Identifying Top-k Most Influential Nodes by using the Topological Diffusion Models in the Complex Networks

Maryam Paidar[1], Sarkhosh Seddighi Chaharborj[1, 2], Ali Harounabadi[3]

[1]Department of Computer Software, Islamic Azad University, Bushehr branch, Bushehr, Iran
[2]School of Mathematics and Statistics, Carleton University, Ottawa, Canada
[3]Department of Computer Software, Islamic Azad University Central Tehran Branch, Iran

Abstract—**Social networks are sub-set of complex networks, where users are defined as nodes, and the connections between users are edges. One of the important issues concerning social network analysis is identifying influential and penetrable nodes. Centrality is an important method among many others practiced for identification of influential nodes. Centrality criteria include degree centrality, betweenness centrality, closeness centrality, and Eigenvector centrality; all of which are used in identifying those influential nodes in weighted and weightless networks. TOPSIS is another basic and multi-criteria method which employs four criteria of centrality simultaneously to identify influential nodes; a fact that makes it more accurate than the above criteria. Another method used for identifying influential or top-k influential nodes in complex social networks is Heat Diffusion Kernel: As one of the Topological Diffusion Models; this model identifies nodes based on heat diffusion. In the present paper, to use the topological diffusion model, the social network graph is drawn up by the interactive and non-interactive activities; then, based on the diffusion, the dynamic equations of the graph are modeled. This was followed by using improved heat diffusion kernels to improve the accuracy of influential nodes identification. After several re-administrations of the topological diffusion models, those users who diffused more heat were chosen as the most influential nodes in the concerned social network. Finally, to evaluate the model, the current method was compared with Technique for Order Preferences by Similarity to Ideal Solution (TOPSIS).**

Keywords—Topological Diffusion; TOPSIS; Social Network; Complex Network; Interactive and Non-interactive Activities; Heat Diffusion Kernel

I.   INTRODUCTION

Most networks existing around us are of complex type. These include neural networks, social networks, organizational networks, computer networks, etc. [1]. Today, social networks have drawn attention more than the others. Every social network is composed of two elements of users and relationships: Users are defined as any entity participating in a relationship and are called Nodes; relationships are the connections between entities and are called Edges. Different types of relationships (work, family, friends, etc.) can exist between nodes [2].

Development of social networks accelerates the spread of different types of information, including rumors, news, ideas, advertisements, etc. People's decisions to refuse or accept information depends on the diffusion or spread method of the information [3]. Therefore, choosing the individuals intended for spreading the information gains much importance. So far, several models have been proposed in social networks for identification of those individuals who have social influence among people. Many existing social influence models for the definition of influence diffusion are based solely on the topological relationship of social networks nodes. The ideas of topological diffusion models can be used in the process of diffusion and spread of influence, and it can be evaluated through topological relationships among nodes in a social network [4], [5].

In this paper, it is attempted to examine the social influence regarding the heat diffusion kernel phenomenon, i.e., the dynamical equations are modeled based on heat diffusion. In fact, the heat diffusion process finds the influential nodes in complex networks using heat transfer laws based on interactive and non-interactive activities. Therefore, users who receive heat more than others and have a greater increase in temperature curve are identified as the most influential nodes in a social network. Then, through defining the modified heat diffusion kernels, the accuracy of the identification of the influential nodes can be increased.

II.   RELATED WORK

Doo & Liu in [3] proposed a model of social influence based on activities. Activity-based social influence is very effective in finding nodes in the social networks. In their paper, three types of topological diffusion, i.e., the Linear Threshold models (LT), Independent Cascade model (IC) and heat diffusion model have been fully expressed.

The authors in [6] carried out in 2015 for the first time the problem of influence maximization as a combination optimization problem. They considered two influences spread models, i.e. independent cascade and the linear threshold and evaluated these models extensively. But authors in [7] focused on the linear threshold model and presented a standard greedy algorithm in which the selection of a node with the maximum edge is increased repeatedly. The authors in [8] discussed mostly the time-critical influence maximization, where each node wants to reach maximize influence spread within a given deadline.

The authors in [9] presented three diffusion models together with three algorithms for selection of the best people. Their paper presented a new approach for analyzing social networks; subsequently, complexity analysis shows that the proposed model is also scalable to great social networks.

The authors in [10] designed a two-stage greedy algorithm (GAUP[1]) to find the most effective nodes in a network; GAUP initially computes the preferences of the user to a pattern that has latent feature model based on SVD[2] or a model based on vector space, then to find top-K nodes in the second stage, it utilizes a greedy algorithm.

A new technique [11] known as Technique for Order Preferences was presented by Similarity to Ideal Solution (TOPSIS weighted) to improve the ranking of node spread. With this method, the authors in [12] not only considered different centrality measures as the multi-attribute to the network but also proposed a new algorithm to calculate the weight of each attribute; and to evaluate its performance in four real networks they used the Susceptible– Infected–Recovered (SIR) model to do the simulation. Hu, *et al.*'s experiments on four real networks showed that the proposed method could rank the spreading ability of nodes more accurately than the original method.

III. METHOD

Social network models can be described by mathematical tools, such as graph and matrix. The most important property of graphs is their topological capability in which a vertex is created for each user, and if two users are friends with each other, the two vertices are connected. If a direction is attributed to each edge, the said network is called "directed", in which the order of vertices in an edge is important. Each row in the matrix corresponds to a vertex, and each column in the matrix also corresponds to a vertex. If there is a relationship/edge between two vertices of an edge, number 1 is used; otherwise, number 0 is inserted [13].

