Multi-criteria approach to viral marketing campaign planning in social networks, based on real networks, network samples and synthetic networks

Artur Karczmarczyk*, Jarosław Jankowski* and Jarosław Wątróbski†
*Faculty of Computer Science and Information Technology
West Pomeranian University of Technology in Szczecin, Żołnierska 49, 71-210 Szczecin, Poland
Email: {artur.karczmarczyk,jaroslaw.jankowski}@zut.edu.pl
†Faculty of Economics and Management
University of Szczecin
Mickewicza 64, 71-101 Szczecin, Poland
Email: jwatrobski@usz.edu.pl

Abstract—Spreading of information within social media and techniques related to viral marketing take more and more attention from companies focused on targeting audiences within electronic systems. Recent years resulted in extensive research centered around spreading models, selection of initial nodes within networks and identification of campaign characteristics affecting the assumed goals. While social networks are usually based on complex structures and high number of users, the ability to perform detailed analysis of mechanics behind the spreading processes is very limited. The presented study shows an approach for selection of campaign parameters with the use of network samples and theoretical models. Instead of processing simulations on large network, smaller samples and theoretical networks are used. Results showed that knowledge derived from relatively smaller structures is helpful for initialization of spreading processes within the target network of larger size. Apart from agent based modeling, multi-criteria methods were used for evaluation of results from the perspective of costs and performance.

I. INTRODUCTION

Online platforms evolved from early stage technical systems to social media with integrated mechanics of social communication and interactions close to the real world [1]. Together with growing audiences, they attracted more attention of marketers. Apart from typical digital marketing channels based on display advertising and search engines new strategies focused on social media emerged. They include mechanism based on detailed targeting, consumer behavior analysis and commercial content dissemination with the use mechanisms of information spreading.

Results delivered from viral campaigns usually outperform traditional campaigns because of the utilized social influence and ability to induce high dynamics even with low budgets [2]. Social recommendations have high impact on customer decisions and, properly integrated with marketing communication [3], help to further increase performance [4].

The recent studies focused on viral marketing take into account data from real platforms as well as theoretical network models [5]. One of the goals is to increase campaign dynamics

and coverage with properly selected initial customers during the seeding process [6]. Apart from static networks, dynamic networks with varying structures are taken into account [7]. Other approaches take into account multi-layer structure of networks representing specifics of real social relations based on different networks, for example private and professional contacts [8].

Theoretical and simulation models are used for prediction of network coverage. They can be derived from analytic models used in epidemiology [9] or can be more focused on network structures and characteristics [10]. Other possibility is to use theories and models related to the diffusion of innovations [11].

While most of the research is focused on coverage and number of infected nodes within the network, from the practical point of view, marketing campaigns can have different goals and specifics. They are planned within assumed budget constraints and timing. A different strategy can be used to acquire high number of potential customers in a very short time than for a long term planning and organic growth of customer database. Campaign budget influences the number of initially infected nodes (seeds) and demographic characteristics. The quality of seeds and their number can be a key factor of campaign coverage and overall results. Additional budgets can be used to increase campaign dynamics or lifespan. To take into account various goals multi-criteria campaign evaluation can be used to select campaign parameters and goals according to preferences and priorities [12]. Earlier research has shown that in order to reduce computational complexity, campaigns can be planned with the use of simulations within smaller synthetic networks based on theoretical model. However, since the theoretical models might not always fit the real networks, the current study proposes the use of network samples for the initial simulations and detection of campaign parameters. Both approaches were compared with results obtained from the complete network and showed the ability to obtain approximate results with network samples.

The paper comprises of five main sections. After this introduction, in Section II literature review is presented. It is followed by the methodology presented within Section III and results in Section IV. Paper is concluded in Section V.

II. LITERATURE REVIEW

Social platforms gather detailed information about user behavior and social relations with the main goal to better address commercial messages and properly target products and services [13]. The growing complexity and volumes of the collected data is a direct result of the growing number of users and that their activities moved to electronic systems [14], [15]. Social platforms are treated as tools to use social influence mechanisms to spread information between friends with the impact strengthened by social recommendations. Contacts within social networks are used to pass the information and it often induces information cascades as a main driver of viral marketing campaigns. Multidisciplinary nature of phenomena connected with information diffusion integrates efforts from scientists from various fields like sociology, computer science, physics and management with a different theoretical and practical goals [4] [9] [6].

For better understanding of the information spreading processes, theoretical models are used and they are often implemented within agent based environment or used for analytic studies [16]. Methodological background of studies is often based on models initially created for epidemic research like SIR or SIS with taken into account analytic view on processes and their dynamics [9]. Apart from them, more dedicated solutions were created to create models on microscopic level using information about network structures and relations between users. They are based on two key mechanisms represented by linear threshold models [11] and independent cascades [10]. Linear threshold model, with its later extensions, assumes the social influence induced by neighbors with the network and information flow when the number of neighbors exceeds assumed threshold. Cascading models use different mechanics with spreading based on propagation probabilities and communication with surrounding neighbors and passing content to them. These approaches can be treated as pull and push spreading models. Spreading models can be also used for analysis based on aggregated and macroscopic level [17].

Apart from the mechanics of the information spreading, the dynamics of processes are related to network models and their structures. For the simplest approaches, static networks of non-varying structures are used. More closer to reality are approaches focused on dynamic networks with a changing number of social connections or availability of nodes [18]. For better representation of real systems multi-layer networks are used with spreading dependent on connections between layers, their structures or similarities [8].

Many studies related to information spreading take into account the selection of initial customers, in a form of a seeding process, targeted with product samples or other marketing content with the main goal to motivate them to spread the information to friends within the network [6]. Proper selection of seeds is crucial for successful campaigns, but the problem identified as influence maximization problem is NP-hard [10]. Greedy solutions deliver effective results, but with the high computational cost they are difficult to use within real networks [10]. More practical approaches base on heuristics and a selection of nodes with the use of the network metrics like degree or betweenness. Centrality measures can be used for selection of initial influencers with assumed characteristics [19] [20].

Apart from seeding only once at the beginning of the process, knowledge about the process performance can be gathered and used for additional actions to improve the process characteristics. Adaptive approaches can be used [21] to increase the reach and better utilize the available knowledge. Other possibility is to spread the seeds over the time and better utilize the natural spreading processes. It can be applied in a form of sequential seeding [22] or its extension with recomputed nodes' rankings at every simulation step [23]. Further improvement of seeding can be performed with the use of knowledge about community structures within the network [24], voting mechanics [25] or k-shell based approach dedicated for identification of nodes with high spreading potential [26].

Apart from single campaigns spreading, processes can interact or compete [27]. For such scenarios seeding can be planned to increase the chance of process to survive among competitors or reach audiences in a shortest time before other processes acquire them. Similar situation takes place in epidemic research where two or more pathogens are competing with each other or conditional infections are observed with activity of first virus required for next viruses. Competing scenarios are observed when awareness spreading is deceasing dynamics of epidemic [28]. It lead to extension of the single campaign models to multi-spreading processes for viral marketing studies[29].

Another studies take into account content specifics and network structures [30], proper ways to motivate users to forward the content [31], influence of emotions on content propagation processes [32] [33] and other structural or functional factors [34] [35].

The earlier studies focused mainly on influence maximization to increase coverage within the network. Campaign evaluation was was also discussed as a multi-criteria problem [12]. Campaigns performed within agent based simulation environment were evaluated with the use of set of criteria related to budgets, campaign costs and the number of target nodes. Model output was delivering solutions with defined number of seeds or propagation probabilities. Study also showed the ability to perform simulations with theoretical models and apply selected strategies to real network. The current study extends the presented approach and uses network samples created with the use of snowball sampling [36].

III. METHODOLOGY

Viral marketing campaigns can be based on various strategies. During the campaign planning, decisions are taken about

optimal number of initial seeds, methods used for their selection, motivation techniques used for users to increase their willingness to spread the content and type of incentives used to increase the propagation probabilities. Similar problems are related to campaign evaluation and selection of campaign metrics dependent on campaign goals. Other performance metrics can be used for campaigns focused on high network coverage than on highly targeted processes addressed only to specific customers.

