Proceedings of the 18th Conference on Computer Science and Intelligence Systems pp. 1023–1027
DOI: 10.15439/2023F539
ISSN 2300-5963 ACSIS, Vol. 35

Dynamic SITCOM: an innovative approach to re-identify social network evaluation models

Bartłomiej Kizielewicz
0000-0001-5736-4014
Dept. of Artificial Intelligence Methods and Applied Mathematics,
West Pomeranian University of Technology in Szczecin
ul. Żołnierska 49, 71-210 Szczecin, Poland
Email: bartlomiej-kizielewicz@zut.edu.pl

Jarosław Jankowski
0000-0002-3658-3039
Dept. of Information Systems Engineering,
West Pomeranian University of Technology in Szczecin
ul. Żołnierska 49, 71-210 Szczecin, Poland
Email: jjankowski@zut.edu.pl

***Abstract*—Complex networks attract attention in various scientific fields due to their ability to model real world phenomena and potential for problem-solving. It is essential to evaluate these networks to simulate and solve various issues. Evaluating social networks is challenging due to the unequal status of nodes and their unknown impact on everall characteristics. Existing measures of centrality often need to consider the global structure of the network, which requires the involvement of experts and creates space for multi-criteria decision-making methods usage. Unfortunately, more access to established decision-making models is often needed for various reasons. In this article, we propose an innovative approach called Dynamic Stochastic IdenTifiCation Of Models (Dynamic SITCOM), which considers the preferences of characteristic objects and the characteristic values of criteria, enabling the re-identification of multi-criteria decision models. The approach evaluates nodes in Facebook's complex social network, focusing on prediction accuracy using similarity measures and Mean Absolute Error. The study shows that a stable decision model can be created and applied to evaluate nodes in complex networks.**

I. Introduction

COMPLEX networks have attracted the interest of researchers from various scientific fields, such as biology, sociology, physics, and computer science [1]. Their occurrence in a wide variety of fields makes evaluating and analyzing these networks of great importance. Evaluating complex networks is essential in many fields because it allows for simulating and solving various problems [2].

For example, assessing the importance of nodes in the context of power networks makes it possible to identify critical points whose failure could lead to network shutdown. In the case of communication networks, assessing the importance of nodes makes it possible to optimally maintain connections and prevent disruptions in the flow of information. Complex networks also have applications in preventing the spread of rumors or diseases [2]. By identifying and assessing the importance of crucial nodes, it is possible to influence the control and limit the propagation of such phenomena. In addition, complex networks are widely present in the social domain, where assessing the importance of nodes is crucial for identifying opinion leaders, influential individuals, or experts in a community.

Due to the growing popularity of social media, it has become one of the most effective marketing tools. Using visual content on platforms such as Twitter, Facebook, and Instagram can help companies build their image and increase the reach of their brand. Social media allows companies to connect directly with customers and monitor product and service feedback [3]. However, one of the main problems with social networks is viral marketing, which involves using social networks to spread information about a product or service through users who pass the information on to their friends.

In addition, companies promoting themselves on social media have begun using virtual influencers to advertise products and services. Through them, companies can reach new audiences and increase the reach of their brand. However, the use of virtual influencers is controversial among consumers, who believe it is a scam and lacks authenticity.

Therefore, evaluating social networks is particularly important to identify critical nodes that play a crucial role in the network. This, in turn, allows us to control and limit the spread of rumors or use them for marketing purposes [4]. By identifying key nodes, we can also increase the reach of positive information and make positive community changes.

Due to the statuses of nodes found in complex networks, which are unequal and different, a problem arises in evaluating them. The main methods used are methods of evaluating nodes based on measures of node centrality. The most popular centrality measures used to evaluate complex networks are degree centrality, inter-node centrality, proximity centrality, PageRank centrality [5], Katz centrality [6], and k-shell [7].

Unfortunately, although centrality criteria are widely used, they have some shortcomings and deficiencies [2]. Measures of node centrality, such as degree centrality, do not always consider the global structure of the network [8], [9]. Therefore, it is common to use the knowledge of domain experts to evaluate key nodes based on the information gathered by selected centrality metrics. Multi-criteria decision analysis methods are also used, using some aggregation of centrality metrics to determine network node ratings [10].