Users can send text, photo or video and like their friends, or write comments for them; hence, the users' activities can be divided into two categories: the interactive activities, i.e., those user activities that include nodes other than themselves, and the non-interactive activities, which include only himself and not someone else. As an example, permission for commenting on the profile picture of another person is an interactive activity because it is implemented between two nodes and the picture in the profile page is a non-interactive activity because it includes only the node itself. If i is taken as a row and j as a column of a matrix, assume that IA_{ij} represents some interactive activities from node v_i to the neighboring node v_j and NIA_i are some non-interactive activities at node v_i. Upon combination of interactive and non-interactive activities, there are several methods for spreading heat diffusion kernel. Interactive and non-interactive activities are defined as $\left(IA_{ji} / \sum_{k:(v_j,v_k)\in E} IA_{jk}\right)$ and $\left(1-\left(NA_i / MAX(NA)\right)\right)$, respectively. The *MAX(NA)* is defined as the largest number of non-

---

[1] Greedy Algorithm based on Users' Preferences (GAUP)

[2] Singular Value Decomposition (SVD)

interactive activity in V [3]. Fig. 1 is an example of the topological structure of Twitter social network displaying top 10 users. The positive integer numbers on the edges represent the number of interactive activities, and positive integer numbers with underline at each node represent the number of non-interactive activities implemented by that node.
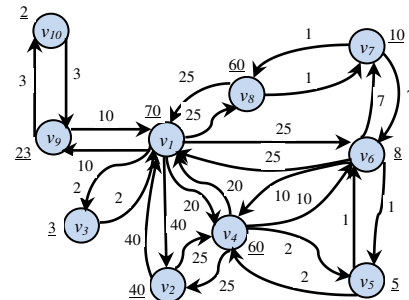

Fig. 1. A sample of social network (interactive and non-interactive activities).

Topological diffusion, as one of the processes of influence diffusion in social networks, exhibits the spread of influence through topological relationships between nodes. Heat diffusion model is one of such topological diffusion models. In this model, it is assumed that at the initial time t_0, in a social network of n nodes, all heat nodes except the node v_i has primary heat of zero. Node v_i which has some heat, is selected as a heat source and at time t_1, v_i diffuses the heat equally between all its neighbors. At time t_2, nodes with non-zero heat diffuser their heat to all their neighbors. With the repetition of this process for a period t for all n nodes, the influence of each node is found.

Suppose that $G = (V, E)$ represents a directional graph of social network, $V = \{v_1, v_2, ..., v_n\}$ is a set of vertices representing the number of users, and $E = \{(u,v)|u,v \in V\}$ is a set of edges representing the friendship relationship between users [14]. The heat at vertex v_i at time t is defined by the function $H_i(t)$; heat flows from a high-temperature node to a low-temperature node following the edges between vertices. In the directional graph, at time interval Δt, vertex v_i diffuses the heat to the amount of $DH_i(\Delta t)$ through output edges to next nodes. At the same time, the vertex v_j, receives the heat $RH_i(\Delta t)$ through input edges. Heat variations at vertex $V \in v_i$ between the time interval t and $t+\Delta t$ is defined with the sum of the differences between the heat that received, and the heat diffuses to, all its neighbors.

$$H_i(t + \Delta t) - H_i(t) = RH_i(\Delta t) - DH_i(\Delta t). \tag{1}$$

$DH_i(\Delta t)$ and $RH_i(\Delta t)$ are defined as (2) and (3):

$$RH_i(\Delta t) = \alpha \Delta t \sum_{j:(v_j,v_i)\in E} H_j(t) \left(\frac{IA_{ji}}{\sum_{k:(v_j,v_k)\in E} IA_{jk}} \right). \tag{2}$$

$$DH_i(\Delta t) = \alpha \Delta t \, H_i(t) \left(1 - \beta \frac{NA_i}{MAX(NA)} \right). \tag{3}$$

Parameter α, called heat diffusion coefficient, controls the rate of heat transfer. Value of α is a real number between 0 and 1. If α tends to 0, heat transfer would be difficult, and the heat will not spread everywhere. If α tends to 1, without heat loss, all heat will be distributed among all neighbors, i.e., if α is big, the heat is diffused rapidly from one node to another. In this paper, the value of α is taken as 0.5, which means that the heat is transferred with a little loss between nodes. β in (3) is used as a weight for non-interacting activities and takes a real number between 0 and 1. If β is considered 0, it means that some of the non-interacting activities are ignored and have no effect on the heat diffusion process. But if β is set to 1, in heat diffusion process, node v_i loses its heat with a lower rate; in this paper, in all curves, the value of β is taken to be 0.5. Equation (2) and (3) can be combined with equation of heat variation (1) and obtain (4).

$$H_i(t+\Delta t)-H_i(t)=\alpha\Delta t\left(\sum_{j:(v_j,v_i)\in E}\left(H_j(t)\left(\frac{IA_{ji}}{\sum_{k:(v_j,v_i)\in E}IA_{jk}}\right)-H_i(t)\left(1-\beta\frac{NA_i}{MAX(NA)}\right)\right)\right). \quad (4)$$

In general, for n nodes, it can be written as (5):

$$H(t+\Delta t)-H(t)=\alpha\Delta t KH(t). \quad (5)$$

In (5), K is an $n\times n$ matrix representing the heat diffusion kernel from graph G, and $H(t)$ is a column vector representing heat distribution at time t in graph G, which is defined for primary heat source $H(0)$. Now the limit of (5) is taken on Δt when Δt approaches zero and gives (6).

$$\lim_{\Delta t\to 0}\frac{H(t+\Delta t)-H(t)}{\Delta t}=\lim_{\Delta t\to 0}\alpha KH(t),$$

$$\to\frac{dH(t)}{dt}=\alpha KH(t), \quad (6)$$

Then we take integral of both sides of (6):

$$\int\frac{dH(t)}{dt}=\int\alpha K dt. \quad (7)$$

Now if we take the *ln* of both sides of (7), the heat equation is defined as (8):

$$\to H(t)=e^{\alpha Kt}H(0). \quad (8)$$

Then, using heat change equations *RH* and *DH*, two types of heat diffusion kernel are provided. Now, through MATLAB software, curves are drawn and with the help of curves, the most influential node can easily be found.