While social networks store information about users, connections and network structures, it is possible to analyze information before campaign to optimize the strategy and maximize results. With the assumed campaign scenarios and goals it is possible to simulate and test different strategies for selection of campaign parameters. Due to high computational complexity it would be difficult for larger networks.

The approach proposed in this paper assumes the generation of synthetic networks based on theoretical models, generation of network samples based on real network, performing simulations focused on verification of different seeding strategies and campaign parameters and evaluation of results with the use of MCDA methods and, finally, launching the campaign within the real network (see Fig. 1)

Simulations can be performed within synthetic networks based on theoretical models like Barabasi-Albert model (BA) [37], Watts-Strogatz (WS) model [38] and Erdos-Renyi model (ER) [39]. The size of synthetic networks can be adjusted with reference to the size of real network and it can be a fraction of the real network e.g. 10%, 20%, 30% etc. It is also important to select proper network model with high similarity to real network. The presented approach uses Kullback-Leibler measure (KL) to compare network similarities [40]. Number of nodes and edges within synthetic network can be scaled for better performance and accuracy.

Since a real network not always must be similar to idealized theoretical models, another approach can be based on network samples generated as a fraction of the real network. Snowball sampling can be used to obtain smaller structures, which would allow to perform simulations easier, yet with assumed similarity to the full network structures. Samples can be scaled from lower to higher fraction of the complete network. It is assumed that accuracy of simulations in the bigger samples is more close to the real network but computational complexity is lower for the smaller samples.

The simulations for all samples and synthetic networks are performed with the use of various campaign parameters. The number of seeds represented by the seeding fraction (SF) and its effect on total coverage can be verified and is the representation of a campaign budget. Another decisions are related to seed selection strategy (SS). It can be based on different network metrics and it is also related to campaign costs. For example, targeting high degree nodes can be more expensive than low degree nodes.

From the other point of view, the selection of nodes with high closeness can be more expensive than the selection of nodes with high degree because of higher computational

complexity required to compute closeness metrics than degree. Another tested parameters are based on propagation probabilities (PP). For lower propagation probabilities, coverage within the network will be lower, but higher probabilities require higher motivation of users to forward the content. It may require incentives and is related to increased budgets.

To compare results from samples and synthetic networks, the proposed study performs analysis for all networks used. The MCDA module takes into account possible campaign success evaluation criteria like coverage, dynamics, campaign costs. In the subsequent step, the performance table obtained from the samples, as well as the criteria and preferences, are used to produce a ranking of possible advertising strategies with the selected MCDA method. After analyzing the ranking and performing robustness / sensitivity analysis, the analyst provides the campaign parameters recommendation for real network campaign.

In the prior research [12], the authors successfully used the PROMETHEE II method [41], [42] to evaluate viral marketing campaign strategies. However, in the proposed research the authors' wanted to emphasize the effect that the marketers' weights assigned to particular criteria have on the final strategies evaluation. Therefore, it was decided that full sensitivity analysis of the obtained solutions should be performed, which eliminated aspect of uncertainty of the decision maker's criteria preference. Moreover, since in the proposed approach the input data comes from simulations, data uncertainty can also be disregarded. However, the evaluation problem at hand still is characterized by weights and data expressed on a quantitative scale. Last, but not least, the obtained solution to the strategy evaluation problem should take the form of a complete ranking to allow the choice of the best strategy. Therefore, based on the analysis of 65 MCDA methods [43], [44] and the guidelines included in [45] and [46], the authors decided to found their approach on the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method [47].

The TOPSIS method is a representative of the American MCDA school [48] which transforms all decision-making problem criteria into a single score value. In case of the TOPSIS method, based on the criterial performance of the evaluated criteria, a positive and negative ideal strategies are created, i.e. one which tops at each criterion and one that bottoms at all criteria. Subsequently, the score of each appraised strategy is computed as a relative distance between the strategy and both the positive and negative ideal solutions. Therefore, the best strategy would be the one which is closest to the positive ideal strategy, yet as far as possible from the negative ideal strategy in terms of criterial performance values.

IV. RESULTS AND DISCUSSION

A. Evaluation of viral marketing campaign strategies on a real network

The empirical study was based on a real network, a part of the topology of the Gnutella network as mapped in 2002 in the
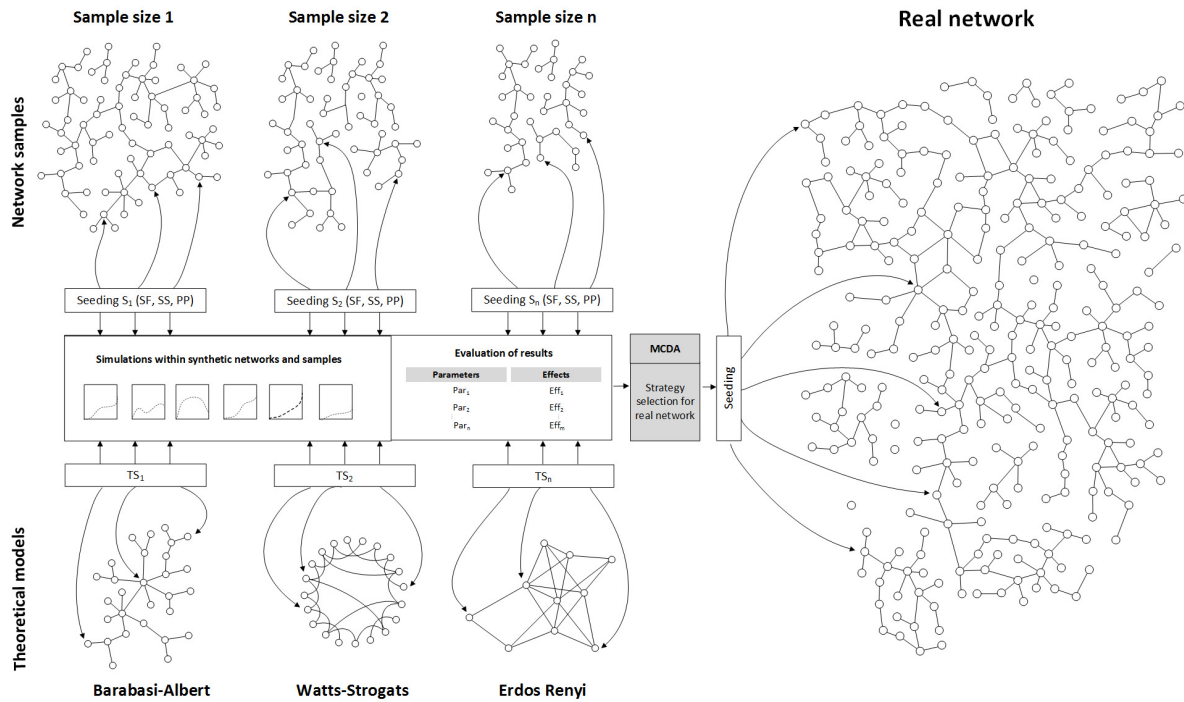
Fig. 1. Conceptual framework for real network strategy selection based on simulation results within network samples and theoretical models

[49] research. The mapped network comprises of 8846 nodes and 31839 edges. The nodes represent hosts in the Gnutella network topology and the edges represents connections between the Gnutella hosts in a single of the network snapshots collected in August 2002. The average values of the main network's metrics are as follows:

1) total degree $D = 7.1985$
2) closeness $C = 1.587441e - 07$
3) Page Rank $PR = 0.0001130454$
4) Eigen Vector $EV = 0.01602488$
5) clustering coefficient $CC = 0.0001130838$
6) betweenness $B = 19104.87$

During the empirical study, the authors used the proposed framework to plan and simulate a viral marketing campaign. Ten simulation scenarios were generated to assure repeatability of the results regardless of the input parameters. Each scenario was composed of the weights drawn for each edge, ranging $< 0; 1 >$. These weights were later compared with the propagation probability of each node to determine whether or not the actual information propagation would occur.