Therefore, this article proposes a new Dynamic SITCOM approach to re-identify the decision model based on the evaluated decision variants. The main novelty of the Dynamic

SITCOM approach is the two-stage optimization. In the case of the baseline Stochastic IdenTifiCation Of Models (SITCOM) approach, optimization is based only on the preferences of characteristic objects that reflect the decision maker's preferences [11], [12]. In contrast, the proposed approach also optimizes the characteristic values responsible for the position of characteristic objects in the decision grid. This leads to the possibility of considering more non-linear problems.

The proposed approach will be applied to the problem of evaluating Facebook nodes in a complex network. For this problem, the expert evaluates nodes with a specific model based on four criteria reflected in the form of centrality metrics. This model is unavailable, so the Dynamic SITCOM approach used is to re-identify it. The study focuses on the accuracy of this approach using similarity measures of rankings and Mean Absolute Error (MAE).

The rest of the article is organized as follows. Section II presents the state of the art of MCDA/MCDM methods related to the topic of node evaluation in complex networks and a brief introduction of Stochastic IdenTifiCation Of Models (SITCOM). Section III presents a proposal for the Dynamic SITCOM approach. Section IV presents research related to the accuracy of the Dynamic SITCOM approach in the problem of evaluating the nodes of the Facebook complex network. The V section presents conclusions and future research directions.

II. Preliminaries

A. State of the art

Multi-criteria decision analysis/Multi-criteria decision-making (MCDA/MCDM) methods assess node importance in complex networks by using centrality metrics, such as node degree, betweenness, and closeness. These methods aggregate multiple criteria to provide a comprehensive evaluation of node significance, aiding in the identification of crucial nodes for network performance. Table I presents examples of the use of MCDA/MCDM methods for complex network evaluations. Khaoula et al. proposed a novel seed-centered approach based on TOPSIS and the k-means algorithm to find communities in a social network [13]. Zhang and Ng proposed a ranking method based on entropy weights and the TOPSIS approach, named EWM-TOPSIS, to evaluate the criticality of nodes considering various node characteristics in complex public transportation networks (PTNs) [14]. Lu used the TOPSIS method to evaluate and compare the ARPA network and the standard IEEE 39 bus system [15]. Meng et al. used the WTOPSIS approach to evaluate complex networks in urban rail transit (URT) [16]. Mi et al. used the VIKOR approach to evaluate a road network with 28 intersections in Shenzhen [17]. Kharanagh et al. proposed using SAW, TOPSIS, and ELECTRE I approaches to analyze social networks in water resources management [18]. Lin et al. used the CRITIC approach to assess the importance of nodes in reconfiguring the electricity grid backbone network [19].

B. Stochastic IdenTifiCation Of Models (SITCOM)

Stochastic IdenTifiCation Of Models (SITCOM) is a new approach to re-identify a decision model based on evaluated decision variants [11], [12]. This approach's operation mechanism is based on the logic of the selected stochastic optimization algorithm and Characteristic Object METhod (COMET). The stochastic algorithm determines the preferences of the Characteristic Objects (CO), which in the case of the COMET method, represents the preferences of the decision maker. Then, when selecting appropriate values of characteristic object preferences is over, it is possible to evaluate new decision variants. A full description of the algorithm can be found in the initial papers [11], [12].

III. Dynamic SITCOM

In this article, we propose to extend the above SITCOM approach with additional optimization. Since the base SITCOM approach only uses characteristic object preference values for optimization, the model may not consider some non-linearity occurring in expert knowledge. Therefore, the present proposition is based on the characteristic objects building factor, i.e., the characteristic values of the criteria. The characteristic values are mainly responsible for the model's grid and irregularity. In addition, the starting and ending values included in the vector of characteristic values of the COMET method are responsible for the boundaries of the model. Therefore, in the proposed approach, in addition to optimizing the preference of characteristic objects, we will focus on optimizing the middle characteristic values of the model.

The proposed method is based on a two-stage optimization. The first optimization, as in the case of the original SITCOM, will be based on the search for the best possible preferences of characteristic objects. The second optimization, on the other hand, will focus on the search for the best possible middle values for the characteristic values of the considered criteria. Due to the grid change, a loop was applied to query the optimized models to change the preferences of the characteristic objects with the newly found middle values for the characteristic values and vice versa.