The first defined kernel is called heat diffusion kernel based on interactive and non-interactive activities. Matrix K is defined as:

$$K(i,j)=\begin{cases} IA_{ji}/\sum_{k:(v_j,v_i)\in E}IA_{jk} & ; \ (v_j,v_i)\in E \\ -\left(1-\beta\left(NA_i/MAX(NA)\right)\right) & ; \ i=j \text{ and } d_i>0 \\ 0 & ; \ \text{otherwise.} \end{cases} \quad (9)$$

Fig. 2(a) illustrates the topological structure of heat diffusion based on interactive and non-interactive activities. This graph has been calculated and plotted using (9) and Fig. 1. As shown in the figure node v_1 neighbors six nodes v_2, v_3, v_4, v_6, v_8 and v_9. Two nodes v_4 and v_6 have four neighboring nodes; nodes v_1, v_2, v_5 and v_6 neighbor node v_4 and nodes v_1, v_4, v_5 and v_7 are four neighbors of node v_6. Each of nodes v_2, v_5, v_7, v_8 and v_9 is directly related to two nodes. Finally the nodes v_3 and v_{10} have one neighbor, it means that the node v_1 is the only neighbor of node v_3 and node v_9 is the only neighbor of node v_{10}.

Fig. 2(b) is the diagram of heat diffusion in which, as an example, node v_7 has been chosen as the heat source and is plotted by using (8) the diagram of heat diffusion. X-axis is the time line and Y-axis is the amount of heat at each node. The heat diffusion in this diagram is such that initially (at time zero), the heat source has a lot of heat and the rest of the nodes are without heat. At time 1, v_7 diffused heat to its neighbors (v_6 and v_8), and these two nodes gain heat. But node v_6 receives more heat and goes up. In this manner, the heat source reduces with time due to diffusion of heat to other nodes and the other nodes also receive heat from their neighbors and increase. Node v_1 receives more heat because it is directly linked with six nodes and increase more than all nodes; so it is obvious that it is ascending in the diagram. But nodes v_3 and v_{10} with just one neighbor will be the lowest limit of the diagram. A point worthy to note in this diagram is that in the time interval 10, two nodes v_2 and v_4 are higher than node v_6; the reason for that is that these two nodes receive more heat than node v_6; hence in the diagram, nodes v_2 and v_4 have higher increases.

The degree of each node is one of the influential factors for information diffusion. If d_j represents the degree of each node, the weight on the edges connected to node v_j is calculated as (10):

$$Weight_{v_j}=\frac{1}{d_j}. \quad (10)$$
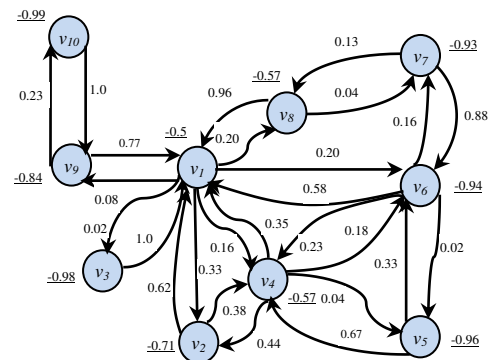
(a)   Topological structure
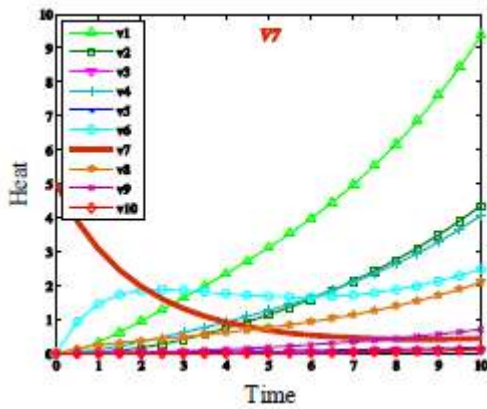
(b) Diagram of heat diffusion with heat source v_7

Fig. 2.   Heat diffusion based on interactive and non-interactive activities.

Therefore, $RH_i(\Delta t)$ can be changed into (11):

$$RH_i(\Delta t) = \alpha \Delta t \sum_{j:(v_j, v_i) \in V} \frac{H_j(t)}{d_j}. \quad (11)$$

Now a new kernel called heat diffusion kernel can define through (11) and (3) based on the non-interactive activities.

$$RH_i(\Delta t) = \alpha \Delta t \sum_{j:(v_j, v_i) \in V} \frac{H_j(t)}{d_j}. \quad (11)$$

$$DH_i(\Delta t) = \alpha \Delta t \, H_i(t) \left(1 - \beta \frac{NA_i}{MAX(NA)}\right). \quad (3)$$

Matrix K changes as (12):

$$K(i,j) = \begin{cases} 1/d_j & ; \quad (v_j, v_i) \in E \\ -\left(1 - \beta\left(NA_i / MAX(NA)\right)\right) & ; \quad i = j \text{ and } d_i > 0 \\ 0 & ; \quad \text{otherwise.} \end{cases} \quad (12)$$

Fig. 3(a) is drawn using (12) and Fig. 1. As it was mentioned above, d_j represents the degree of each node meaning that if the degree of each node is calculated based on (10), weight of the edge between nodes is obtained. As an example, vertices v_6 and v_8 are two neighbors of node v_7; therefore, the weight of 0.5 is assigned to $E(v_7, v_6)$ and $E(v_7, v_8)$, respectively, which means half of the heat of v_7 is transferred to v_6 and the other half to v_8. Node v_{10} has just one friend, which is vertex v_9. Thus, the weight of $E(v_{10}, v_9)$ is equal to 1. It means that the whole heat of v_{10} is diffused to v_9. Similarly, the edge weight between vertices is calculated.