As part of the simulations, a total of 400 sets of parameters were tested, built as a Cartesian product of the following simulation parameter values:

1) Par1 - 0.01, 0.02, ..., 0.09, 0.10;
2) Par2 - 0.01, 0.10, 0.20, ..., 0.90;
3) Par3 - degree (1), closeness (2), eigenvector centrality (3), betweenness (4) – the value is the rank of the method based on its computation speed.

Rank	Alt	CCi	SF	PP	Last Iter	Coverage	Measure
1	A11	0.7494	0.01	0.20	14.4	0.5174	1
2	A10	0.7244	0.01	0.20	14.2	0.5176	2
3	A7	0.7218	0.01	0.10	16.8	0.1334	1
4	A51	0.7148	0.02	0.20	13.2	0.5189	1
5	A50	0.7033	0.02	0.20	13.6	0.5201	2
6	A8	0.6975	0.01	0.10	19.6	0.1127	3
7	A12	0.6960	0.01	0.20	14.7	0.5172	3
8	A6	0.6901	0.01	0.10	15.6	0.1359	2
9	A47	0.6883	0.02	0.10	14.5	0.1625	1
10	A48	0.6878	0.02	0.10	18.9	0.1218	3
11	A52	0.6839	0.02	0.20	14.5	0.5181	3
12	A46	0.6805	0.02	0.10	15.1	0.1638	2
13	A91	0.6799	0.03	0.20	12.2	0.5213	1
14	A88	0.6732	0.03	0.10	18.1	0.1313	3
15	A92	0.6722	0.03	0.20	14.5	0.5194	3
16	A90	0.6635	0.03	0.20	12.4	0.5221	2
17	A128	0.6588	0.04	0.10	17.6	0.1464	3
18	A87	0.6561	0.03	0.10	13	0.1870	1
19	A131	0.6560	0.04	0.20	11.8	0.5230	1
20	A132	0.6537	0.04	0.20	14.3	0.5207	3

Fig. 2. Visualization of the top 20 alternatives from the TOPSIS evaluation of the [49] real network.

Consequently, 4000 simulations were performed for the [49] network. The results of each simulation run were registered, including inter alia the iteration during which the last infection occurred as well as the total coverage achieved, which values were labelled for the further evaluations as Eff4 and Eff5.

After the simulations concluded, the TOPSIS method was

used to evaluate all 400 campaign scenarios. Initially, the weights of all criteria well set equal. The preference direction of the **Par1-Par3** criteria was minimum and of the **Eff4-Eff5** was maximum. Intuitively that would mean the decision maker would prefer low cost of the enterpreneurship, yet long duration and maximum coverage. The top 20 strategies are presented on Fig. 2. The best strategy, A11, obtained ϕ_{net} score of 0.7494. This strategy is based on low values of SF and PP (0.01 and 0.20 respectively) and degree as the method of seeding nodes selection. The runner-up alternative, A10, is based on the same SF and PP values, but uses closeness as the method for selecting the seeding nodes. As a result, slightly broader coverage was achieved in minutely less iterations (0.02s difference). The third-best strategy, A7, maintains the degree measure and the SF of 0.01, however it reduces the PP by half, to 0.10. Such strategy would non-negligibly reduce the costs of the campaign (lower investment in incentives), and, since less nodes at each step would get infected, the procedure would take longer (16.8 iterations on average). However, the obtained coverage is significantly lower, equal to 0.1334 of the network, which is over three-fold worse than the winning A11 strategy.

For the purposes of comparison, the worst strategy, A400, was based on high SF (0.10), high (ignitable) PP (0.9) and eigenvector centrality as the measure. As a result, the contamination process averagely finished within 5.1 iterations, with the mean coverage of 0.9722. Although almost full network gets covered with that strategy, it is important to note that the incentive costs for such strategy would be very high to achieve 90% propagation probability. Also the duration of the campaign would be low, which is against the DM's preferences.

One of the benefits of the TOPSIS method is the fact it allows to build an ideal reference model for the given evaluation problem. In case of the problem at hand, the ideal strategy would be based on degree for selecting the nodes to seed information to and only 1% nodes would be seeded. Incentives would be in place to generate an average propagation probability of 0.01%. With such parameters of the network, the DM would like the outcomes of the marketing campaign to be 19.6 iterations resulting in 97.22% coverage. It is important to note, however, that although ideal, such strategy is only a reference model and does not exist.

The rank presented on Fig. 2 is based on an assumption that the weight of each criterion on the final outcome is equal. However, the DM often gives more significance to some criteria over the others. One of the tremendous benefits of the utilisation of MCDA in the evaluation of viral marketing campaign strategies is the possibility to perform a sensitivity analysis, to learn how even slight changes in preferences of each criterion would affect the final outcome. Therefore, in a subsequent part of the research, a sensitivity analysis was performed to show how the ranking relations between the top 20 alternatives would change if the weights of each criterion would change. The analysis was divided into five parts, one for each criterion. During each phase, the weight of a single criterion was changed from 1 to 100, while the weights of the remaining criteria were set equally to 50.

The results of the sensitivity analysis are presented on Fig. 3. The top row of the figure (A-E) presents how the score of each strategy changed, resulting from each criterion's weight change, whereas the bottom row of the figure (F-J) presents how that change affected the strategies' positions in the ranking. The analysis of Fig. 3A,F shows that no matter how the weight of the criterion Par1 changed, strategy A11 remained the leading one. On the other hand, if the weight of this criterion dropped slightly below 40, strategy A7 would outrun strategy A10. Strategy A51 rank is not affected by the changes of weight of criterion Par1, whilst strategy A50 (ranked fifth) would be outrun by strategy A12 (ranked 7) if its weight was higher than 75. The analysis of the chart on Fig. 3A allows to observe, that while the score of alternatives A128, A131 and A132 is not significantly affected by the changes of Par1 weight, the remaining strategies gain more score as the weight of this criterion increases. A similar tendency can be observed on Fig. 3B, where the scores of all strategies increase along with the increase of significance of criterion Par2. When the weight of that criterion would exceed 90, the runner-up strategy A7 would outrun the strategy A11. An opposite tendency can be observed on Fig. 3E, where all alternatives lose score when the weight of Eff5 grows. Along with this criterion's weight growth, there are only little changes in the order of the three leading alternatives, however, if the weight of that criterion dropped close to 0, the leading strategy A11 would drop six positions to rank 7. This demonstrates the fact that strategy A11 is considerably supported by criterion Eff5. The observation of Fig. 3F-J shows that while for the criteria Par1 and Par2 the majority of rank changes occur when the weight of the criterion changes considerably, in case of criteria Par3 – Eff5, most of the rank changes occur with even minute changes of these criteria's weights.

B. Selection of synthetic networks

As it was presented in section IV-A, the proposed MCDA framework allows to successfully evaluate various viral marketing campaign strategies performed over a real network. However, full networks are rarely available for the entities ordering campaigns. Often, only characteristics of a network are provided. Moreover, running comprehensive simulations on a real networks containing multitude of nodes is also time consuming. Therefore, it is beneficial to perform simulations on smaller synthetic networks before launching the actual campaign on a real network.