IV. Study case

In this paper, we will focus on investigating the accuracy of the Dynamic SITCOM approach based on the problem of evaluating nodes of complex networks. First, it will get the selected dataset presented, and the study will be conducted in the next section.

A. Dataset description

The selected dataset concerns a complex network, which consists of nodes that are Facebook profiles. This dataset is anonymized and derived from the paper [20]. It consists of 4039 nodes and 88234 edges that connect the selected nodes. The network and the degree of the given nodes can be represented by Fig. 1.

TABLE I
EXAMPLES OF THE USE OF MCDA/MCDM METHODS IN EVALUATING NODES OF COMPLEX NETWORKS.

MCDA approach	No. of nodes	No. of criteria	Problem	Year	Reference
TOPSIS	4039	4	Evaluation of nodes from Facebook's network	2023	[13]
EWM–TOPSIS	95	3	Evaluation of the MTR network in Hong Kong	2021	[14]
TOPSIS	21	7	Evaluation of the ARPA network and the standard IEEE 39-bus system	2020	[15]
WTOPSIS	118, 132, 166	4	Evaluation of the Shenzhen Metro System	2020	[16]
VIKOR	28	3	Analysis of traffic safety at intersections	2020	[17]
SAW, TOPSIS, ELECTRE I	54, 30, 32, 26	12	Social network analysis of water resources management	2019	[18]
CRITIC	66	7	Evaluation of the Guangdong power system in China	2017	[19]

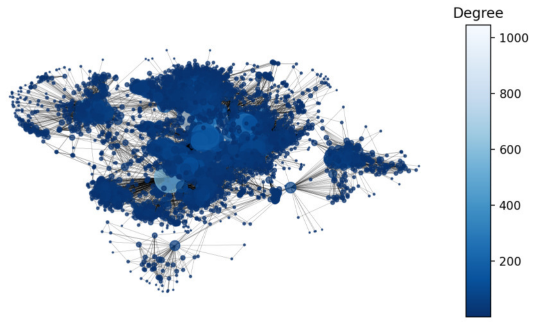

Fig. 1. Facebook's complex network of anonymized profiles [20].

Due to the problem of evaluating the nodes of the present network, four metrics were selected to serve as criteria. The selected network centrality metrics are degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality.

A decision matrix incorporating criteria and preferences, including expert ratings, will guide the re-identification of the decision model. It utilizes min-max normalization for all criteria, employing stochastic optimization in the training process. Table II displays the first ten decision variants with normalized values for criteria and preferences.

TABLE II
EXAMPLE 10 ALTERANTIVES.

$\mathbf{A_i}$	$\mathbf{C_1}$	$\mathbf{C_2}$	$\mathbf{C_3}$	$\mathbf{C_4}$
A_1	3.044754e-01	0.331418	0.000356	0.622104
A_2	5.792237e-06	0.015326	0.000006	0.295339
A_3	1.580590e-07	0.008621	0.000002	0.294918
A_4	3.506768e-06	0.015326	0.000007	0.295339
A_5	3.829891e-07	0.008621	0.000002	0.294918
A_6	4.590804e-06	0.011494	0.000012	0.295098
A_7	5.106522e-08	0.004789	0.000002	0.294678
A_8	3.544060e-04	0.018199	0.000269	0.342924
A_9	5.744837e-07	0.006705	0.000002	0.294798
A_{10}	3.424270e-05	0.053640	0.000023	0.297750
...	...	...	...	...

B. Research on the accuracy of the approach

This research case will investigate the accuracy of the proposed Dynamic SITCOM approach. For this purpose, the selected stochastic algorithm for re-identification of the multi-criteria model is the genetic algorithm. For optimization in the genetic algorithm in determining the preferences of characteristic objects and the means of characteristic values, 50 chromosomes were selected. On the other hand, for each criterion, the characteristic values were defined as a set of 0, 0.5, 1 because the criterion values were normalized. The implementations used in this study are from the *mealpy* library (genetic algorithm: `BaseGA`) [21] and the *pymcdm library* (COMET method) [22]. The related study evaluates subsets of test collections derived from tenfold cross-validation. The entire set in this study is divided into two parts, i.e., the training part (90 percent of the original set) and the testing part (10 percent of the original set). This division made was 10 times, where the selection of Folds is generated using the sklearn library. The subsets of the train and test sets are drawn 1,000 times and have 10,15,25,50,100 decision variants. These subsets evaluated were to use the learned SITCOM model on the selected Fold training set. Their evaluation is then compared with a reference evaluation determined subjectively by the expert using the MAE measure. In addition, the output evaluation from the learned decision variant model and the expert evaluation ranked is, and their similarity examined is using the r_w and WS measures. The training set has high quality and was similar to the test set, so it was decided to present only the research on the test set.