For example, node v_7 is considered as a heat source and the heat diffusion diagram is plotted in Fig. 3(b) using (8). The heat diffusion process in this diagram is as follows: at time interval zero, the heat source has high heat, and the rest of the nodes are without heat.
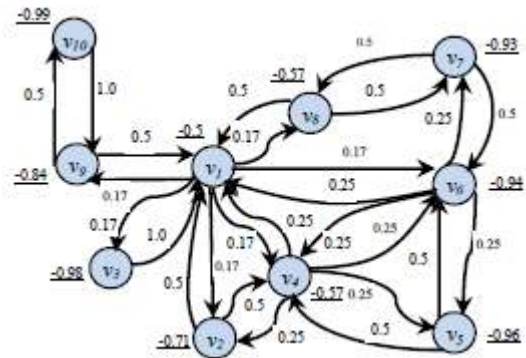
(a)   Topological structure
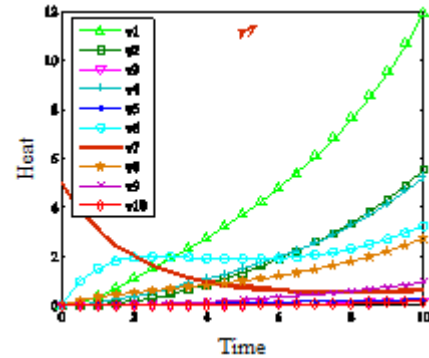
(b)   Diagram of heat diffusion with heat source v7

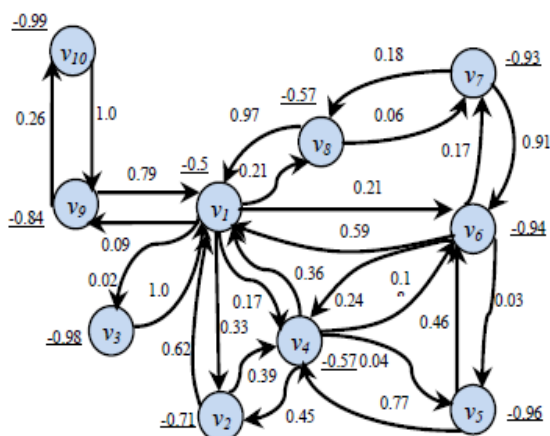Fig. 3.   Heat diffusion based on non-interactive activities.

With the passage of time, heat of the source reduces due to heat diffusion to other nodes; accordingly, heat increases in two nodes v_6 and v_8, which are directly related to heat source (v_7). But as shown, finally, it is node v_6, with more links to other nodes, which finally receives more heat and goes higher than node v_8. From time interval 4 onwards, node v_1 increases and surpasses all other nodes, even the heat source. Node v_1, with links to more nodes, receives higher heat. Therefore, expectedly, it follows an ascending trend.

A comparison of Fig. 2(b) and Fig. 3(b) diagrams indicates that in both diagrams, node v_1 is higher than all others; therefore, it is selected as the most influential node. But, as previously mentioned, the y-axis represents the receipt of the amount of heat; where each node receiving more heat goes higher.

Now, a comparison of node v_1 in both diagrams reveals that node v_1 in diagram Fig. 2(b) is higher than in diagram Fig. 3(b); therefore, it can be stated that diagram Fig. 2(b) has received more heat. It is thereupon implied that the heat diffusion kernel based on interactive and non-interactive activities (9) acts better than the heat diffusion kernel based on non-interactive activity (12); because the nodes in heat diffusion kernel based on interactive and non-interactive activity have received more heat. It is now attempted to obtain a modified kernel by combination of kernels in such a way that the most influential node receives more heat (i.e., it is higher in the y-axis). This core is expressed as (13):

$$K(i,j) = \begin{cases} (1+IA_{ji}) / \left(\sum\limits_{k: (v_j, v_k) \in E} IA_{jk} + \dfrac{1}{d_j} \right); & (v_j, v_i) \in E \\ -\left(1 - \beta\left(NA_i / MAX(NA)\right)\right) & ; \quad i = j \text{ and } d_i > 0 \\ 0 & ; \quad \text{otherwise.} \end{cases} \quad (13)$$

Fig. 4(a), is the topological structure of the proposed kernel drawn using Fig. 1 and (13). As can be seen, the weight on the edges varies when (13) is applied. Heat diffusion diagram with heat source v_7 using (8) is indicated in Fig. 4(b). In this diagram also the heat source has a descending trend over time due to heat diffusion to other nodes; and still, node v_1 with more links has higher heat and increases more than other nodes. Accordingly, node v_1 is selected as the most influential node in the proposed kernel similar to the previously presented kernels.