Consequently, in the empirical research, apart from evaluating campaign strategies based on full, real network, the authors also used the proposed framework to perform simulations on synthetic networks, similar to the real one, but of a reduced size. The strategies' rankings obtained for synthetic networks were then compared to the ranking obtained for the real network.
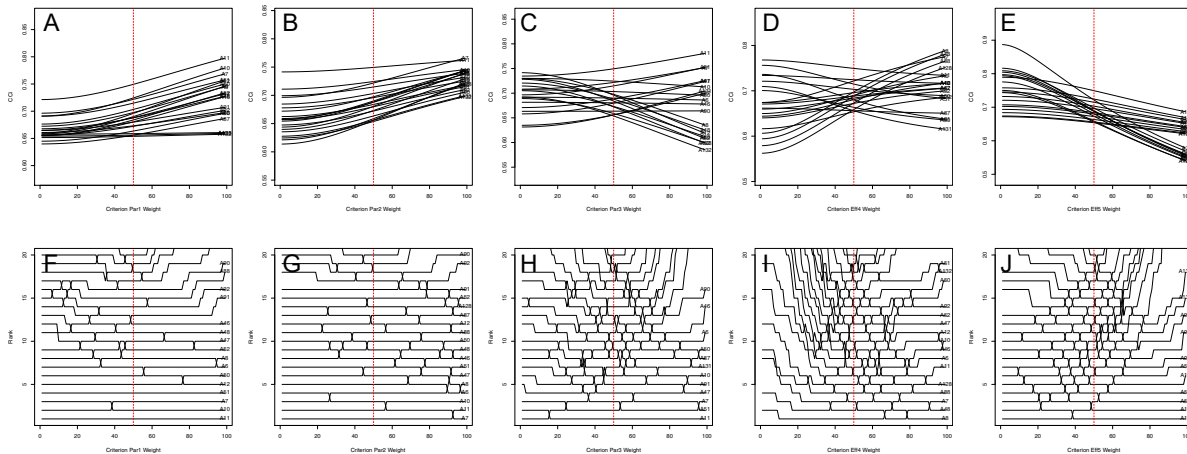
Fig. 3. Ranking sensitivity analysis for the top 20 alternatives from the TOPSIS evaluation of the [49] real network.

For the 10%, 30% and 50% size of the real network, BA, WS and ER networks were generated with the following parameters:

1) BA - number of nodes equal to 10%, 30% and 50% of the real network; number of edges m to add in each step equal to $1, 2, \ldots, 5$ – a total of 15 networks;

2) WS - number of nodes equal to 10%, 30% and 50% of the real network; the neighborhood within which the vertices of the lattice will be connected equal to $1, 2, \ldots, 5$ – a total of 15 networks;

3) ER - number of nodes equal to 10%, 30% and 50% of the real network; number of edges equal to the chosen number of nodes multiplied by $1, 2, \ldots, 5$ – a total of 15 networks.

As a result, a set of 45 networks was generated. In order to avoid arbitrary decisions which network to run the simulations on, the Kullback-Leibler divergence measure was used to compare the degree distribution of all generated networks to the real one. Based on the smallest value of the KLD measure, three networks were selected for further simulations. The selected networks are presented in Table IV-B.

C. Viral marketing campaign strategies planning with synthetic networks

The results of the viral marketing campaign strategies planning with the use of the three aforementioned BA networks is presented in table on Fig. 4. The analysis of the table allows to notice that regardless of the selected network size, in all three cases the same strategy A11 was chosen as the superior one, similarly to the real network case. While in case of the real network this strategy lasted averagely for 14.4 iterations and resulted in 0.5174 coverage, in case of the synthetic networks, the process averagely lasted $10 - 11.2$ iterations (slightly shorter) and resulted in $0.5783 - 0.7049$ coverage (slightly higher). The second best strategy in all three synthetic networks was strategy A51, which above, in case of the real network, was ranked fourth. This strategy is based

on small values of SF and PP (0.02 and 0.20 respectively) and lasts averagely in $9.5 - 10.7$ iterations resulting averagely in $0.5783 - 0.7049$ coverage. The measure used here is also degree, as in the winning alternative.

The strategy A10, which for the real network evaluation was ranked second, in case of the synthetic networks reached place 3 for the 50% network and rank 4 for the remaining networks. More interesting is the case of strategy A7. On the real network it is ranked third, for the 30% network it remained at the same ranking position, however, for the 50% network it dropped to the fourth rank, whilst for the 10% network its ranking fell to 15th position. The strategy A7 is characterized by its very low SF and PP values (0.01 and 0.10 respectively) and degree as the measures which makes it one of the cheapest, with maximally extended information propagation process duration, on the cost of small final coverage. The duration of the process is very long for this strategy on the real network and the 30% and 50% networks (16.8, 12.8 and 13 iterations averagely, while the maximum average duration was 19.6, 12.9 and 13.7 iterations respectively). In case of the 10% synthetic network, the average duration is 9.7 iterations and the yielded coverage is lower, equal to 0.1784, which resulted in reduction of the A7's rank.

In case of the strategy A15 which for the 10% network is ranked third, it does not occur on the real network top-twenty list, and on the remaining synthetic networks it is below the first top-ten. This is an interesting difference, which can be further analyzed with the use of the sensitivity analysis (see Fig. 5). In case of the 10% network, the strategy is slightly supported by Par1 criterion. If the weight of criterion Par2 was increased, the strategy A15 would significantly drop in the ranking, down to rank 17. On the other hand, if the weight of the Par3 criterion became insignificant, strategy A15 would be ranked 10th. Regarding the efficiency rankings, Eff5 supports the strategy A15 (rank 11 to rank 1 increase when Eff5 weight increases from 1 to 100) and Eff4 is in conflict with A15 (rank 1 to rank 6 decrease when Eff4 weight increases from 1 to

TABLE I
KULLBACK-LEIBLER DIVERGENCE MEASURE FOR THE SELECTED SYNTHETIC NETWORKS

Expected %	Network	Num. of nodes	Perc. of nodes	Num. of edges	Perc. of edges	KLD
10	BA, $m = 4$	885	0.100045218%	3530	0.110870316%	0.000935498
30	BA, $m = 5$	2654	0.300022609%	13255	0.416313326%	0.000800703
50	BA, $m = 5$	4423	0.5%	22100	0.694117278%	0.000521317

		BA-885-4							BA-2654-5							BA-4423-5					
Rank	Alt	CCi	SF	PP	Last Iter	Coverage	Measure	Alt	CCi	SF	PP	Last Iter	Coverage	Measure	Alt	CCi	SF	PP	Last Iter	Coverage	Measure
1	A11	0.8202	0.01	0.20	10.3	0.5783	1	A11	0.8202	0.01	0.20	10	0.7049	1	A11	0.8190	0.01	0.20	11.2	0.7026	1
2	A51	0.8005	0.02	0.20	9.5	0.5783	1	A51	0.8005	0.02	0.20	9.8	0.7049	1	A51	0.7958	0.02	0.20	10.7	0.7026	1
3	A15	0.8002	0.01	0.30	8.7	0.8129	1	A7	0.8002	0.01	0.10	12.8	0.2567	1	A15	0.7926	0.01	0.20	11.3	0.7026	2
4	A10	0.7905	0.01	0.20	10.3	0.5783	2	A10	0.7905	0.01	0.20	10	0.7049	2	A7	0.7758	0.01	0.10	13	0.2686	1
5	A50	0.7705	0.02	0.20	9.5	0.5783	2	A91	0.7705	0.03	0.20	9.8	0.7050	1	A50	0.7679	0.02	0.20	10.7	0.7026	2
6	A14	0.7705	0.01	0.30	8.7	0.8129	2	A50	0.7705	0.02	0.20	9.8	0.7049	2	A91	0.7647	0.03	0.20	10.2	0.7028	1
7	A91	0.7689	0.03	0.20	8.8	0.5791	1	A47	0.7689	0.02	0.10	11.5	0.2732	1	A6	0.7606	0.01	0.10	13.7	0.2686	2
8	A55	0.7651	0.02	0.30	7.8	0.8129	1	A6	0.7651	0.01	0.10	12.8	0.2567	2	A47	0.7558	0.02	0.10	11.8	0.2831	1
9	A95	0.7454	0.03	0.30	7.7	0.8129	1	A131	0.7454	0.04	0.20	9.7	0.7051	1	A90	0.7430	0.02	0.10	12.4	0.2828	2
10	A19	0.7414	0.01	0.40	7.6	0.9173	1	A90	0.7414	0.03	0.20	9.8	0.7050	2	A131	0.7379	0.04	0.20	10	0.7033	1
11	A131	0.7414	0.04	0.20	8.6	0.5801	1	A46	0.7414	0.02	0.10	11.5	0.2757	2	A12	0.7372	0.01	0.20	11.3	0.7026	3
12	A90	0.7399	0.03	0.20	8.8	0.5786	2	A15	0.7399	0.01	0.30	8.2	0.8867	1	A90	0.7354	0.03	0.20	10.1	0.7027	2
13	A12	0.7369	0.01	0.20	10.6	0.5783	3	A87	0.7369	0.03	0.10	10.3	0.2898	1	A87	0.7225	0.03	0.10	10.6	0.2961	1
14	A54	0.7365	0.02	0.30	7.8	0.8129	2	A12	0.7365	0.01	0.20	10.1	0.7049	3	A15	0.7205	0.01	0.30	8.3	0.8882	1
15	A7	0.7334	0.01	0.10	9.7	0.1784	1	A8	0.7334	0.01	0.10	12.9	0.2583	3	A8	0.7193	0.01	0.10	13.5	0.2681	3
16	A135	0.7233	0.04	0.30	7.7	0.8129	1	A130	0.7233	0.04	0.20	9.6	0.7055	2	A52	0.7156	0.02	0.20	10.7	0.7027	3
17	A59	0.7205	0.02	0.40	7.2	0.9173	1	A55	0.7205	0.02	0.30	7.9	0.8867	1	A130	0.7124	0.04	0.20	10	0.7031	2
18	A94	0.7197	0.03	0.30	7.8	0.8129	2	A171	0.7197	0.05	0.20	9.3	0.7058	1	A55	0.7050	0.02	0.30	8.1	0.8882	1
19	A47	0.7184	0.02	0.10	8.8	0.2058	1	A14	0.7184	0.01	0.30	8.2	0.8867	2	A86	0.7025	0.03	0.10	10.7	0.2960	2
20	A52	0.7177	0.02	0.20	9.7	0.5783	3	A52	0.7177	0.02	0.20	9.8	0.7049	3	A14	0.7008	0.01	0.30	8.4	0.8882	2

