Dear Prof. Jie Lu,
Editor in Chief, Knowledge-Based Systems Journal

We would like to thank you for providing us with the opportunity to revise our paper (Bozorgi et al, "Community-based Influence Maximization in Social Networks under a Competitive Linear Threshold Model"). As per your suggestion, we have thoroughly revised our paper in response to the detailed comments provided by the two anonymous reviewers. Below is given descriptions of how we have addressed each of the reviewer comments.

Reviewer #1

Comment #1: The contribution of the manuscript is unclear for reader. If the main contribution is a competitive version of linear threshold model, a proper experiment should be designed for assessment of the new diffusion model.

Answer: We thank the reviewer for pointing this out. We see the main contributions of our paper as (1) the DCM propagation model, which gives nodes the ability to think about incoming influence in a competitive version of the LT propagation model, and (2) the CI2 algorithm, in which social network community structure is exploited to allow the second competitor to select small seed sets for influence spread. To highlight these contributions better, we have made changes to the **Contributions** subsection on page 3 of the Introduction section. Namely, the first and third items of the contribution list have been changed as follows:

> *List item #1*: "We propose the DCM propagation model, which gives decision-making power to nodes based on the incoming influence in a competitive version of LT propagation model."
> *List item #3*: "We propose the CI2 algorithm to find the minimum number of the most influential nodes for competitor C_2 by the knowledge of the nodes selected by competitor C_1 so that C_2 could achieve more influence spread by spending less budget. Computing the spread of seed nodes is done locally inside communities of the input graph, which causes a remarkable decrease in running time."

An additional contribution of our paper is a detailed investigation of the effects of the d parameter in the DCM model on the quality of seed sets derived by and the running time of the CI2 algorithm (see Figures 3 and 4 and explanations of the results reported in these figures in Section 4.2). As the former was demonstrated in the submitted version of the paper, we have performed additional experiments and added discussion of the results of these experiments in the revised paper to demonstrate the latter.

Comment #2: The descriptions of community detection method and influence maximization are minimal and could be enhanced. So, I carefully investigated for several state-of-the-art community detection method, in order to enrichment of this section, I suggest to cite these state-of-the-art methods:

a. Mirsaleh, M. R., & Meybodi, M. R. (2016). A Michigan memetic algorithm for solving the community detection problem in complex network. Neurocomputing.
b. Khomami, M. M. D., Rezvanian, A., & Meybodi, M. R. (2016). Distributed learning automata-based algorithm for community detection in complex networks. International Journal of Modern Physics B, 30(8), 1650042.
c. Hosseini-Pozveh, M., Zamanifar, K., & Naghsh-Nilchi, A. R. (2016). A community-based approach to identify the most influential nodes in social networks. Journal of Information Science, 0165551515621005.
d. Zhao, Y., Li, S., & Jin, F. (2016). Identification of influential nodes in social networks with community structure based on label propagation. Neurocomputing.

Answer: We very much thank the reviewer for their efforts in helping us to improve our literature citations. We have modified the **Community Detection** subsection in Section 3.3 by to add a new paragraph that briefly explains the approaches used in the first two suggested references. The changed paragraph is:

> "Many approaches have been proposed to solve the community detection problem in online social networks. MLAMA-Net [30] is an evolutionary algorithm, which solves the community detection problem in a network of chromosomes using evolutionary operators and local searches. In MLAMA-Net, each node including a chromosome, which represents the community of the node, and a learning automata, which represents a meme, and saves the histories of the exploitation. Very related to MLAMA-Net, Khomami et al. proposed DLACD [31], which extracts the community structure of complex networks based on distributed learning automata."

As for the final two suggested references, we thought it most appropriate to cite them as References [17] and [18] in the last paragraph of the Introduction section before the **Contributions** subsection.

Comment #3: While three datasets are used for evaluation, they are all real dataset. However, the community structure of network may be different (network with low/high community). It is reasonable to evaluate algorithm on artificial modular networks (LFR benchmark) with different structure (varying mixing parameters in LFR benchmark).