The results of the test set presented are in Tables III, IV and V for different numbers of randomly selected alternatives: 10, 15, 25, 50, and 100. The tests repeated were 1000 times. The Tables contain information on accuracy, expressed by the r_w, WS, and MAE metrics. Analyzing the MAE metric in the present case, the minimum MAE was smallest for 10 alternatives and was 0.002259, while the largest minimum error occurred for 100 alternatives and was 0.003226. The average values of the MAE for all the numbers of alternatives considered ranged from 0.021608 to 0.021710. As for the maximum values, the largest value was reached for 10 alternatives, while the smallest value occurred for 100 alternatives. The standard deviation indicates the spread of the results around the mean value of the MAE, which was approximately 0.012 for all cases.

TABLE III
MAE VALUES FOR SELECTED 1000 DRAWS OF GIVEN NUMBERS OF ALTERNATIVES FROM THE TEST SET.

No. of alts.	Min	Mean	Max	Std
10	0.002259	0.021665	0.059992	0.012677
15	0.002419	0.021710	0.057319	0.012469
25	0.002580	0.021608	0.055091	0.012236
50	0.002832	0.021657	0.050384	0.012166
100	0.003226	0.021674	0.048271	0.012082

The results of the r_w ranking similarity metric for randomly selected alternatives from the test set shown are in the following table. The table contains the minimum, mean, and maximum values of the r_w metric and the standard deviation. The table shows that the smallest minimum values of the r_w metric achieved were for 10 random alternatives, where they amounted to 0.388430. In comparison, the most significant minimum values occurred for 100 alternatives, where they reached a value of 0.929653. The average values of the r_w metric for all the considered numbers of alternatives range from 0. 985462 to 0.992024. Virtually all maximum values obtained were equal to 1, except for 100 alternatives, where the highest value was 0.999972. The standard deviation indicates the spread of results around the average value of the r_w metric, which ranged approximately from 0.010770 to 0.037956 for the different cases.

TABLE IV
r_w VALUES FOR SELECTED 1000 DRAWS OF GIVEN NUMBERS OF ALTERNATIVES FROM THE TEST SET.

No. of alts.	Min	Mean	Max	Std
10	0.388430	0.985462	1.000000	0.037956
15	0.638393	0.987587	1.000000	0.028707
25	0.656923	0.989946	1.000000	0.019518
50	0.871222	0.991282	1.000000	0.014129
100	0.929653	0.992024	0.999972	0.010770

Examining Table V, it can be seen that the minimum values of the WS metric for the various numbers of randomly selected alternatives range from 0.418711 to 0.888070. The average values of the WS metric for all the numbers of alternatives considered ranged from 0. 978925 to 0.994398. All the obtained maximum values of the WS metric equal 1. The standard deviation shows the spread of the results around the average value of the WS metric, which ranged from 0.007709 to 0.048627 for the different cases.

TABLE V
WS VALUES FOR SELECTED 1000 DRAWS OF GIVEN NUMBERS OF ALTERNATIVES FROM THE TEST SET.

No. of alts.	Min	Mean	Max	Std
10	0.418711	0.978925	1.0	0.048627
15	0.493393	0.980259	1.0	0.041311
25	0.511516	0.983272	1.0	0.032187
50	0.571031	0.988155	1.0	0.020861
100	0.888070	0.994398	1.0	0.007709

The visualizations associated with the r_w, WS, and MAE measures for randomly selected alternatives from the test set, repeated 1,000 times, are shown in Figs. 2, 3, and 4. Analyzing the MAE measure, the most accurate model was obtained for Fold number 2, while the least accurate model obtained was for Fold number 10. The smallest number of outliers was observed for Fold number 8, while the most significant was for Fold number 9. Turning to the similarity measure of rankings r_w, the lowest similarity of rankings occurred for Fold numbers 6 and 10, while the highest similarity of rankings observed was for Fold number 7.