Fig. 4. Visualization of the top 20 alternatives from the TOPSIS evaluation of the campaign strategy planning on synthetic networks.

100).

The sensitivity analysis can also provide information about the overall stability of the obtained solution. In case of the 10% network, the ranking is very stable and the A11 strategy either remains on the winning rank or drops to the second position if the weight of Par2 drops below 40%, Par3 drops below 10%, Eff4 drops below 25%. The only significant change occurs for the Eff5 criterion, where A11 would drop to rank 2 if the Eff5's weight increased to over 60% and even further if the weight increased to over 75%. If exclusively Eff5 was considered, the A11 strategy would be ranked 13th.

Similar stability for Par1-Par3 can be observed for the 30% network, however if the weight of Eff4 increased significantly or the weight of Eff5 increased significantly, A11 would be ranked 6th.

Last, but not least, in case of the 50% synthetic network, A11 would remain ranked 1st regardless of Par1 weight, would drop to 2nd position if Par2 had weight exceeding 90 or would drop to 3rd position if Par3 had negligible weight. In case of Eff4, the stability interval of the obtained solution is $0 - 80$, whilst in case of Eff5 the stability interval is $35 - 100$.

D. Viral marketing campaign strategies planning with network samples

As it was stated in the methodology section of this paper, although synthetic networks allow to minimize the computational efforts, their resemblance to the actual real network might be insufficient. Therefore in the subsequent step of the research, the original real network [49] was sampled, resulting in 3 networks containing 10%, 30% and 50% of the original

network. The sampling procedure was performed with the *snowball.sampling* R function from the netdep R library [50].

The results of the viral marketing campaign planning based on the real network [49] samples are presented in table on Fig. 6. Contrary to the synthetic networks' results, where the same strategy A11 was best in case of all three networks, in case of the samples of the real network, the rankings are more diversified.

When the 50% network is considered, the best-ranked strategy is the strategy A15, based on very low SF, higher PP (0.30), degree measure mediocre process length (14.1 iterations) and satisfying coverage (0.5075). Strategy A15 is followed by strategy A11, which uses smaller PP (0.20), which resulted in simulations in less dynamic process, leading to extending its duration to 17.9 iterations, but reducing the coverage almost by half, to 0.2685. The third position in the ranking belongs to strategy A55, which is based on 0.02 SF and 0.30 PP and results in efficiency results similar to the leading A15 strategy - 13.3 iterations and 0.5106 coverage respectively. However, the costs of such approach are higher due to the increase of the SF. When the 30% and 10% networks are considered, the A15 strategy is ranked second in the former and sixteenth in the latter, which, as mentioned earlier, is in contrast to the observations made for synthetic BA networks.

The equal-weights TOPSIS analysis was followed by a sensitivity analysis of the top 20 strategies for each of the sampled networks (see Fig. 7). An overall observation of the figures allow to see that the rankings for the 50% and 30% networks are much more stable than in case of the 10% network. To illustrate that fact, one can notice that
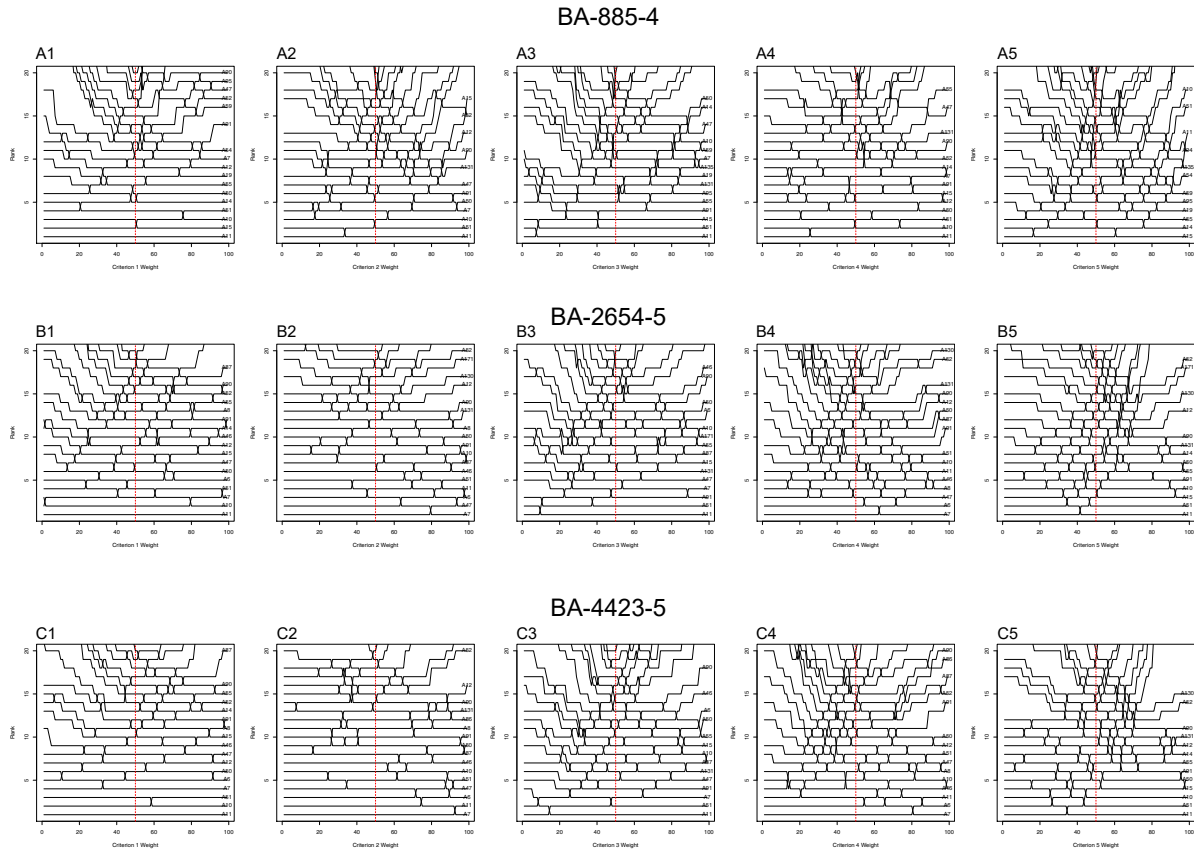
BA-885-4

BA-2654-5

BA-4423-5

Fig. 5. Ranking sensitivity analysis for the top 20 alternatives from the TOPSIS evaluation of the synthetic networks. A1-A5 – 10% network, B1-B5 – 30% network, C1-C5 – 50% network.

in case of figures C1-C5 and B1-B5 only minute or none changes in the rank of the leading alternative can be observed when the weight of Par1-Eff5 criteria are modified. On the other hand, in case of the 10% network, if Par1 criterion weight was decreased significantly, the leading A23 strategy would drop to position 20 (see Fig. 7A1). Moreover, Fig. 7A2 and A5 demonstrate multiple leader changes in case of even slightest fluctuations of the Par2 and Eff5 criteria. When compared to the stability of the rankings obtained for the actual real network (see Fig. 3), this might suggest that a network obtained as a 10% sample of a real network is too small to maintain the stability of evaluation.