(a)   The topological structure of proposed kernel
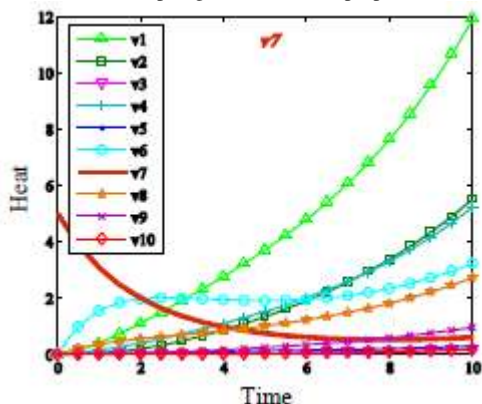
(b) Diagram of heat diffusion with heat source v7

Fig. 4.   Heat diffusion of proposed kernel.

Heat diffusion kernel based on interactive and non-interactive activities is compared with the proposed kernel in two diagrams Fig. 2(b) and 4(b) in Fig. 5. In this diagram, too, node v_7 is selected as a heat source; continuous line represents the heat source of proposed method, and a dotted line represents the heat source of the method of interactive and non-interactive activities. Node v_1 (as the most influential node) for the proposed method is indicated as triangular and for the case

of interactive and non-interactive method is shown as a circle. As seen in the diagram of Fig. 5, the triangular curve is higher than the circle curve, i.e. node v_1 in the proposed method receives more heat, and therefore triangular curve has higher increase. So the proposed kernel acts better than the heat diffusion kernel based on the interactive and non-interactive activities.
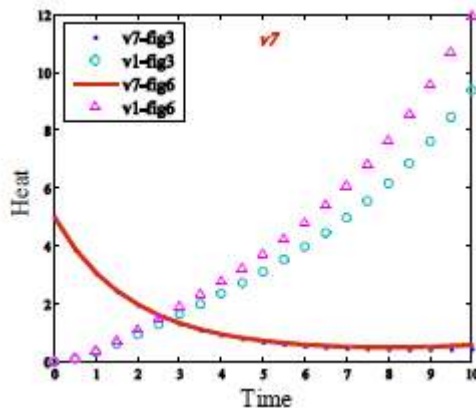

Fig. 5.   Comparing best kernel with proposed kernel.

IV.   EVALUATION

In this section, two real data sets, Revije.paj and EIES.paj, are selected; then using Pajek software, the data is analyzed and implemented using MATLAB software. After that, TOPSIS method, which is a multiple criterion decision-making method, and is defined based on positive ideal solution and negative ideal solution, is used for evaluation. TOPSIS model evaluation and prioritization procedure will be as follows [4]:

The first step is development of a decision matrix; that is, using m criteria, n indexes will be evaluated.

The second step is normalization of the decision matrix. To that end, (14) is used, which is a vector method.

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum\limits_{j=1}^{m} x_{ij}^2}} \quad ; \quad i = 1,\ldots,m \quad ; \quad j = 1,\ldots,n. \quad (14)$$

The third step is making weighted normalized decision matrix where the weight of criteria is calculated by (15).

$$v_{ij} = w_j \times r_{ij} \quad ; \quad i = 1,\ldots,m \quad ; \quad j = 1,\ldots,n. \quad (15)$$

In the fourth step, the positive and negative ideals are calculated. The highest performance of each index (positive ideal) and the lowest performance of each index (negative ideal) are represented by A^+ and A^-, respectively. These two indexes are calculated in (16) and (17).

$$A^+ = \left\{ \left(\max_i \ v_{ij} \mid j \in K_b \right) \left(\min_i \ v_{ij} \mid j \in K_c \right) \right\}. \quad (16)$$

$$A^- = \left\{ \left(\min_i \ v_{ij} \mid j \in K_b \right) \left(\max_i \ v_{ij} \mid j \in K_c \right) \right\}. \quad (17)$$

In the fifth step, Euclidean distances of each criterion from the positive ideal and negative ideal are calculated by (18) and (19), respectively.

$$S_i^+ = \sqrt{\sum_{j=1}^{n}\left(v_j^+ - v_{ij}\right)^2} \quad ; \quad i = 1,\dots,m \quad ; \quad j = 1,\dots,n. \quad (18)$$

$$S_i^- = \sqrt{\sum_{j=1}^{n}\left(v_j^- - v_{ij}\right)^2} \quad ; \quad i = 1,\dots,m \quad ; \quad j = 1,\dots,n. \quad (19)$$

The sixth step is the calculation of ideal solution, i.e., the relative closeness of each criterion to the ideal solution, which is obtained by (20).

$$C_i = \frac{S_i^-}{S_i^- + S_i^+} \quad ; \quad i = 1,\dots,m. \quad (20)$$

Prioritization is based on the value of C_i^+ where this value can be $0 \le C_i^+ \le 1$. When this value is closer to one, it indicates the highest rank, and when this value is close to zero it indicates the lowest rank.

A. Revije Data Set

Revije is a data set of Slovenian magazines and journals published in 1999 and 2000[3] [15]. 124 different magazines and journals were listed, and over 100,000 people were asked to read magazines and journals. A typical network is created from this data set in which the magazines are the vertices. In this network, edges are directed and weighted; the ring at vertices corresponds to the readers of the magazine. If a reader has read two or more magazines, then those magazines are linked together, and the weight on that indicates the number of readers of two magazines. The topological structure of this data set is shown in Fig. 6.
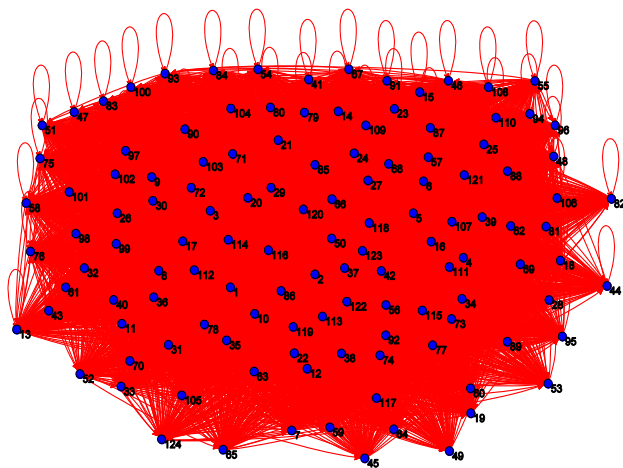

Fig. 6.   The topological structure of Revije data set.