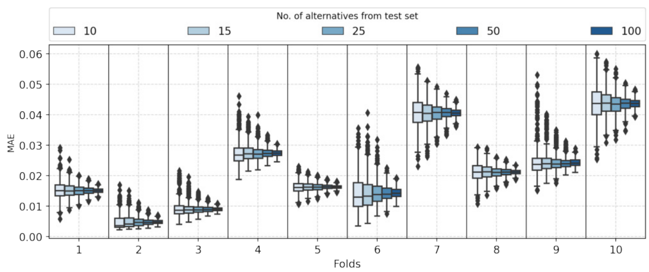

Fig. 2. Distributions of MAE values for selected 1000 draws of given numbers of alternatives from the test set for 10-fold crosvalidation.

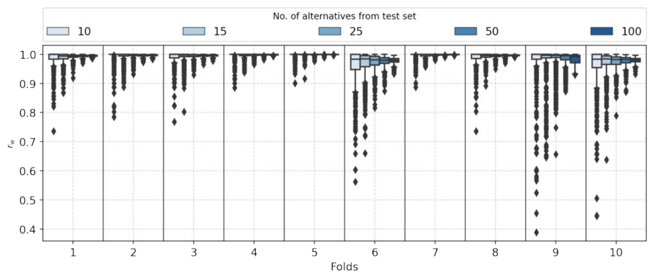

Fig. 3. Distributions of r_w values for selected 1000 draws of given numbers of alternatives from the test set for 10-fold crosvalidation.

V. CONCLUSIONS AND FUTURE RESEARCH

The study related to Dynamic SITCOM shows that it is possible to create a stable decision model for complex network nodes. Several conclusions can be taken by analyzing the presented research results related to the proposed Dynamic SITCOM approach. The first is that the larger the number of randomly selected alternatives, the smaller the value of the

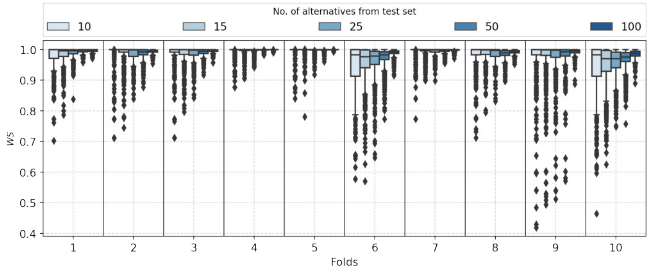

Fig. 4. Distributions of WS values for selected 1000 draws of given numbers of alternatives from the test set for 10-fold crosvalidation.

maximum MAE, suggesting that a more significant number of alternatives leads to better prediction accuracy. However, the average values of the MAE for all the numbers of alternatives considered are very close, indicating the stability of the model.

A similar trend observed is for the similarity measures of the r_w and WS rankings, where a more significant number of alternatives has larger minimum, average and maximum values. The maximum values are close to 1 for all cases, meaning the model represents reality well.

The conclusion is that a more significant number of randomly selected nodes presented as decision alternatives lead to better prediction accuracy and a more accurate reflection of the decision maker's preferences in the Dynamic SITCOM approach. At the same time, the models achieve stable results, as evidenced by the low variability of mean values and low standard deviation.

Future research directions could focus on considering more characteristic values for optimization. In addition, compromise solutions should be considered for characteristic objects with similar criterion values. It is also necessary to consider the applicability of the Dynamic SITCOM approach to other multi-criteria problems, such as selecting suppliers or creating a stable recommendation system.

REFERENCES

[1] Z. Liu, C. Jiang, J. Wang, and H. Yu, "The node importance in actual complex networks based on a multi-attribute ranking method," *Knowledge-Based Systems*, vol. 84, pp. 56–66, 2015.

[2] Y. Yang, L. Yu, Z. Zhou, Y. Chen, T. Kou *et al.*, "Node importance ranking in complex networks based on multicriteria decision making," *Mathematical Problems in Engineering*, vol. 2019, 2019.