E. Comparison of rankings' evaluation accuracy

The research was concluded by a pairwise comparison of rankings based on equal weights for all analyzed networks. In the comparison, the scores and ranks of all strategies for each network were combined into a single table, ordered by the strategy name. This allowed to obtain correlation matrices for all the networks, presenting how correlated are the ranks (Table IV-E) and scores (Table IV-E) for each pair of networks.

The analysis of the correlation matrices allows to observe that the rankings for BA networks are highly correlated to the ranking for the real network with $0.9390 - 0.9799$ correlation

coefficient for scores and $0.9631 - 0.9800$ coefficient for ranks, which means that the relation between them is almost linear. In turn, for the sampled networks, only the ranking for the 50% network achieved high correlation coefficient with the real network, equal to 0.8797 for scores and 0.9222 for ranks. This shows, that the results of the evaluation for the real network and the 50% sampled network are very similar, yet the computational power required to perform the evaluation is significantly smaller. On the other hand, the correlation coefficient values for scores and ranks for the 30% network are much lower, i.e. 0.6043 and 0.6837 respectively, and for the 10% even lower, i.e. 0.4171 and 0.4629 respectively. Such positive yet low values of correlation coefficients indicate there is a positive relation between the rankings obtained for the real network and its 10% and 30% snowball samples. However, the margin of error there might be too high to base the actual campaign on the strategies obtained for such small network samples.

V. Conclusions

Nowadays, when over 45% of the world population are active social media users [51], information spreading in complex social networks begins to bring better results than traditional online advertising campaigns. Online marketers have begun

		Snowball Sample 10%							Snowball Sample 30%							Snowball Sample 50%						
Rank	Alt	CCi	SF	PP	Last Iter	Coverage	Measure	Alt	CCi	SF	PP	Last Iter	Coverage	Measure	Alt	CCi	SF	PP	Last Iter	Coverage	Measure
1	A23	0.6298	0.01	0.50	9.1	0.1958	1	A19	0.7257	0.01	0.40	14.5	0.4418	1	A15	0.7521	0.01	0.30	14.1	0.5075	1
2	A27	0.6293	0.01	0.60	9.1	0.2434	1	A15	0.7137	0.01	0.30	16.3	0.2755	1	A11	0.7491	0.01	0.20	17.9	0.2685	1
3	A19	0.6266	0.01	0.40	9	0.1523	1	A59	0.7079	0.02	0.40	13.6	0.4448	1	A55	0.7302	0.02	0.30	13.3	0.5106	1
4	A22	0.6262	0.01	0.50	10.7	0.2018	2	A23	0.7003	0.01	0.50	12.7	0.5560	1	A10	0.7281	0.01	0.20	17.3	0.2748	2
5	A26	0.6213	0.01	0.60	10.2	0.2495	2	A18	0.6990	0.01	0.40	14.1	0.4424	2	A14	0.7258	0.01	0.30	13.9	0.5085	2
6	A18	0.6206	0.01	0.40	10.2	0.1590	2	A55	0.6958	0.02	0.30	14.7	0.2816	1	A51	0.7183	0.02	0.20	14.8	0.2841	1
7	A35	0.6182	0.01	0.80	9.4	0.3411	1	A14	0.6954	0.01	0.30	16.5	0.2781	2	A54	0.7064	0.02	0.30	13.3	0.5109	2
8	A31	0.6159	0.01	0.70	8.4	0.2896	1	A63	0.6832	0.02	0.50	12.1	0.5584	1	A50	0.7034	0.02	0.20	15.2	0.2883	2
9	A67	0.6145	0.02	0.60	8.7	0.2445	1	A99	0.6801	0.03	0.40	12.4	0.4509	1	A95	0.6982	0.03	0.30	12.3	0.5148	1
10	A30	0.6118	0.01	0.70	9.8	0.2959	2	A58	0.6792	0.02	0.40	13.2	0.4461	2	A91	0.6965	0.03	0.20	13.8	0.3017	1
11	A39	0.6101	0.01	0.90	10.1	0.3828	1	A95	0.6765	0.03	0.30	13.6	0.2963	1	A12	0.6912	0.01	0.20	17.3	0.2702	3
12	A34	0.6084	0.01	0.80	10.2	0.3485	2	A22	0.6739	0.01	0.50	12.4	0.5570	2	A16	0.6908	0.01	0.30	14.4	0.5077	3
13	A62	0.6084	0.02	0.50	9.6	0.2087	2	A27	0.6661	0.01	0.60	11.5	0.6388	1	A90	0.6835	0.03	0.20	14.4	0.3022	2
14	A14	0.6083	0.01	0.30	10.2	0.1050	2	A20	0.6658	0.01	0.40	14.8	0.4409	3	A52	0.6813	0.02	0.20	16.9	0.2762	3
15	A63	0.6079	0.02	0.50	8.1	0.1974	1	A103	0.6631	0.03	0.50	11.6	0.5624	1	A56	0.6744	0.02	0.30	13.9	0.5104	3
16	A15	0.6062	0.01	0.30	8	0.1031	1	A54	0.6609	0.02	0.30	13.4	0.2851	2	A16	0.6723	0.01	0.40	10.2	0.6304	1
17	A66	0.6044	0.02	0.60	9.4	0.2547	2	A139	0.6597	0.04	0.40	12.1	0.4573	1	A131	0.6723	0.04	0.20	13.1	0.3126	1
18	A75	0.6041	0.02	0.80	8.9	0.3423	1	A16	0.6588	0.01	0.30	16.4	0.2732	3	A94	0.6721	0.03	0.30	12.2	0.5151	2
19	A107	0.6015	0.03	0.60	8.4	0.2527	1	A62	0.6582	0.02	0.50	12	0.5598	2	A135	0.6715	0.04	0.30	11.8	0.5174	1
20	A102	0.5998	0.03	0.50	9.5	0.2207	2	A98	0.6507	0.03	0.40	12.1	0.4523	2	A92	0.6665	0.03	0.20	16.4	0.2825	3

Fig. 6. Visualization of the top 20 alternatives from the TOPSIS evaluation of the campaign strategy planning on the real network [49] samples.

TABLE II
CORRELATION MATRIX BETWEEN THE RANKS OF EACH OF THE ANALYZED NETWORKS.

Rank	Real	BA-885-4	BA-2654-5	BA-4423-5	SS 10%	SS 30%	SS 50%
Real	x	0.9631	0.9794	0.9800	0.4629	0.6837	0.9222
BA-885-4	0.9631	x	0.9840	0.9812	0.4806	0.7760	0.9703
BA-2654-5	0.9794	0.9840	x	0.9980	0.3809	0.6706	0.9289
BA-4423-5	0.9800	0.9812	0.9980	x	0.3647	0.6585	0.9191
SS 10%	0.4629	0.4806	0.3809	0.3647	x	0.8227	0.6159
SS 30%	0.6837	0.7760	0.6706	0.6585	0.8227	x	0.8718
SS 50%	0.9222	0.9703	0.9289	0.9191	0.6159	0.8718	x

TABLE III
CORRELATION MATRIX BETWEEN THE SCORE VALUES OF EACH OF THE ANALYZED NETWORKS.