The data set of magazines and journals have interactive and non-interactive activities and include 124 vertices and 12068 edges. The highest degree of nodes is 123, which is related to

nodes v_1, v_2, v_4, v_6, v_{27}, v_{42}, v_{113} and v_{119} and the lowest degree is 18, which belongs to node v_{50}. Ten highest nodes prioritized by TOPSIS model are shown in Table 1.

Node v_4 which has the highest priority in TOPSIS model is considered as the heat source. Top 10 nodes with the heat source v_4 for Revije data set are shown in Table 2, in the order of priority.

TABLE. I.   THE ARRANGEMENT OF 10 SUPERIOR NODES BY THE TOPSIS MODEL FOR REVIJE DATA SET

| Priority / Model	1	2	3	4	5	6	7	8	9	10
Model TOPSIS	v_4	v_1	v_{119}	v_{116}	v_3	v_{27}	v_{122}	v_2	v_{120}	v_{39}

TABLE. II.   THE ARRANGEMENT OF 10 SUPERIOR NODES WITH THE HEAT SOURCE V_4 FOR REVIJE DATA SET

| Priority / Method	1	2	3	4	5	6	7	8	9	10
Heat diffusion kernel based on interactive and non-interactive activity	v_4	v_{119}	v_1	v_{116}	v_{27}	v_{122}	v_{120}	v_{39}	v_2	v_3
Heat diffusion kernel based on the non-interactive activity	v_4	v_{119}	v_1	v_3	v_{116}	v_{39}	v_2	v_{120}	v_{27}	v_{122}
Proposed heat diffusion kernel	v_4	v_{119}	v_1	v_{116}	v_{27}	v_{122}	v_{120}	v_{39}	v_2	v_3

By comparing Tables 1 and 2, it can be seen that in all methods, node v_4 is the priority, but other priorities are different. For example, the second priority in TOPSIS method is node v_1 and in the heat diffusion method, node v_{119}. Hence, the heat diagrams of Revije dataset for two nodes v_1 and v_{119} are drawn in Fig. 7 and from that, the most influential node will be determined. For readability of the diagrams, the heat source (node v_4) is drawn only for the proposed heat diffusion kernel; heat diffusion kernel based on interactive and non-interactive activities are shown with φ_1, heat diffusion kernel based on non-interactive activity is shown by φ_2, and the proposed heat diffusion kernel is shown by φ_3.

As shown in Fig. 7, nodes v_1 and v_{119} are compared for the above-mentioned heat diffusion models, in both diagrams (a) and (b). Fig. 7(b) which is related to the node v_{119}, receives more heat compared with diagram Fig. 7(a) and thus has a higher ascending trend. So node v_{119} is more influential than node v_1 and has a higher priority. Hence, for advertisement and news diffusion, firstly node v_{119} and then node v_1 are selected.
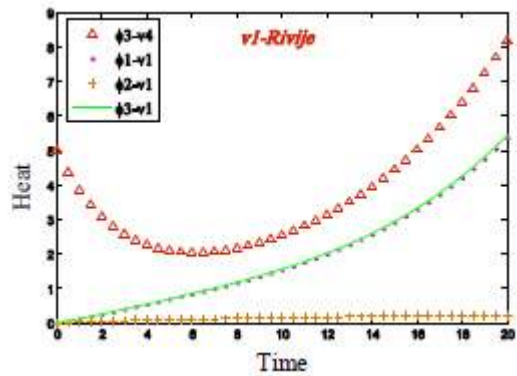
(a) Comparing node v_1 for all three diffusion models
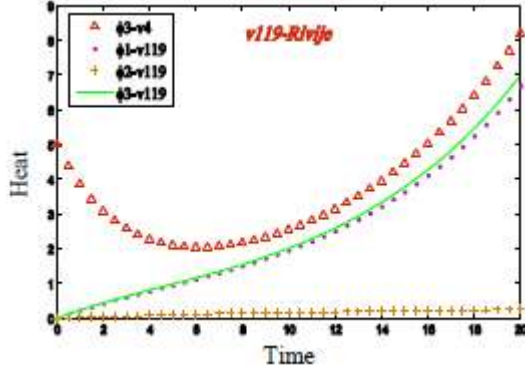
(b) Comparing node v_{119} for all three diffusion models

Fig. 7.   Comparing the second priority for models presented in Revije.