[3] L. Wang, Z. Yu, F. Xiong, D. Yang, S. Pan, and Z. Yan, "Influence spread in geo-social networks: a multiobjective optimization perspective," *IEEE Transactions on Cybernetics*, vol. 51, no. 5, pp. 2663–2675, 2019.

[4] A. Zareie, A. Sheikhahmadi, and K. Khamforoosh, "Influence maximization in social networks based on TOPSIS," *Expert Systems with Applications*, vol. 108, pp. 96–107, 2018.

[5] M. Zhang, T. Huang, Z. Guo, and Z. He, "Complex-network-based traffic network analysis and dynamics: A comprehensive review," *Physica A: Statistical Mechanics and its Applications*, p. 128063, 2022.

[6] J. Zhao, T.-H. Yang, Y. Huang, and P. Holme, "Ranking candidate disease genes from gene expression and protein interaction: a Katz-centrality based approach," *PloS one*, vol. 6, no. 9, p. e24306, 2011.

[7] 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, pp. 888–893, 2010.

[8] P. Yang, X. Liu, and G. Xu, "A dynamic weighted TOPSIS method for identifying influential nodes in complex networks," *Modern Physics Letters B*, vol. 32, no. 19, p. 1850216, 2018.

[9] J. Zhang, Q. Zhang, L. Wu, and J. Zhang, "Identifying influential nodes in complex networks based on multiple local attributes and information entropy," *Entropy*, vol. 24, no. 2, p. 293, 2022.

[10] Y. Du, C. Gao, Y. Hu, S. Mahadevan, and Y. Deng, "A new method of identifying influential nodes in complex networks based on TOPSIS," *Physica A: Statistical Mechanics and its Applications*, vol. 399, pp. 57–69, 2014.

[11] B. Kizielewicz, "Towards the identification of continuous decisional model: the accuracy testing in the SITCOM approach," *Procedia Computer Science*, vol. 207, pp. 4390–4400, 2022.

[12] B. Kizielewicz and W. Sałabun, "A new approach to identifying a multi-criteria decision model based on stochastic optimization techniques," *Symmetry*, vol. 12, no. 9, p. 1551, 2020.

[13] A. Khaoula, M. MACHKOUR, and J. ANTARI, "Unsupervised Learning-based New Seed-Expanding Approach using Influential Nodes for Community Detection in Social Networks," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 1, 2023.

[14] Y. Zhang and S. T. Ng, "Identification and quantification of node criticality through EWM–TOPSIS: a study of Hong Kong's MTR system," *Urban Rail Transit*, vol. 7, no. 3, pp. 226–239, 2021.

[15] M. Lu, "Node importance evaluation based on neighborhood structure hole and improved TOPSIS," *Computer Networks*, vol. 178, p. 107336, 2020.

[16] Y. Meng, X. Tian, Z. Li, W. Zhou, Z. Zhou, and M. Zhong, "Exploring node importance evolution of weighted complex networks in urban rail transit," *Physica A: Statistical Mechanics and its Applications*, vol. 558, p. 124925, 2020.

[17] X. Mi, C. Shao, C. Dong, C. Zhuge, and Y. Zheng, "A framework for intersection traffic safety screening with the implementation of complex network theory," *Journal of advanced transportation*, vol. 2020, pp. 1–12, 2020.

[18] S. G. Kharanagh, M. E. Banihabib, and S. Javadi, "An MCDM-based social network analysis of water governance to determine actors' power in water-food-energy nexus," *Journal of Hydrology*, vol. 581, p. 124382, 2020.

[19] Z. Lin, F. Wen, H. Wang, G. Lin, T. Mo, and X. Ye, "CRITIC-based node importance evaluation in skeleton-network reconfiguration of power grids," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 2, pp. 206–210, 2017.

[20] J. Leskovec and J. Mcauley, "Learning to discover social circles in ego networks," *Advances in neural information processing systems*, vol. 25, 2012.

[21] N. V. Thieu and S. Mirjalili, "MEALPY: a Framework of The State-of-The-Art Meta-Heuristic Algorithms in Python," Jun. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.6684223

[22] B. Kizielewicz, A. Shekhovtsov, and W. Sałabun, "pymcdm—The universal library for solving multi-criteria decision-making problems," *SoftwareX*, vol. 22, p. 101368, 2023.