CCi	Real	BA-885-4	BA-2654-5	BA-4423-5	SS 10%	SS 30%	SS 50%
Real	x	0.9390	0.9749	0.9799	0.4171	0.6043	0.8797
BA-885-4	0.9390	x	0.9757	0.9729	0.4954	0.7730	0.9688
BA-2654-5	0.9749	0.9757	x	0.9974	0.3807	0.6373	0.9152
BA-4423-5	0.9799	0.9729	0.9974	x	0.3674	0.6266	0.9049
SS 10%	0.4171	0.4954	0.3807	0.3674	x	0.8204	0.6043
SS 30%	0.6043	0.7730	0.6373	0.6266	0.8204	x	0.8480
SS 50%	0.8797	0.9688	0.9152	0.9049	0.6043	0.8480	x

to invest greater effort into seeding information into social networks and providing incentives to increase the information propagation probability within the networks. These increased efforts have opened the research area for providing evaluation of various social network advertising campaign strategies as well as supporting the process of their planning.

The approach presented in this paper provides a framework for multi-criteria planning of viral marketing campaigns in social networks and their evaluation, in which various preferences and criteria of the marketer are taken into account. The example criteria provided in this paper allow to choose the satisfactory campaign strategy considering the costs related to the seeding of the information and providing incentives to

increase its propagation probability in relation to their effect on the process dynamics and obtained coverage.

The authors' contributions in this paper include:

- multi-criteria framework for evaluation of viral marketing campaigns in social networks;
- simulation engine and usage of synthetic network models and real network samples of limited size allowed to provide a viral marketing campaigns planning tool of reduced computational requirements;
- an example set of criteria was provided that allows to choose a satisfactory viral marketing campaign strategy based on multi-criteria consideration of its costs, dynamics and coverage;

Snowball Sample 10%

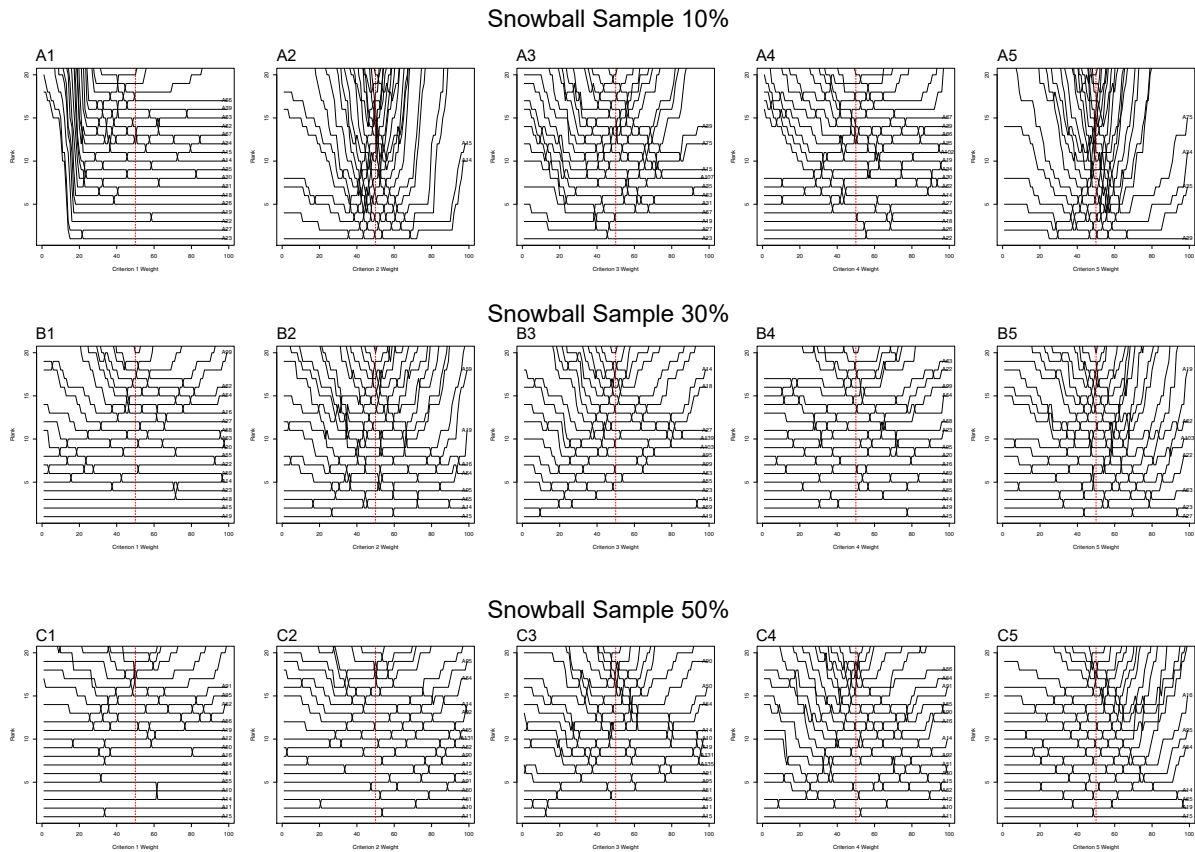## Snowball Sample 30%

Snowball Sample 50%

Fig. 7. Ranking sensitivity analysis for the top 20 alternatives from the TOPSIS evaluation of the real network [49] samples. A1-A5 – 10% network, B1-B5 – 30% network, C1-C5 – 50% network.

- the strategies' evaluation accuracy was compared between a full-size real network and a set of reduced-size synthetic and sample networks derived from the original network.

In practical terms, the empirical study has shown that while the synthetic networks, which were selected based on their Kullback-Leibler divergence, provided very similar results to the real network even when as little as 10% of nodes were used, in case of the sampled networks obtained with the snowball sampling approach provided satisfactory results only when the number of nodes was still relatively high. Also, while the rankings obtained from synthetic networks were stable, there was little stability of the rankings from the snowball sample networks.

All in all, the research has identified possible areas of improvement and future works. First of all, a more numerous set of sizes of sample network could be studied to verify how the network size affects its rankings' correlation to the real network's rankings. Secondly, only snowball sampling approach was used in the research. It would be beneficial to explore networks obtained with other sampling approaches. Last, but not least, the list of criteria could be expanded to allow more precise adjustment of the selected strategy to the marketer's needs.

VI. ACKNOWLEDGMENTS

REFERENCES

[1] W. Chmielarz and O. Szumski, "Digital distribution of video games-an empirical study of game distribution platforms from the perspective of polish students (future managers)," in *Information Technology for Management: Emerging Research and Applications.* Springer, 2018, pp. 136–154.

[2] D. J. Watts, J. Peretti, and M. Frumin, *Viral marketing for the real world.* Harvard Business School Pub., 2007.

[3] E. Ziemba, *Towards a sustainable information society: People, business and public administration perspectives.* Cambridge Scholars Publishing, 2016.

[4] K. Szopik-Depczynska, A. Kedzierska-Szczepaniak, K. Szczepaniak, K. Cheba, W. Gajda, and G. Ioppolo, "Innovation in sustainable development: an investigation of the EU context using 2030 agenda indicators," *Land Use Policy*, vol. 79, pp. 251–262, Dec. 2018. doi: 10.1016/j.landusepol.2018.08.004. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0264837718306203

[5] W. Chmielarz and O. Szumski, "Analysis of users of computer games," in *2016 Federated Conference on Computer Science and Information Systems (FedCSIS).* IEEE, 2016, pp. 1139–1146.

[6] O. Hinz, B. Skiera, C. Barrot, and J. U. Becker, "Seeding strategies for viral marketing: An empirical comparison," *Journal of Marketing*, vol. 75, no. 6, pp. 55–71, 2011.

[7] J. Tang, M. Musolesi, C. Mascolo, V. Latora, and V. Nicosia, "Analysing information flows and key mediators through temporal centrality metrics," in *Proceedings of the 3rd Workshop on Social Network Systems*. ACM, 2010, p. 3.