B. EIES Data Set

EIES data set[4] [15] of Wasserman and Faust data collection is the second set of real data that has been considered in this paper. This data set also has interactive and non-interactive activities; its communication network is directional, weighted and has 32 nodes and 460 edges. Node v_1 with the degree 29 and node v_{25} with the degree 6 are highest and lowest degree in this data set, respectively. Fig.8 shows the topological structure of the data set.
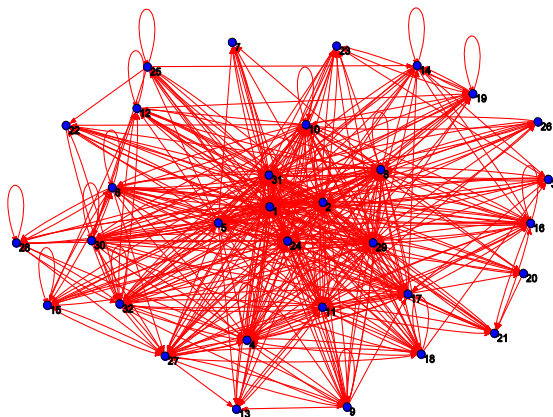

Fig. 8.   The topological structure of EIES data set.

---

[4] http://vlado.fmf.uni-lj.si/pub/networks/data.

TABLE. III.   THE ARRANGEMENT OF 10 SUPERIOR NODES BY THE TOPSIS MODEL FOR EIES DATA SET

Priority / Model	1	2	3	4	5	6	7	8	9	10
Model TOPSIS	v_1	v_{29}	v_8	v_2	v_{32}	v_{31}	v_{11}	v_{24}	v_{10}	v_4

TABLE. IV.   THE ARRANGEMENT OF 10 SUPERIOR NODES WITH THE HEAT SOURCE V_4 FOR EIES DATA SET

| Priority / Method	1	2	3	4	5	6	7	8	9	10
Heat diffusion kernel based on interactive and non-interactive activity	v_1	v_{32}	v_{29}	v_8	v_2	v_{31}	v_{11}	v_{24}	v_{27}	v_{10}
Heat diffusion kernel based on the non-interactive activity	v_1	v_{32}	v_{24}	v_{31}	v_2	v_{29}	v_5	v_8	v_{11}	v_{10}
Proposed heat diffusion kernel	v_1	v_{29}	v_{32}	v_8	v_2	v_{31}	v_{11}	v_{24}	v_{27}	v_{10}

In Table 3, the top 10 nodes based on the TOPSIS model are shown in the order of priority.

According to Table 3, node v1 in TOPSIS model has the highest priority; accordingly, in Table 4, the order of priorities of 10 superior nodes for data set EIES is indicated using v_1 as the heat source.

Comparing Tables 3 and 4, it can be seen that the priority in all methods belongs to node v_1, but other priorities vary. For example, the second priority in some methods is v_{29} while in some other ones, it is v_{32}. Thus, the diagram of heat diffusion of data set EIES for two nodes v_{29} and v_{32} is drawn in Fig. 9. Then, the most influential node is found from the diagrams. For readability of diagrams, the heat source (node v_1) is drawn just for the proposed heat diffusion kernel; heat diffusion kernel based on interactive and non-interactive activities are shown by φ_1, the heat diffusion kernel based on non-interactive activity is shown by φ_2 and proposed heat diffusion kernel is shown by φ_3.

By comparing the two diagrams (a) and (b) in Fig. 9, it is clear that the two diagrams have little difference. Now, if this is compared to diagram (c), which is related to node v_8 (third priority of TOPSIS model); for data set EIES, it can be stated that node v_8 has lower priority compared with node v_{32}. Therefore, nodes v_{29}, v_1 and v_{32} have the best condition for diffusion in social networks and they are more appropriate for spreading ideas and news or advertisements.
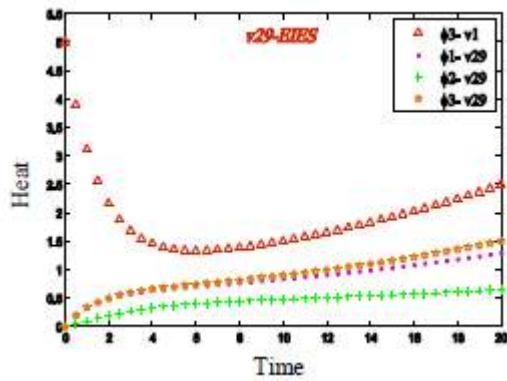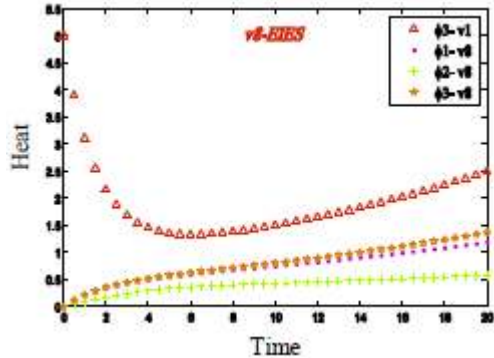
(a) Comparing node v_{29} for all three diffusion models
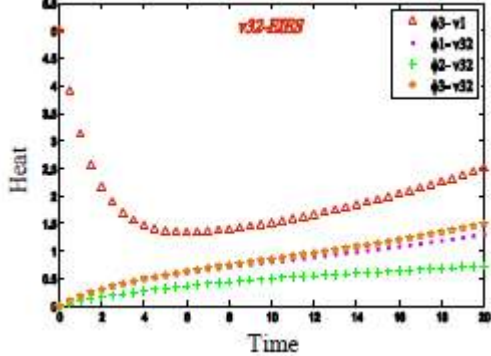
(b) Comparing node v_{32} for all three diffusion models

(c) Comparing node v_8 for all three diffusion models

Fig. 9. Comparing top nodes with heat source v_1 for models presented in EIES.

V. CONCLUSIONS

The most obvious problem in the field of social networks is finding k influential nodes in a network of individuals so as to benefit from the influence of these individuals in the entire network and diffuse the news in entire network to the most neighbors in the best and fastest possible way. Earlier, in the Methods section, three different heat diffusion kernels were defined: heat diffusion kernel based on interactive and non-interactive activities, heat diffusion kernel based on non-interacting activity and proposed heat diffusion kernel; this was followed in Experiments and Evaluation section by finding influential nodes for two data sets Revije and EIES (more specifically, node v_4 for Revije data set and node v_1 for EIES data set). Now in this section, the best heat diffusion kernel is depicted using Fig. 10. As stated previously, for readability of diagrams, the heat diffusion kernel based on interactive and

non-interactive activities are shown by φ_1, the heat diffusion kernel based on non-interactive activity is shown by φ_2 and proposed heat diffusion kernel is shown by φ_3.

In Fig. 10(a), the diagram with the heat source for node v_4 in Revije data set is shown, where the diffusion kernel related to φ_3, which is the proposed diffusion kernel, is higher; in other words, since the proposed kernel has higher heat compared to the other two kernels, it has an ascending trend. In this diagram, heat diffusion kernel based on the non-interacting activity is lower than the other kernels.

In Fig. 10(b), the diagram with the heat source for node v_1 in data set EIES is drawn in which again the proposed diffusion kernel (φ_3) has increased more compared to the other two kernels, i.e., in EIES data set, the proposed kernel has higher heat. On the other hand, in this diagram also the heat diffusion kernel based on non-interactive activity is lower than the other kernels.
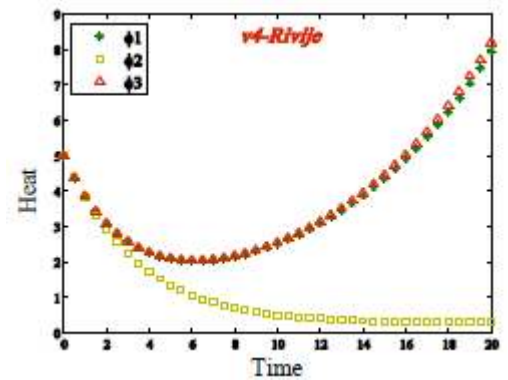
(a) The heat source of node v_4 for Revije data set
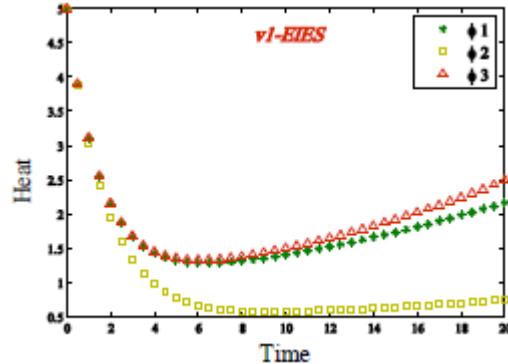
(b) The heat source of node v_1 for EIES data set

Fig. 10. Comparison of kernels to determine the highest priority of the kernel.

Therefore, proposed heat diffusion kernel in both data set Revije and EIES have higher increase and more heat. Thus, the order of kernels from the lowest to the highest priority can be concluded as the following:

1) Heat diffusion kernel based on the non-interactive activity.
2) Heat diffusion kernel based on interactive and non-interactive activity.
3) Proposed heat diffusion kernel.

REFERENCES

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, "Complex networks: Structure and dynamics," *Physics Reports*, vol. 424, no. 4, pp. 175-308, Feb. 2006.

[2] M. E. Newman, "The structure and function of complex networks", SIAM review, vo1. 45, no. 2, pp. 167-256. و 2003.

[3] M. Doo, and L. Liu, "Extracting top-k most influential nodes by activity analysis," *In Information Reuse and Integration (IRI), 2014 IEEE 15th International Conference on*, 2014, pp. 227-236.

[4] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. Kramer, C. Marlow, J. E. Settle, and J. H.Fowler, "A 61-million-person experiment in social influence and political mobilization", *Nature*, vol. 489, no. 7415, pp. 295-298, September 2012.

[5] I. Anger, and C. Kittl, "Measuring influence on Twitter", In Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies, ACM, no. 31, September 2011.

[6] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," *Theory of Computing*, vol. 11, no. 4, pp. 105-147, April 2015.

[7] Z. Lu, L. Fan, W. Wu, B. Thuraisingham, and K. Yang, "Efficient influence spread estimation for influence maximization under the linear threshold model," *Computational Social Networks*, vol. 1, no. 1, pp. 2, Dec 2014.

[8] W. Chen, W. Lu, and N. Zhang, "Time-critical influence maximization in social networks with time-delayed diffusion process," *In AAAI*, vol. 2012, pp. 1-5, July 2012.

[9] H. Ma, H. Yang, M. R. Lyu, and I. King, "Mining social networks using heat diffusion processes for marketing candidates selection," *In Proceedings of the 17th ACM conference on Information and knowledge management*, 2008, pp. 233-242.

[10] J. Zhou, Y. Zhang, and J. Cheng, "Preference-based mining of top-K influential nodes in social networks," *Future Generation Computer Systems*, vol. 31, pp. 40-47, Feb 2014.

[11] 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, April 2014.

[12] J. Hu, Y. Du, H. Mo, D. Wei, and Y. Deng, "A modified weighted TOPSIS to identify influential nodes in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 444, pp. 73-85, Feb 2016.

[13] J. Scott, "Social network analysis", *Sage*, 2012.

[14] P. Van Mieghem, "Graph spectra for complex networks", Cambridge University Press. 2010, pp. 13-17.

[15] V. Batagelj and A. Mrvar (2006). Pajek datasets.