to 21.41. All comparison results are presented in the Table 2.

Table 2. Wilcoxon two variants test divided into static and dynamic rankings

Comparison based on group analysis										
Differences between intervals	1 vs 2	1 vs 4	1 vs 8	1 vs 16	2 vs 4	2 vs 8	2 vs 16	4 vs 8	4 vs 16	8 vs 16
	26.68	14.61	17.99	21.28	14.91	17.44	20.76	19.11	21.41	35.24

5 Conclusions

The main goal of this study was to analyse the effects of seed selection for sequential seeding with the use of dynamic rankings generated with the Vote-Rank algorithm. In the typical approach, network nodes are ranked once at the beginning of the process and seeds are selected according to their rank. Together with ongoing spreading processes within network changes, nodes with high potential for seeding at the beginning may no longer be effective. This occurs, for example, if a high fraction of their neighbours are already activated. In the proposed approach, nodes are ranked with the use of a network reduced by already activated nodes. Votes are gathered only from nodes able to be activated. The results demonstrated the performance of the proposed approach with a revival mode when additional seeding occurs after the process dies out. The results were dependent on network characteristics and the increase in performance when compared to the static version was above 10%. The best results were observed for low propagation probabilities. High performance was observed for a low number of seeds used in each step, with best result for one seed per step. Recomputation frequency increased the performance, with the best results obtained for recomputation in every step, but this results in higher computational costs. In many cases, larger intervals between recomputations still improved the performance with lower computational costs.

From the perspective of real applications, it is observed that recent marketing solutions focus on adaptive approaches with the use of knowledge gathered from earlier stages of campaigns. The same can be applied to viral marketing with more natural strategies based on spreading budgets and seeds allocation over the time. It creates potential for dynamic Vote Rank usage, with the ability to cope with large networks, same like it was proved for its static version.

The presented findings provide several future directions for adaptive seeding and usage of knowledge from network states observed when information spreading occurs. Future work could extend the proposed approach toward a more adaptive version with the ability to estimate the time when the recomputation should be performed to maximise the outcome. Another direction is modification of the vote counting method with the use of information about activations within the network.

6 Acknowledgments

This work was supported by the National Science Centre, Poland, grant no. 2016/21/B/HS4/01562.

References

1. Abebe, R., Adamic, L.A., Kleinberg, J.: Mitigating overexposure in viral marketing. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)

2. Acemoglu, D., Ozdaglar, A., ParandehGheibi, A.: Spread of (mis) information in social networks. *Games and Economic Behavior* **70**(2), 194–227 (2010)
3. Bharathi, S., Kempe, D., Salek, M.: Competitive influence maximization in social networks. In: *International workshop on web and internet economics*. pp. 306–311. Springer (2007)
4. Boichak, O., Jackson, S., Hemsley, J., Tanupabrungsun, S.: Automated diffusion? bots and their influence during the 2016 us presidential election. In: *International conference on information*. pp. 17–26. Springer (2018)
5. Chen, W., Yuan, Y., Zhang, L.: Scalable influence maximization in social networks under the linear threshold model. In: *2010 IEEE international conference on data mining*. pp. 88–97. IEEE (2010)
6. Frias-Martinez, E., Williamson, G., Frias-Martinez, V.: An agent-based model of epidemic spread using human mobility and social network information. In: *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*. pp. 57–64. IEEE (2011)
7. Goyal, A., Lu, W., Lakshmanan, L.V.: Celf++: optimizing the greedy algorithm for influence maximization in social networks. In: *Proceedings of the 20th international conference companion on World wide web*. pp. 47–48. ACM (2011)
8. Guimera, R., Danon, L., Diaz-Guilera, A., Giralt, F., Arenas, A.: Self-similar community structure in a network of human interactions. *Physical review E* **68**(6), 065103 (2003)
9. Harlow, S.: It was a facebook revolution”: Exploring the meme-like spread of narratives during the egyptian protest. *Revista de comunicaci3n* (12), 59–82 (2013)
10. Hinz, O., Skiera, B., Barrot, C., Becker, J.U.: Seeding strategies for viral marketing: An empirical comparison. *Journal of Marketing* **75**(6), 55–71 (2011)
11. Holme, P.: Temporal network structures controlling disease spreading. *Physical Review E* **94**(2), 022305 (2016)
12. Jankowski, J.: Dynamic rankings for seed selection in complex networks: Balancing costs and coverage. *Entropy* **19**(4), 170 (2017)
13. Jankowski, J., Br3dka, P., Kazienko, P., Szymanski, B.K., Michalski, R., Kajdanowicz, T.: Balancing speed and coverage by sequential seeding in complex networks. *Scientific reports* **7**(1), 891 (2017)
14. Jankowski, J., Br3dka, P., Michalski, R., Kazienko, P.: Seeds buffering for information spreading processes. In: *International Conference on Social Informatics*. pp. 628–641. Springer (2017)
15. Jankowski, J., Michalski, R., Kazienko, P.: The multidimensional study of viral campaigns as branching processes. In: *International Conference on Social Informatics*. pp. 462–474. Springer (2012)
16. Jankowski, J., Michalski, R., Kazienko, P.: Compensatory seeding in networks with varying availability of nodes. In: *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*. pp. 1242–1249. IEEE (2013)
17. Jankowski, J., Szymanski, B.K., Kazienko, P., Michalski, R., Br3dka, P.: Probing limits of information spread with sequential seeding. *Scientific reports* **8**(1), 13996 (2018)
18. Joshi-Tope, G., Gillespie, M., Vastrik, I., D’Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G., Wu, G., Matthews, L., et al.: Reactome: a knowledgebase of biological pathways. *Nucleic acids research* **33**(suppl_1), D428–D432 (2005)
19. Kandhway, K., Kuri, J.: How to run a campaign: Optimal control of sis and sir information epidemics. *Applied Mathematics and Computation* **231**, 79–92 (2014)

20. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 137–146. ACM (2003)
21. Leskovec, J., Adamic, L.A., Huberman, B.A.: The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)* **1**(1), 5 (2007)
22. Leskovec, J., Kleinberg, J., Faloutsos, C.: Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)* **1**(1), 2 (2007)
23. Leskovec, J., McAuley, J.J.: Learning to discover social circles in ego networks. In: Advances in neural information processing systems. pp. 539–547 (2012)
24. Ley, M.: The dblp computer science bibliography: Evolution, research issues, perspectives. In: International symposium on string processing and information retrieval. pp. 1–10. Springer (2002)
25. Libai, B., Muller, E., Peres, R.: Decomposing the value of word-of-mouth seeding programs: Acceleration versus expansion. *Journal of marketing research* **50**(2), 161–176 (2013)
26. Liu, Q., Hong, T.: Sequential seeding for spreading in complex networks: Influence of the network topology. *Physica A: Statistical Mechanics and its Applications* **508**, 10–17 (2018)
27. Long, C., Wong, R.C.W.: Minimizing seed set for viral marketing. In: 2011 IEEE 11th International Conference on Data Mining. pp. 427–436. IEEE (2011)
28. Newman, M.E.: Scientific collaboration networks. i. network construction and fundamental results. *Physical review E* **64**(1), 016131 (2001)
29. Nguyen, H., Zheng, R.: On budgeted influence maximization in social networks. *IEEE Journal on Selected Areas in Communications* **31**(6), 1084–1094 (2013)
30. Ni, C., Yang, J., Kong, D.: Sequential seeding strategy for social influence diffusion with improved entropy-based centrality. *Physica A: Statistical Mechanics and its Applications* p. 123659 (2019)
31. Opsahl, T.: Why anchorage is not (that) important: Binary ties and sample selection. online] <http://toreopsahl.com/2011/08/12/why-anchorage-is-not-that-important-binary-tiesand-sample-selection> (accessed September 2013) (2011)
32. Sahneh, F.D., Scoglio, C., Van Mieghem, P.: Generalized epidemic mean-field model for spreading processes over multilayer complex networks. *IEEE/ACM Transactions on Networking (TON)* **21**(5), 1609–1620 (2013)
33. Seeman, L., Singer, Y.: Adaptive seeding in social networks. In: 2013 IEEE 54th Annual Symposium on Foundations of Computer Science. pp. 459–468. IEEE (2013)
34. Sela, A., Ben-Gal, I., Pentland, A.S., Shmueli, E.: 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. pp. 629–632. ACM (2015)
35. Šubelj, L., Bajec, M.: Software systems through complex networks science: Review, analysis and applications. In: Proceedings of the First International Workshop on Software Mining. pp. 9–16. ACM (2012)
36. Wang, X., Zhang, X., Zhao, C., Yi, D.: Maximizing the spread of influence via generalized degree discount. *PloS one* **11**(10), e0164393 (2016)
37. Watts, D.J., Strogatz, S.H.: Collective dynamics of ‘small-world’ networks. *nature* **393**(6684), 440 (1998)
38. Zhang, J.X., Chen, D.B., Dong, Q., Zhao, Z.D.: Identifying a set of influential spreaders in complex networks. *Scientific reports* **6**, 27823 (2016)