Answer: We thank the reviewer for pointing out this oversight on our part. To investigate the effect of community structure on our algorithm, we generated and did experiments relative to three LFR networks which vary in their mixing parameters. The introduction of these synthetic datasets has necessitated changes in the abstract, the Introduction section, and the first paragraph of Section 4. A complete description of these datasets is given in the (new) second paragraph of Section 4.1 and these datasets are visualized in Figure 2. The added paragraph is as following:

> "To generate our synthetic datasets, we used LFR benchmark [33], which clarify the heterogeneity of the networks in the distributions of node degrees and community sizes. The node degrees and community sizes are taken from power lows distribution with exponents γ and β respectively. By assigning three different values 0.03, 0.08 and 0.15 to parameter μ which is the mixing parameter and setting $N = 1000$ as the number of nodes, $\gamma = 2$ and $\beta = 1$, in-degree of nodes ranging from 0 to 50 with average 15 and the community size between 20 and 50, we generated three different datasets, which are visualized in Figure 2. Mixing parameter determines the fraction of one node's links to other nodes inside its community and nodes outside its community. More specifically, each node shares a fraction of $1 - \mu$ of its links with the nodes inside its community and a fraction μ with other nodes, which belong to the other communities."

To address the effect of community structure on our algorithm, we performed additional experiments relative to the synthetic datasets (whose results are shown in Figure 6) and added a new subsection **The effect of community structure on seed selection** to Section 4.2 of the paper. This new subsection is as follows:

> "To study that how the structure of communities can affect the quality of seed nodes, we ran our algorithm on three synthetic LFR networks [33] varying in their community structures, which is due to the different values assigned to mixing parameter in LFR benchmark. As the mixing parameter would be smaller, the communities are loosely connected to each other and there are few links between nodes in different communities. The results of our runs on LFR networks in Figure 6 show that our algorithm acts better in finding seed nodes in networks, which their community structures are more significant. In the first network with smallest mixing parameter, the influence spread achieved by the extracted seed set is higher than the two other networks with larger mixing parameter values. One of the ways to help our algorithm acts better in networks with less significant community structures is considering the effect of border nodes on influence spread computations, introduced in [14]. Border nodes have at least one link to the other nodes in another community, which allow the spread of influence from

their own community to others and vice versa. As the main point of this paper is to propose DCM propagation model and our aim of using community structure in CI2 algorithm is to improve the running time of finding seed nodes, we will address the issue of border nodes in the future work."

Comment #4: Curiously, the authors should evaluate the results of influence spread with respect to running time.

Answer: ***** FIX THIS ***** you for your attention to this matter. To simulate the thinking ability of the nodes, we have added parameter d to our DCM propagation model. When we run our algorithm, the process of influence propagation would be repeated more times depending on value of parameter d. So, it affects the overall running time of our algorithm that causes an increment in the running time in comparison with other proposed algorithms, which have not considered a situation in which the nodes can think about the incoming influence spread, and their process of finding seed nodes is repeated less than that of ours. Therefore, we believe that comparing our algorithm based on its running time with other algorithms is not reasonable. One of the main reasons, which we get benefit from communities of the graph is to localize the searches to compute the spread values to decrease the running time. In total, by adding parameter d, we increase the quality of found seed nodes and we simulate more realistic model of influence propagation. Moreover, and by localizing the spread value computations, we decrease the running time. Using these two approaches will make our algorithm and propagation model practical for even large datasets. We explained the effect of parameter d on the running time and quality of seed nodes in the first and second paragraphs of section 4.2 of the paper.

Comment #5: It is unclear that the superiority of influence spread (e.g., Fig. 2) is resulting from considering community structure of network or the chosen community detection algorithm. To ameliorate this, the authors could compare the influence spread for different community detection method and also with and without considering community structure.

Answer: We thank the reviewer for this valuable suggestion. The superiority of influence spread is the result of using parameter d in the DCM propagation model, which causes our community-based algorithm to find seed nodes with better quality. This claim is demonstrated in Figures 3 and 4 of the revised paper. In Figure 3(a), we vary the value of parameter d on the x-axis and see its results in influence spread on the y-axis. To show that our algorithm and propagation model find groups of seed nodes which achieve influence spread near to that achieved by the standard greedy algorithm, we performed a new experiment to compare the influence spreads achieved by our approach and the greedy algorithm, and the results of this experiment are shown in Figure 4.

The reason that we used the approach of [32] for our community detection algorithm is its low running time in finding the communities, as we want to decrease the running time as much as possible. To make this clearer, we added the following sentence to the last paragraph of the **Community detection** subsection of Section 3.3:

> "The main reason of using the approach of [32] to find the communities is its less time consuming process to find the communities with a high quality. The experiments done in [32] confirm those specifications."

To consider the effect of communities on spread computations, we ran a new experiment in which we determined seed nodes in the NetHEPT dataset both with and without considering community structure. The results of this experiment are described in the following new paragraph added to the **The efficiency of proposed algorithm** subsection in Section 4.2:

> "Calculating the spread of nodes locally inside the communities they belong to, causes a huge decrease in running time. To proof that, we ran our CI2 algorithm to find a seed set of size 50 from NetHEPT dataset by 1) considering existing communities and calculating the spread values locally inside communities, 2) calculating the spread of each node in the whole graph without considering their own community. In the former case, CI2 finds the seed nodes in about 22 seconds, while it finds such seed nodes in about 70 minutes in the later case. This clearly shows the effect of localizing the spread calculations in running time, which is the result of considering community structure in CI2 algorithm."

Comment #6: Curiously, the authors can discuss about the effects of the temporal evolution of networks for their analysis.

Answer: We thank the reviewer for this intriguing suggestion. In order to investigate the effects of temporal evolution of networks, we would have to make many changes to the structure of and algorithms and experiments reported in our paper. Hence, we see this as future work, and have added an appropriate sentence to that effect at the end of the second paragraph of the Conclusions section as follows:

> "Analyzing the effect of temporal evolution of networks on influence maximization problem is another scope of studies which we should think about."

Comment #7: The writing of the manuscript should be improved. There are some typos around the manuscript, for example page 17, line 346, line 349 . . .

Answer: We thank the reviewer for pointing this out and regret their need to do so. We have thus worked very hard as a group to both improve the quality of the quality of the writing and eliminate all typos in the revised paper.

Comment #8: The conclusions section is just a recap of the proposed model. The authors should also stress there the benefits of the proposed model and the influence maximization algorithm.

Answer: On rereading the submitted paper, we agree with the reviewer that the main points of our paper were not recapped as clearly as they could have been. To address this, we have both modified the text of and added a new sentence stressing the main points of our paper to the first paragraph of the Conclusions section. The revised paragraph is as follows:

> "In this paper we studied competitive influence maximization from the follower's perspective and introduced an extended version of the LT model called DCM for influence propagation in a competitive fashion. To find the influential nodes in a social network graph, we proposed an efficient algorithm, which extracts the communities of the input graph and finds the most influential node in each community as a seed candidate. Then the final seed nodes are selected from the set including seed candidates. The size of the final seed set should be as small as possible, i.e. we assign the seed nodes to the second competitor so as to achieve higher influence spread comparing with the spread achievement of the first competitor's seed set by spending less budget. The ability of nodes to think about incoming influence in the DCM propagation model simulates a realistic situation in which nodes' tendency is toward the spread of influence, which has been adopted by the majority of their neighbors after d time steps. Adding parameter d to simulate the thinking ability of nodes results in finding influential nodes with higher quality and by calculating the spread values of each node locally inside its community, we achieved an acceptable running time. The results of our experiments on different real and synthetic datasets proof the efficiency of our proposed algorithm and propagation model."

Comment #9: A good discussion would be more appropriate: such as "under what circumstance, this model is proper for such social network or nature of a social network, might have advantage over other algorithms" etc.

Answer: On rereading the submitted paper, we agree with reviewer that detailed discussion of the circumstances under which our proposed algorithm performs well were lacking in the submitted paper. However, we believe that the changes we have made in our revision that are described in the comment responses above (particularly those responses describing new experiments that we have performed with respect to synthetic datasets) address the issues raised by the reviewer.

Reviewer #2

Comments #1–7: We thank the reviewer for giving a succinct overview of the contributions of our paper as well as important points within these contributions, as well suggestions for additional references; many of our contributions were indeed not highlighted as well as they should have been in the submitted paper. The suggested references are now cited as references [6] and [10] in the Introduction section. As for issues regarding contributions, we believe that these issues have now been addressed in the revisions described in our responses to the comments of Reviewer #1 above.

Comment #8: Test is not sufficient. Artificial datasets and Contrast test are also in need.

Answer: We agree with the reviewer that described testing was insufficient in the submitted paper, and believe that the various experiments done with respect to synthetic datasets described in the responses to our responses to the comments of Reviewer #1 above address this deficit.