[8] 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*, vol. 2, no. 2, pp. 65–83, 2015.

[9] K. Kandhway and J. Kuri, "How to run a campaign: Optimal control of sis and sir information epidemics," *Applied Mathematics and Computation*, vol. 231, pp. 79–92, 2014.

[10] 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*. ACM, 2003, pp. 137–146.

[11] E. M. Rogers, *Diffusion of innovations*. Simon and Schuster, 2010.

[12] A. Karczmarczyk, J. Jankowski, and J. Wątróbski, "Multi-criteria decision support for planning and evaluation of performance of viral marketing campaigns in social networks," *PLOS ONE*, vol. 13, no. 12, p. e0209372, Dec. 2018. doi: 10.1371/journal.pone.0209372. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0209372

[13] E. Ziemba, "The contribution of ict adoption to the sustainable information society," *Journal of Computer Information Systems*, vol. 59, no. 2, pp. 116–126, 2019.

[14] G. Bello-Orgaz, J. J. Jung, and D. Camacho, "Social big data: Recent achievements and new challenges," *Information Fusion*, vol. 28, pp. 45–59, 2016.

[15] J. Wątróbski, E. Ziemba, A. Karczmarczyk, and J. Jankowski, "An index to measure the sustainable information society: the polish households case," *Sustainability*, vol. 10, no. 9, p. 3223, 2018.

[16] J. Jankowski, J. Hamari, and J. Wątróbski, "A gradual approach for maximising user conversion without compromising experience with high visual intensity website elements," *Internet Research*, vol. 29, no. 1, pp. 194–217, 2019.

[17] R. Pfitzner, A. Garas, and F. Schweitzer, "Emotional divergence influences information spreading in twitter." *ICWSM*, vol. 12, pp. 2–5, 2012.

[18] R. Michalski, T. Kajdanowicz, P. Bródka, and P. Kazienko, "Seed selection for spread of influence in social networks: Temporal vs. static approach," *New Generation Computing*, vol. 32, no. 3-4, pp. 213–235, 2014.

[19] C. Kiss and M. Bichler, "Identification of influencers: measuring influence in customer networks," *Decision Support Systems*, vol. 46, no. 1, pp. 233–253, 2008.

[20] Y. Liu-Thompkins, "Seeding viral content: The role of message and network factors," *Journal of Advertising Research*, vol. 52, no. 4, pp. 465–478, 2012.

[21] L. Seeman and Y. Singer, "Adaptive seeding in social networks," in *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*. IEEE, 2013, pp. 459–468.

[22] 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*, vol. 7, no. 1, p. 891, 2017.

[23] J. Jankowski, "Dynamic rankings for seed selection in complex networks: Balancing costs and coverage," *Entropy*, vol. 19, no. 4, p. 170, 2017.

[24] 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*, vol. 10, no. 12, p. e0145283, 2015.

[25] J.-X. Zhang, D.-B. Chen, Q. Dong, and Z.-D. Zhao, "Identifying a set of influential spreaders in complex networks," *Scientific reports*, vol. 6, p. 27823, 2016.

[26] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature physics*, vol. 6, no. 11, p. 888, 2010.

[27] C. Granell, S. Gómez, and A. Arenas, "Competing spreading processes on multiplex networks: awareness and epidemics," *Physical review E*, vol. 90, no. 1, p. 012808, 2014.

[28] C. Granell, S. Gómez, and A. Arenas, "Dynamical interplay between awareness and epidemic spreading in multiplex networks," *Physical review letters*, vol. 111, no. 12, p. 128701, 2013.

[29] X. Wei, N. C. Valler, B. A. Prakash, I. Neamtiu, M. Faloutsos, and C. Faloutsos, "Competing memes propagation on networks: A network

[30] science perspective," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 6, pp. 1049–1060, 2013.

[30] M. Bampo, M. T. Ewing, D. R. Mather, D. Stewart, and M. Wallace, "The effects of the social structure of digital networks on viral marketing performance," *Information systems research*, vol. 19, no. 3, pp. 273–290, 2008.

[31] J. Y. Ho and M. Dempsey, "Viral marketing: Motivations to forward online content," *Journal of Business research*, vol. 63, no. 9-10, pp. 1000–1006, 2010.

[32] S. Stieglitz and L. Dang-Xuan, "Emotions and information diffusion in social media: sentiment of microblogs and sharing behavior," *Journal of management information systems*, vol. 29, no. 4, pp. 217–248, 2013.

[33] A. Dobele, A. Lindgreen, M. Beverland, J. Vanhamme, and R. Van Wijk, "Why pass on viral messages? because they connect emotionally," *Business Horizons*, vol. 50, no. 4, pp. 291–304, 2007.

[34] C. Camarero and R. San José, "Social and attitudinal determinants of viral marketing dynamics," *Computers in Human Behavior*, vol. 27, no. 6, pp. 2292–2300, 2011.

[35] J. Berger and K. L. Milkman, "What makes online content viral?" *Journal of marketing research*, vol. 49, no. 2, pp. 192–205, 2012.

[36] A. Rezvanian, B. Moradabadi, M. Ghavipour, M. M. D. Khomami, and M. R. Meybodi, "Social network sampling," in *Learning Automata Approach for Social Networks*. Springer, 2019, pp. 91–149.

[37] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.

[38] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *nature*, vol. 393, no. 6684, p. 440, 1998.

[39] P. Erdös and A. Rényi, "On random graphs, i," *Publicationes Mathematicae (Debrecen)*, vol. 6, pp. 290–297, 1959.

[40] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[41] J. Wątróbski, K. Małecki, K. Kijewska, S. Iwan, A. Karczmarczyk, and R. Thompson, "Multi-criteria analysis of electric vans for city logistics," *Sustainability*, vol. 9, no. 8, p. 1453, 2017.

[42] P. Ziemba, "Neat f-promethee–a new fuzzy multiple criteria decision making method based on the adjustment of mapping trapezoidal fuzzy numbers," *Expert Systems with Applications*, vol. 110, pp. 363–380, 2018.

[43] J. Wątróbski, J. Jankowski, P. Ziemba, A. Karczmarczyk, and M. Zioło, "Generalised framework for multi-criteria method selection," *Omega*, vol. 86, pp. 107–124, 2019.

[44] J. Wątróbski, J. Jankowski, P. Ziemba, A. Karczmarczyk, and M. Zioło, "Generalised framework for multi-criteria method selection: Rule set database and exemplary decision support system implementation blueprints," *Data in brief*, vol. 22, p. 639, 2019.

[45] J. Wątróbski, J. Jankowski, and Z. Piotrowski, "The selection of multicriteria method based on unstructured decision problem description," in *International Conference on Computational Collective Intelligence*. Springer, 2014, pp. 454–465.

[46] J. Wątróbski and J. Jankowski, "Guideline for mcda method selection in production management area," in *New frontiers in information and production systems modelling and analysis*. Springer, 2016, pp. 119–138.

[47] W. Chmielarz and M. Zborowski, "Analysis of e-banking websites' quality with the application of the topsis method–a practical study," *Procedia computer science*, vol. 126, pp. 1964–1976, 2018.

[48] M. Jankowski, A. Borsukiewicz, K. Szopik-Depczynska, and G. Ioppolo, "Determination of an optimal pinch point temperature difference interval in ORC power plant using multi-objective approach," *Journal of Cleaner Production*, vol. 217, pp. 798–807, Apr. 2019. doi: 10.1016/j.jclepro.2019.01.250. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0959652619302756

[49] M. Ripeanu, I. Foster, and A. Iamnitchi, "Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design," *arXiv:cs/0209028*, Sep. 2002, arXiv: cs/0209028. [Online]. Available: http://arxiv.org/abs/cs/0209028

[50] "Snowball Sampling Function - R Documentation." [Online]. Available: https://www.rdocumentation.org/packages/netdep/versions/0.1.0/topics/snowball.sampling

[51] S. Kemp, "Digital 2019: Global Internet Use Accelerates," Jan. 2019. [Online]. Available: https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates