A Hurst-based Diffusion Model using Time Series Characteristics for Influence Maximization in Social Networks

Bhawna Saxena, Vikas Saxena *Senior Member, IEEE*, Nishit Anand, Vikas Hassija, Vinay Chamola *Senior Member, IEEE*, and Amir Hussain *Senior Member, IEEE*

***Abstract*—Online Social Networks have grown exponentially in the recent years whilst finding applications in real life like marketing, recommendation systems, and social awareness campaigns. An important research area in this field is Influence Maximization, which pertains to finding methods for maximizing the spread of information (influence) across an OSN. Existing works in IM widely use a pre-defined edge propagation probability for node activation. Hurst exponent (H), which depicts the self-similarity in the time series depicting a user's past interaction behaviour, has also been used as activation criteria. In this work, we propose a Time Series Characteristic based Hurst-based Diffusion Model (TSC-HDM), which calculates H based on the stationary or non-stationary characteristic of the time series. TSC-HDM selects a handful of seed nodes and activates a seed node's inactive successor only if H>0.5. The proposed model has been tested on 4 real-world OSN datasets. The results have been compared against 4 other IM models - Independent Cascade, Weighted Cascade, Trivalency, and Hurst-based Influence Maximisation. TSC-HDM is found to have achieved as much as 590% higher expected influence spread as compared to the other models. Moreover, TSC-HDM has attained 344% better average influence spread than other state-of-the-art models namely LIR, A-Greedy, LPIMA, Genetic Algorithm with Dynamic Probabilities, NeighborsRemove, DegreeDecrease, IGIM, IRR, and PHG.**

***Index Terms*—online Social Networks, Influence Maximization (IM), Hurst-based Diffusion Model, Self-Similarity.**

I. Introduction

In terms of size and reach, Online Social Networks (OSNs) have grown exponentially during the last few years. OSNs enable umpteen users to connect and share information. OSNs have become an integral part of our lives and its immense usage has led to the generation of unprecedented volumes of user-related data [1]. The availability of this data has presented researchers with newer opportunities, pertaining to the research on user behaviour in social networks [2]. The widespread popularity of OSNs is owing to their applicability in numerous areas like finance, bio-informatics, politics, healthcare, social awareness, etc. Using OSNs for diffusing (spreading) information is a part of many real-life activities such as viral marketing, online recommendation systems, online advertising, influential blogger identification, healthcare communities, and political and social awareness campaigns [3–9].

Information gets spread or diffused through interactions between individuals in the society [10], [11]. A real-world social group can be conceptualized as a huge social network (graph) wherein "a node is an abstract representation of an individual user in the real world [12]." Interaction between two users indicates a relation between them and is depicted by an edge connecting the two corresponding nodes [13, 14].

Social Interactions Analysis is all about analysing social interactions for a better comprehension of social structures and user relations [15]. *Social Influence Analysis* is a favoured research domain under social interactions analysis, "which focuses on studying the process of information diffusion in social networks along with the identification of influential users [5]." *Influence Maximization* (IM), a prominent research topic within the domain of social influence analysis, aims at finding an answer to the question, "how to maximize the spread of influence across an OSN [16]."

A predominant commercial application of IM is viral (influencer) marketing, wherein an organization collaborates with influencers to promote their product or services [17]. Influencer marketing has seen a continuous growth over the past few years, with 93% of businesses confirming the use of influencer marketing as a major marketing strategy [18]. In the U.K. Google searches for influencer marketing grew by 400% during 2016-2021 [18]. Statistics reveal that in 2022, the U.S. influencer marketing industry was around $16.4 billion and in 2023 around 89% of marketers currently utilising influencer marketing plan to raise/ maintain their influencer marketing budget in 2023 [19]. Recommendations from friends/ family are often a one-to-one interaction, however influencer marketing has the potential of reaching thousands, or even millions of people. It has been found that 49% of consumers rely on influencer recommendations and about 82% consumers take purchasing advice from social networks

Bhawna Saxena, Vikas Saxena, and Nishit Anand are with the Department of Computer Science and Engineering & IT, Jaypee Institute of Information Technology, Noida, India 201304 (e-mail: bhawna.saxena@jiit.ac.in, vikas.saxena@jiit.ac.in, nishitanand01@gmail.com, and vikas.hassija@jiit.ac.in).

Vikas Hassija is with the School of Computer Engineering, Kalinga Institute of Industrial Technology, India (e-mail: vikas.hassijafcs@kiit.ac.in).

Vinay Chamola is with the Department of Electrical and Electronics Engineering, BITS-Pilani, Pilani Campus, India 333031 (e-mail: vinay.chamola@pilani.bits-pilani.ac.in)

Amir Hussain is with the School of Computing, Edinburgh Napier University, Scotland, UK (e-mail: A.Hussain@napier.ac.uk).

Digital Object Identifier: XXXXXXXXXXXX

[20]. These statistics depict that influencer marketing is an important aspect of the marketing strategy and contributes significantly towards better marketing outcome.

Consider a scenario such that an organization plans to utilize an OSN for viral marketing of its new offering by utilizing an OSN. The organization would thus aim to maximize the spread (reach) of its campaign and touch as many potential customers as possible. Achievement of this goal is reliant on two core steps. First step is seed node identification [21]. Seed nodes are a small subset of OSN users who are the initial adopters of information and initiate the diffusion process. The second task involves the development of a diffusion model pertaining to the underlying information diffusion process. The diffusion model signifies how information would spread from a node to its neighbours, over the edge connecting the two of them. Studies indicate that "how information is propagating from one user to another heaviliy impacts the influence spread achieved [22]". Hence, selection of diffusion model is very critical as the diffusion model outlines the condition(s) for node activation.

The diffusion model helps in assessing the influence spread anticipated to be achieved by the selected seed nodes. Some existing diffusion models for IM are *Independent Cascade* (IC), *Weighted Cascade* (WC), *Linear Threshold* (LT), *Trivalency* (TV) model, and *Dynamic Independent Cascade* (DIC) model [23–25]. In all these aforesaid diffusion models, diffusion is reliant on some pre-decided value for propagation probability and is either selected randomly or from a pre-defined set of values. In the *Credit Distribution with Node Features* model developed by Deng et al., credits allocated to a node are the deciding factor for its activation [26]. The credits are decided based on the past interactions carried out by the node. Under the Voter model of diffusion, a node chooses a successor, based on the probability derived from the assigned edge weights, and aligns its own opinion with the opinion of the chosen successor [27–30].

Edges surely depict the probable paths for diffusion, but how much diffusion has been carried out over a path, can only be known by studying the connecting node's past temporal behaviour. Thus, "to develop a more realistic diffusion model, node's actual past interactions should also be considered for node activation [31]". Propagation probability signifies likelihood of diffusion, but node's past interaction pattern presents a more truthful view of diffusion that has actually taken place [32].

User behaviour (interactions) being a time dependent phenomenon, can be represented in the form of a time series and this time series can further be analyzed to discover the existence of a pattern/ trend. Although, human are assumed to behave randomly in general, researchers have found that human behaviour tends to repeat over time, thereby displaying statistical similarity. Studies reveal that time series generated analogous to how humans behave in the real-world, has been found to exhibit self-similarity [33]. On similar lines, the behaviour of OSN users over time can also be expected to exhibit statistical self-similarity. Although "the role of self-similarity in edge creation in OSNs" has been explored, exhibition of self-similarity in an OSN users' past interaction pattern is not a much explored aspect [34].

Saxena and Saxena have presented the *Hurst-based Influence Maximization* (HBIM) diffusion model, in which firstly a time series has been generated corresponding to each node's past interactions [35]. Thereafter, *Hurst* exponent (H) is computed for quantifying the self-similarity trend (SST) displayed by each node's generated time series. Under HBIM model node activation is reliant on its degree along with the H value quantifying its past activity's SST.

A time series can be either stationary or non-stationary. Researchers have observed that applying the same method for computing the H value of a stationary as well as non-stationary time series may lead to inaccurate evaluations [36, 37]. Drawing motivation from this notion, a novel diffusion model, called *Time Series Characteristic based Hurst-based Diffusion Model* (TSC-HDM) is being presented in this paper, which augments the aforementioned HBIM model and incorporates the usage of different methods to compute H value, depending upon if the time series is stationary or not.

Under the proposed TSC-HDM model, firstly a time series corresponding to the past interactions of each node is generated, followed by the computation of H value for quantifying the SST displayed by the time series. Based upon the characteristic of the time series under consideration, i.e., whether it is stationary or not, different methods have been used to compute H value. Thereafter, based on the nature of SST, node activation takes place. Thus, node's activation is dependent on its past real-time behaviour. Figure 1 depicts the framework used for the proposed TSC-HDM Model

The proposed work has been presented in 5 sections. A brief discussion about the existing models of diffusion for IM is presented in Section II. Section III presents the work proposed in this paper. Experimental setup and analysis of results are discussed in Section IV. Lastly, section V presents the final conclusion.

A. *Research contributions of this work*

Following are the key contributions of the proposed work:

1. We have proposed a novel diffusion model called Time Series Characteristic based Hurst-based Diffusion Model (TSC-HDM) wherein activation of node relies on the SST displayed by its past interaction pattern, which is quantified by computing the H value for the time series generated based on node's past interactions.
2. Our model takes into account each node's influence potential as a criteria for its activation. It does not rely on a random edge propagation probability like many of the existing works done before in the field of IM.

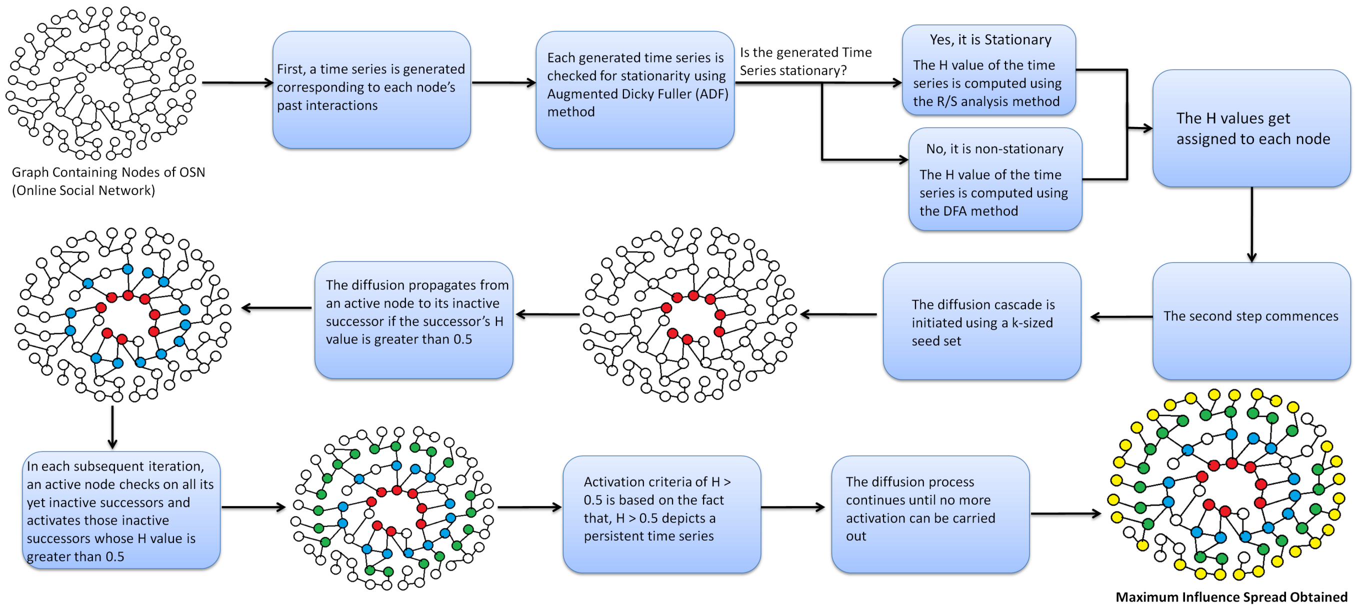


Figure 1: The framework used for our TSC-HDM Model

3. Our model computes H value on the basis of whether the generated time series is stationary or non-stationary. This has not been done before and is the first of its kind in the field of IM.
4. This has led to TSC-HDM model outperforming other existing models by 344%. Our model also achieved anywhere from 27% to 590% higher expected influence spread as compared to IC, WC, TV, and HBIM models.
5. Our model is the first to place so much importance on the characteristics of time series - whether they are stationary or non-stationary, and then calculate H value depending on that and this has led to our model performing much better. Apart from this, our paper will be an important landmark as it proves that node activation during the diffusion process should be dependent on the node's past temporal interaction behaviour.

II. Related Works

IM is explored as an algorithmic problem in the seminal work of Domingos and Richardson [3]. Following that, for addressing the problem of IM in OSNs, many diffusion models along with seed identification algorithms have been designed. Influence Maximization (IM) largely covers two activities, which are development of the diffusion model and identification of the seed node. "The various existing seed identification algorithms can be classified into two types - heuristic based approaches and greedy based approaches" [9, 22]. Seed selection in greedy based approaches is done on the basis of the nodes' marginal gain, whereas in heuristic based approaches seed selection is relent on the fulfilment of some condition(s).

Maximizing the influence spread across a given network is IM's aim. "The extent of influence spread attained by the chosen seed nodes is dependant upon the diffusion model [22]". Hence, quantifying the diffusion process and deciding the criteria for node activation is a critical activity. The following section briefly describes some of the existing diffusion models.

A. Existing Diffusion Models

Three diffusion models have been discussed by Kempe et al. namely *Independent Cascade* (IC),*Weighted Cascade* (WC), and *Linear Threshold* (LT) [23]. Under IC diffusion model, the network edges are assigned a randomly chosen pre-defined propagation probability. In LT, edges are assigned a randomly chosen pre-defined weight and a pre-defined random threshold value is allotted to each node. When the sum of edge weights of the active neighbours of a node exceeds its assigned threshold value, the node gets activated. WC model is a variant of the IC model, wherein each edge's probability of propagation is equivalent to the reciprocal of the recipient node's in-degree, whereas Chen et al. proposed the *Trivalency* (TV) diffusion model, where an edge's probability of propagation is randomly chosen from amongst three values, 0.1, 0.01, and 0.001 [24].

As can be observed, in all these aforementioned models, diffusion is driven by a randomly chosen value or weight assigned to the edges in the network. These models get behind the idea that an inactive node's chances of transitioning to an active state increases when the inactive node's active neighbours increase. Table I given below talks about related works on Influence Maximization

Dynamic Independent Cascade model (DIC), extends the IC model [25]. Under the DIC model, the edge propagation probability is randomly picked from a pre-defined distribution, but unlike IC, the edge's probability of propagation isn't the same for all of the edges. Deng et al. are credited to

Table I: Related works on Influence Maximization

References	Contributions	Strengths	Weaknesses
Kempe *et al.* [23]	Discussed about the widely popular Independent Cascade model (IC) in which a randomly chosen pre-defined propagation probability is given to every edge in a network	For an arbitrary instance of the IC Model, the resulting influence function σ() is submodular	Diffusion is dependent on a pre-defined edge propagation probability, which is not reliant on any node characteristic and is the same for all of the edges in the Online Social Network (OSN) in the Independent Cascade model
Chen *et al.* [24]	Their algorithm gives the users the ability to control the balance between the spread of influence of the algorithm and the running time	Their algorithm not only scales beyond million-sized graphs where greedy algorithm becomes infeasible, but their algorithm also performs consistently in all the size ranges, from small to very large	In their MIA model the authors assume that the seeds which are in S, influence each node v in the graph G through its MIIA(v, θ) (Maximum Influence In(Out)-Arborescence
Tong *et al.* [25]	The paper introduces the concept, which is adaptive seeding strategy and it also presents the Dynamic Independent Cascade (DIC) Model	The Dynamic Independent Cascade model is able to capture both - The dynamic aspects of a real social network as well as the uncertainness of the diffusion process	H-Greedy is a heuristic strategy, which is not effective for all the settings of the Dynamic Independent Cascade model. Moreover, with a round limit, their objective function is no longer submodular
Zhang *et al.* [38]	Proposed Opinion-based Cascading model (OC), which studies the spread of positive opinion across an OSN	An opinion indicator is attached to each node, which depicts a positive, negative, or neutral opinion	The opinion spread function O(), which is under the Opinion-based Cascading model is no longer submodular
Shrivastava *et al.* [39]	They propose a cost-effective diffusion model based on the classic IC model, but each active node tries to infect it's most influential neighbour with a predefined constant probability platform	The conventional IC model is outperformed by the proposed model by more than 5 times for the core super-spreaders and 2 times for the non-core super-spreaders	The paper assumes that for each of the infected edge, there is an equal cost
Our proposed approach	Our proposed diffusion model activates a node on the basis of the self-similarity trend exhibited by the past interaction pattern of the node	Our proposed model outperforms other state-of-art diffusion models by 344 %. Our model identifies the importance of the characteristics of time series - whether they are stationary or non-stationary, and then calculates H value depending on that	Assumption that the node can be expected to continue displaying similar behaviour in the future

have developed the *Credit Distribution with Node Features* model, where credits are assigned to a node on the basis of it's past interactions and depending on these credits, the node is activated [26]. Both, dynamic and static influence are taken into consideration while assigning credits to these nodes.

Voter model of diffusion, first a weight is assigned to every edge in the network [27, 28]. Thereafter, based on the probability proportional to the assigned edge weights, a node selects one of its successors and adopts the opinion held by the chosen successor. The Voter model has been extended by Li et al. in order to incorporate negative relationships [40]. In their model, if the edge to the chosen successor holds positive opinion, then the opinion of the successor is adopted by the node. Whereas if negative opinion is held by the edge, then an opinion opposite to the opinion of the successor is adopted by the node.

The *Opinion-based Cascading* (OC) model, which was developed by Zhang et al., studies the spread of positive opinion across an OSN [38]. Under the OC model, an opinion indicator is attached to each node, which depicts a positive, negative, or neutral opinion. When for a node its active predecessors' collective edge weight exceed the node's pre-defined threshold, activation of the node takes place. On activation, a node adopts the opinion of the incoming influence.

Saxena and Kumar have given *Activity-based IC* (AbIC) and *Activity-based LT* (AbLT) models of diffusion, which draw inspiration from the IC and LT models, respectively [31]. Under AbIC and AbLT models, the number of interactions a node initiated in the past are the basis on which edge propagation probability is computed.

An extension of the classic IC model is the *second-order IC model* [41]. In the second order IC model, activation takes place at both node and edge level. Each of the seed nodes propagates influence to its out-neighbours with a pre-defined influence probability. Thus, node to node activation takes place in the first order. Additionally, whenever a seed node successfully activates it's out-neighbour, the connecting edge also becomes active. Thereafter, in the second-order, edge to edge influence propagation takes place, in which every edge which is active attempts to convert their inactive out-edges to active ones with some constant probability.

Yu and Li, who developed the *CMMI* model for diffusion incorporate user preferences and diffusion enhancement [42]. Under CMMI model, a node on receiving information, changes state from inactive to active. Once a node becomes active, it can then move into accepted or rejected state, if the node's probability of accepting product is greater or less than a pre-defined node transition threshold, respectively. Probability of accepting product is computed based on internal influence (user preference), influence of adjacent nodes, and external influence. Accepted state is indication that the node has accepted the product and will be propagating it further, whereas rejected state is an indication that the node has rejected the product and has refused to propagate it further.

Hudson and Khamfroush have presented the *Behavioral Independent Cascade* diffusion model (BIC) for the purpose of opinion maximization, where user nodes' opinions and their personalities form the basis for propagation probabilities [43]. Though the BIC model makes use of IC model framework, it differs from it in certain aspects. Unlike the classical IC model, propagation probabilities in the BIC model are not static and pre-determined. Rather, they are dynamically computed before each activation using behaviour and opinion parameters assigned to each node. Further, BIC model permits multiple activation attempts.

Li et al. have proposed the *User Behavior Model* (UBM) for undirected networks [44]. Under UBM, a message is sent by every node to all its neighbours with a pre-defined probability (similar to classic IC model [23]). On the successful receipt of the message, a node gets activated and responds to the message as per its personal interest, and may then send out a message (forward or reply) to all its neighbours. Diffusion stops when all nodes have tried responding to the received messages.

With the aim of developing an integrated information diffusion model, Kong et al. have presented the *Diffusion and Influence Model* (DIM), which combines the two aspects - diffusion as well as information influence together [45]. DIM consists of two stages, one is called the diffusion stage and the other is called the influence stage. These two stages are represented using diffusion and influence functions, respectively. The diffusion function is a probabilistic function depicting the chance of information being spread by users. The influence function is also a probabilistic function, representing how users in a network get influenced. DIM presents a unified approach for implementing classic diffusion models, which can be done by varying the settings of the diffusion and influence functions.

Shrivastava et al. have proposed a cost-effective diffusion approach motivated by the classic IC model [39]. Under the proposed model, each active node tries to infect its most influential neighbour with a pre-defined constant probability. In case the node fails to activate its most influential neighbour, it then tries infecting the second most influential neighbour. If it fails again, it tries infecting the third most influential neighbour and so on.

In the Charismatic Transmission in Influence Maximization algorithm developed by Kazemzadeh et al., nodes having high "charismatic power" are selected [46]. In this algorithm, the nodes which have a high correlation with other communities' influential nodes are selected, which leads to optimal diffusion.

Fu et al. targetted the dynamic OSNs and worked on extending the IC model to a Dynamic Social Network Dissemination model which is based on effective links [47]. They have presented a two-stage IM algorithm called Outdegree Effective Link, which utilises node degree and effective links to tackle the ever changing nature of dynamic OSNs. The authors also discuss the influence of node interaction between nodes on information dissemination in dynamic social networks.

Li et al. have proposed the Layered Gravity Bridge Algorithm, wherein a community detection technique has been used for obtaining communities in social networks and thereafter, bridge nodes which can be considered as possible candidate seeds are identified [48]. The detected communities are then amalgamated into larger communities and new bridge nodes are realized. Finally, all the candidate seed nodes are sorted through an improved gravity model, in order to determine the final seed nodes.

Rezvanian et al. have presented a diffusion model based on a stochastic graph, wherein the influence probabilities related with the links are unknown random variables [49]. They then go on to use the set of learning automata in their proposed diffusion model to create an approach which estimates the influence probabilities by sampling the links of the stochastic graph.

III. Proposed Time Series Characteristic based Hurst-based Diffusion Model (TSC-HDM)

This section first of all provides a brief overview of the concepts that have been used for our model and the later part of the section explains the proposed model.

A. Rescaled Range (R/S) analysis

R/S analysis technique helps in assessing a time series's variability over time [50]. For computing H using R/S analysis technique, firstly divide the full-length time series into various shorter, varying length time series. Thereafter, computation of an average value for the R/S is done [51]. For a given time series X_t of length n, where $t \in n$, R/S is calculated as follows:

1) Firstly, mean (x) of given time series (X) is computed using Eqn. 1:
$$x = \frac{1}{n}\sum_{j=1}^{n} X_j \tag{1}$$
2) Then, generate the mean-adjusted series (Z) using Eqn. 2:
$$Z_t = X_t - x, \quad t = 1, 2, ..n \tag{2}$$
3) Then, compute cumulative deviate series (C) using Eqn. 3:
$$C_t = \sum_{j=1}^{t} Z_j, \quad t = 1, 2, ..n \tag{3}$$
4) Generate range (R) and standard deviation (S) series as indicated in Eqn. 4 and 5:
$$R_t = max(C_1, C_2...C_n) - min(C_1, C_2...C_n), \;\; t = 1, 2, ..n \tag{4}$$
$$S_t = \sqrt{\frac{1}{t}\sum_{j=1}^{t}(Y_j - u)^2}, \quad t = 1, 2, ..n \tag{5}$$

Use Eqn. 6 to compute Rescaled Range value:
$$(R/S)_t = \frac{R_t}{S_t}, \quad t = 1, 2, ..n \tag{6}$$

Thereafter, estimation of H is done by fitting a straight line through the plot of the values of *log(R/S) vs. log (t)* and (n being the time series length):
$$(R/S)_t \propto t^H \tag{7}$$

Slope of the fitted line represents H [50, 51].

B. Detrended Fluctuation Analysis

(DFA) DFA is used for quantifying the SST of a non-stationary time series [36, 52, 53]. For computing H using DFA technique, a time series of length K is first integrated. Assume a bounded time series y_t of length K, where $t \in K$. Eqn. 8 first of all converts the bounded time series to unbounded time series X_t [54].

$$x(t) = \sum_{i=1}^{t} \left(y_i - \langle y \rangle\right), \tag{8}$$

where $\langle y \rangle$ denotes the mean of the time series and $x(t)$ denotes the summation profile. Then, integrated time series is divided into shorter time series (boxes) of length k samples each. Next, the fluctuation (mean-squared residual) is calculated using:

$$F(k) = \sqrt{\frac{1}{D}\sum_{t=1}^{D}[x(t) - x_{\Delta k}(t)]^2}, \tag{9}$$

D signifying the total number of data points. Thereafter, using fluctuation the self similarity in the time series is calculated (Eqn. 10), which is further used to calculate the H.

$$F(k) \propto k^{\alpha}, \tag{10}$$

$$ln(F(k)) = \alpha ln(k) + ln(C) \tag{11}$$

where C and α denote constant of proportionality and scaling exponent estimated using least-squares fit, respectively. Finally, H is calculated using Eqn. 12.

$$H = \alpha(2) - 1 \tag{12}$$

Development of a model for diffusion of information (influence) across an OSN is an important aspect of IM. A node is either active (influenced) or inactive. Diffusion model outlines the condition(s) to be fulfilled by a node to transition from inactive to active state.

As stated earlier, in the widely popular IC model, diffusion is not reliant on any node characteristic, rather it is dependent on a pre-decided probability of propagation [23]. Further, all edges are assigned the same propagation probability. In the LT diffusion model, a threshold is assigned to each node, which again is not dependent on any node characteristic and is the same for all nodes in the OSN [23]. This threshold value assigned to a node is to be crossed by the incoming influence, only then the node can get activated. Additionally, influence propagated over an edge is a randomly chosen pre-defined value, which is again same for all edges. The propagation probability in WC model, is equivalent to the reciprocal of the recipient node's degree [23]. TV model picks the propagation probability from the set 0.001, 0.01, 0.1 at random [24] . Under the DIC model, propagation probability is randomly picked from a pre-defined distribution[25].

It can be observed that in most of these popular diffusion models, node activation does not rely on any node characteristics, and is mostly based on a randomly chosen edge propagation probability or weight. Instead of selecting a random propagation probability, a better approach would be to consider each node's influence potential as a criterion for its activation. To assess the influence potential of a node, its structural and temporal characteristics are usually considered.

As mentioned before, an aspect that has not been explored much in context with OSN users (nodes) is the SST displayed by a node's past interactions. Study of human behaviour conducted by Fan et al. suggests that humans have a tendency of repeating their actions in real world, i.e., human behaviour over time can be exhibits statistical self-similarity [33]. On similar lines, OSN users can also be presumed to repeat their behaviour and hence exhibit statistical self-similarity [35]. User behaviour is a time dependent phenomenon and can be represented as a time series and the extent of self-similarity (auto-correlation) in this generated time series can be assessed using Hurst exponent (*H*) [33, 36, 53]. "*H* measures the relative inclination of a time series to either regress strongly to the mean or to cluster in one direction [55]."

H helps figure out whether a time series is exhibiting an anti-persistent, random, or persistent trend. The value of *H* lies between 0 and 1. *H* value between 0 – 0.5 signifies an anti-persistence (long-term switching between high and low values). *H* = 0.5 signifies uncorrelated (random) time series, for which establishment of any trend becomes difficult owing to the non-existence of correlation. *H* between 0.5 – 1 signifies persistence indicating the possibility of a high value being followed by another high value, and the trend will possibly remain so for a long time.

Drawing motivation from the aforementioned notion, Saxena and Saxena developed the HBIM diffusion model which utilizes degree of the node and the *H* value corresponding to the SST of its past interactions, as the criteria for node activation [35]. Under HBIM model, a node gets activated if its degree is greater than the average node degree in the network, and if its *H* value is greater than 0.5. In the HBIM model R/S analysis technique is used to compute the value of *H* corresponding to a generated time series. But, a time series can be either stationary or non-stationary. Kirichenko et al., Resta, and Wairimu have found that applying the same method for computation of H for stationary as well as non-stationary time series may lead to inaccurate evaluations [36, 37, 53]. R/S analysis method has been found to be more suitable for computing H corresponding to a stationary time series, and *Detrended Fluctuation Analysis* (DFA) method has been found to give more accurate evaluations for non-stationary time series [36, 37, 53].

For getting a realistic estimation of a node's influence potential, its past temporal behaviour must be regarded as a significant contributor. Driven by this belief, a novel *Time Series Characteristic based Hurst-based Diffusion Model* (TSC-HDM) is being proposed in this paper, which augments the aforementioned HBIM model and aims to develop a diffusion model for IM, in which diffusion is based on the H value corresponding to the time series of the recipient node's past interactions, and the method used for computing H is reliant on the characteristic (stationary or non-stationary) of the generated interaction time series [35]. The proposed work supports the premise that node activation during the diffusion process should be dependent on its influence potential, instead of some pre-defined probability or threshold [31, 35].

The HBIM model makes use of R/S analysis method for computing *H* value corresponding to each generated time series, while not considering the nature (stationary or non-stationary) of the time series. Though R/S analysis is prevalently used for computing *H*, researchers have been found that DFA method is more suitable for making accurate evaluations pertaining to a non-stationary time series [36, 37, 53]. Hence, in the proposed work, each generated time series is first checked for stationarity, and thereafter depending upon the characteristic of the time series (i.e., stationary or non-stationary). R/S analysis and DFA methods have been used for computation of *H* for stationary and non-stationary time series respectively.

Algorithm 1 TSC-HDM Diffusion Model

```
Graph G = (V, E), seed_nodes []: seed node set of size k
expec_spread: expected spread to be achieved
for each node n do
    Generate time series based on past interactions, TS(n),
    for the time span under consideration  Using Augmented
    Dicky Fuller method, check if TS(n) is stationary or
    non-stationary  if TS(n) is stationary then
        Compute H for TS(n) using the Rescaled Range
        Analysis method
    else
        Compute H for TS(n) using the Detrended Fluctuation
        Analysis method
influenced_nodes[] := seed_nodes[]  expec_spread :=
LEN(influenced_nodes[])
for each node in influenced_nodes do
    for each successor of node do
        if successor not in influenced_nodes then
            if IF H(successor) > 0.5 then
                expec_spread := expec_spread + 1
                influenced_nodes[expec_spread] := successor
Return expec_spread
```

Under the proposed TSC-HDM model, diffusion is modelled as a two-step process. The first step focuses on the computation and assignment of *H* to each node, based on the SST exhibited by the node's past interactions. Thus, for each node a time series is first created corresponding to the node's past interaction pattern. Thereafter, the generated times series is checked for stationarity using the *Augmented Dicky Fuller* (ADF) method. ADF is the most commonly used statistical test used for analyzing the stationary of a time series. If the generated time series is found to be stationary, then the *H* value is computed using the R/S analysis method. However, if the generated time series exhibits non-stationarity, then DFA method is used for computing the *H* value corresponding to the SST depicted by the time series under consideration. Once the *H* values get assigned to each node, the second step commences.

In the second step, the diffusion cascade is initiated using a *k*-sized seed set. The seed set comprises of *k* nodes which are considered to be in an active state at the beginning of the diffusion process, and act as the initiator of the diffusion process. In the TSC-HDM model, diffusion propagates from an active node to its inactive successor(s) if for the successor $H > 0.5$. *influenced_nodes* depicts a list of active nodes. At the beginning of the diffusion process, the list *influenced_nodes* contains only the seed nodes. *expec_spread* denotes the count of influenced nodes, i.e., the number of nodes activated by the diffusion process, and is equal to length of the list *influenced_nodes*.

During the first iteration, the *k* active seed nodes contact each of their successors and check for yet inactive successors. If the *H* value for the inactive successor is greater than 0.5, then the state of the successor is swapped to active, and it is added to the list *influenced_nodes* whilst increasing the *expec_spread* by 1. Thereafter, in each subsequent iteration, an active node (from the list *influenced_nodes*) checks on all its yet inactive successors and activates those whose $H > 0.5$. With each iteration the nodes that change state from inactive to active keep on adding to the *influenced_nodes* list and *expec_spread* keeps updating as per the length of the list *influenced_nodes*. In the proposed TSC-HDM model node activation takes place if the node's predecessor is active and the node's self *H* value is greater than 0.5. "Activation criteria of $H > 0.5$ is based on the fact that, $H > 0.5$ depicts a persistent time series [35]". Diffusion process keeps continuing until no further activation cannot be done. When the process culminates, *expec_spread* denoted the total number of nodes influenced by the diffusion process. Algorithm 1 presents an outline of the proposed TSC-HDM diffusion model.

IV. Experimental Set-up and Results

A. Experimental Set-up

All the experiments were run on Windows Operating System in a Python Environment. The machine on which the experiments were run had an Intel Core i5-8250U CPU, running at a base clock speed of 1.6GHz. The machine had 8GB RAM.

For evaluating the proposed TSC-HDM diffusion model, its performance was compared with four existing diffusion models for IM, namely IC, WC, TV, and HBIM models. In the IC model, the edge's probability of propagation had been set as 0.1 for evaluation purposes.

Model evaluation was done using four publicly available real-world social network datasets. Each of the four datasets is a directed temporal network, wherein an edge (u, v, t) signifies that user u interacted with (sent message to) user v at timestamp t. Following are the datasets used:

- *UC Irvine messages*[1] - The dataset contains messages shared by users on an online social network of University

[1] http://snap.stanford.edu/data/CollegeMsg.html

of California, Irvine. Each directed edge indicates a message exchange between the students of the university.
- *Email EU-Core* [2] - The dataset represents an email network from a large European research institution. A directed edge between two nodes indicates an exchange of email between the two institution members.
- *Math Overflow* [3] - The dataset depicts a network of interactions among users of the Math Overflow website. Interactions among users could be answering a question or commenting on a question or answer.
- *Linux Kernel mailing list* [4] - The dataset depicts the communication network of the Linux kernel mailing list. A directed edge between two nodes represents a reply from a user to another.

Each dataset, for the purpose of evaluation, was split into two parts. For calculating H as well as creating time series, the first part was used. For studying the diffusion process, the second part was used. First 100 days data of the UC Irvine dataset, first 1-year data of the Email EU-Core dataset, first 4 years data of the Math Overflow dataset, and first 5 years data of the Linux Kernel dataset was used to compute H and create the interaction time series. To study the spread of influence, the remaining data was used. Details of all the four datasets have been shown in Table II.

The four datasets used have varying time spans ranging from 193 days to 2922 days. Datasets having varying time spans have been used so as to validate the proposed method for shorter as well as longer interaction periods between the users. Considering UC Irvine dataset, for each node the interaction time series gets generated for 193 days which is a short time span. For Email EU-Core the length of the time series generated for each node is 803 days, which is a medium sized time span. For Math Overflow a time series for 2350 days is generated for each node, and for Linux Kernel mailing list the time series generated for each node covers it a span of 2922 days, both of which are longer time durations.

B. Results

The proposed TSC-HDM diffusion model's performance is compared to the performance of four diffusion models, namely IC, WC, TV, and HBIM models. The four datasets mentioned previously have been used to do evaluation. Initial sets of seeds having different sizes (k = 10, 20, 30, 40, and 50) were created using the below-mentioned algorithms for seed selection (number of nodes in the initial set of seeds is denoted by k):

- *Degree* - The k nodes which have the highest degree are selected as the seed nodes by this seed selection algorithm [23].
- *SingleDiscount* - In this algorithm, the node with highest degree becomes the seed node, and the degree of every inactive neighbour of that node gets discounted by 1 [56].
- *DegreeDiscountIC* - It this scheme, node having highest degree node is chosen as seed node. After that, the degree of the node's yet inactive neighbours is discounted based on the degree of that inactive neighbour and how many of its neighbours have been selected as seeds [56].

The generated seed sets are given as input to the IC, WC, TV, HBIM and proposed TSC-HDM models. The spread attained by the given seeds under the 5 diffusion models under consideration has been computed and compared. If we start the process with k seed nodes, the overall number of nodes which are expected to be influenced when we reach the end of the diffusion process, are denoted by the Influence Spread attained.

The spread of influence is calculated independently for all the five different seed set sizes, i.e., k = 10, 20, 30, 40, and 50. For the purpose of representing, the spread of influence attained by the seed set having k = 50 has been shown in Figures 2(a), 2(b), 2(c), and 2(d) corresponding to the UC Irvine, the Email EU-Core, the Math Overflow, and the Linux Kernel datasets respectively.

Obtained results depict that the spread of influence attained by our proposed TSC-HDM model is far greater than the influence spread attained by the remaining four existing models under consideration, by the generated seed sets. Hence it has been found that, a greater number of nodes are getting influenced in the proposed TSC-HDM model, in comparison with the IC, WC, TV, as well as HBIM models.

Figure 2(a) shows that for the UC Irvine dataset, the spread of influence achieved by the seed set initialized using Degree algorithm under our TSC-HDM model is about 351% more than the influence spread attained under Independent Cascade model, 29% more than WC model, 644% more than TV model, and 114% more than HBIM model. For seed set generated using SingleDiscount algorithm, spread achieved under TSC-HDM is 332% higher than IC, 27% higher than WC, 627% higher than TV model, and 114% more than HBIM model. For seed nodes chosen using DegreeDiscountIC algorithm, influence spread under TSC-HDM is 315% more than IC, 31% more than WC, 616% more than TV, and 110% more than the spread achieved under the HBIM model.

Figure 2(b) shows the spread of influence attained by the seed set generated pertaining to Email EU-Core dataset. Spread achieved by seeds chosen utilizing the Degree algorithm by our TSC-HDM model is about 67% higher than the influence spread attained by the Independent Cascade model, 96% higher than WC model, and 275% higher than TV model. Seed nodes selected using SingleDiscount algorithm, attain 67% more spread under TSC-HDM as compared to IC model, 94% more than WC, and 264% more than TV model. For seeds chosen using DegreeDiscountIC algorithm, spread under TSC-HDM is 65% more than IC, 103% more than WC, and 245% higher than the spread obtained under the TV model. Influence Spread achieved by seed sets chosen by all the seed selection algorithms which are under consideration,

[2]http://snap.stanford.edu/data/email-Eu-core-temporal.html
[3]http://snap.stanford.edu/data/sxmathoverflow.html
[4]http://konect.uni-koblenz.de/networks/lkml-reply

Table II: Details of used datasets

Dataset Name	Total Nodes	Static Edges	Temporal Edges	Data Available For (Duration)	Time Span (in days)
UC Irvine messages	1,899	20,296	59,835	15-04-2004 to 26-10-2004	193
Email EU-Core	986	24,929	3,32334	01-01-1970 to 15-03-1972	803
Math Overflow	24,818	506,550	239,978	29-09-2009 to 06-03-2016	2350
Linux Kernel mailing list	63,399	242,976	1,096,440	01-01-2006 to 01-01-2014	2922

is about 88% higher in our TSC-HDM model, in comparison with the influence spread attained by the same seeds in the HBIM model.

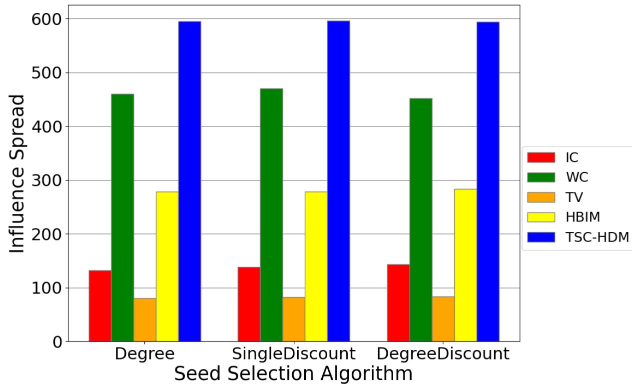
Figure 2(c) shows that for the Math Overflow dataset, influence spread attained by our TSC-HDM model is about 252% more than the influence spread attained by the Independent Cascade model, 134% more than WC, and 852% more than TV model for those seed sets which are initialized by making use of the Degree algorithm. Spread of influence in TSC-HDM is 250% higher than IC, 136% more than WC and 855% more than TV for seed set generated using SingleDiscount algorithm. For seed nodes selected using DegreeDiscountIC algorithm, spread under proposed TSC-HDM is 248% higher than IC, 150% higher than WC, and 840% higher than TV model. Spread of influence attained in our TSC-HDM model is approximately 95% higher than the influence spread attained in the HBIM model by all the seed sets which are under consideration.

Figure 2(d) shows that the spread of influence attained by seeds chosen using Degree and SingleDiscount algorithms is approximately 253% higher under TSC-HDM in comparison with the influence spread attained by the same set of seeds when they are under the Independent Cascade model, and approximately 259% higher for seed set generated using DegreeDiscountIC algorithm. Compared to the spread of influence attained by the WC model, spread attained under proposed TSC-HDM increases by approximately 157% for Degree seed set, 155% for SingleDiscount seed set and 200% for DegreeDiscountIC seed set. Compared to the spread attained under TV model, the spread achieved under TSC-HDM model increases by approximately 590% for seeds of Degree and SingleDiscount algorithms, and 595% for seeds selected using DegreeDiscountIC algorithm. Compared to the spread attained under HBIM model, the spread of influence attained by our TSC-HDM model is approximately 50% more for seeds of Degree and SingleDiscount algorithms, and 49% for seeds selected using DegreeDiscountIC algorithm.

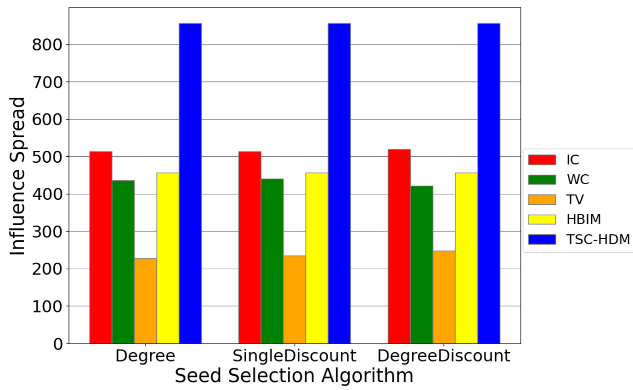
Thus, it has been affirmed by the results obtained, that in comparison with the influence spread attained under four existing diffusion models under consideration, namely IC, WC, TV, and HBIM models, there is a higher spread of influence in our TSC-HDM model. Consequently, on account of the results obtained, we can conclude that aspects of a node's past interaction pattern when taken into consideration, results in a better assessment of the influence potential of a node. Hence, it should be considered as a criterion for node activation during the diffusion process. Furthermore, on comparing the obtained results for HBIM and proposed TSC-HDM models, it can be observed that using different methods for computing H value significantly improves the expected influence spread.

Researchers generally don't present a comparison of research work related to Influence Maximization in Online Social Networks with existing approaches because the results which are obtained are not comparable directly. Results obtained by a particular approach depend on different parameters such as different types of datasets used. Datasets which are used, many times differ with regards to the type of information they consist of as well as the time period for which the dataset's data is being considered. For the purpose of comparison, the proposed TSC-HDM model's performance, with regards to the spread of influence achieved, is compared to nine state-of-the-art existing approaches for IM (where each uses the initial seed set size, k = 50), namely *LIR*, *A-Greedy*, *LPIMA*, *Genetic Algorithm with Dynamic Probabilities*, *NeighborsRemove*, *DegreeDecrease*, *IGIM*, *IRR*, and *PHG* [25, 57–63]. The average percentage of spread of influence attained by an initial seed set of size k = 50, i.e., having 50 seed nodes, has been used as the base for comparison.

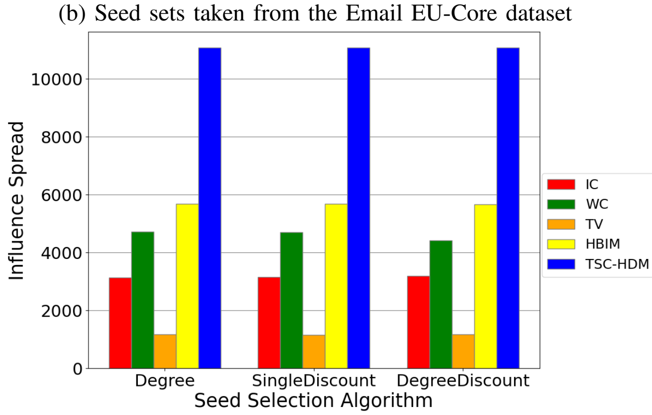
As per the reported results, *LIR* achieves an average percentage of spread of influence of 9.13%, *A-Greedy* attains 1.6%, *LPIMA* achieves 2.7% and *Genetic Algorithm with Dynamic Probabilities* achieves 11.33% [25, 57–59]. *NeighborsRemove* and *DegreeDecrease* have been reported to achieve an average percentage of spread of influence of 18.91% as well as 18.58% respectively [60]. Average spread percentage for IGIM, as per reported results, is 13.27%, for IRR it is 3.39%, and for PHG it is 7.98% [61–63]. However, an average percentage of spread of influence of 84% is achieved under the proposed TSC-HDM model. Figure 3 compares the average percentage of influence spread attained under these ten aforementioned approaches towards IM in OSN.



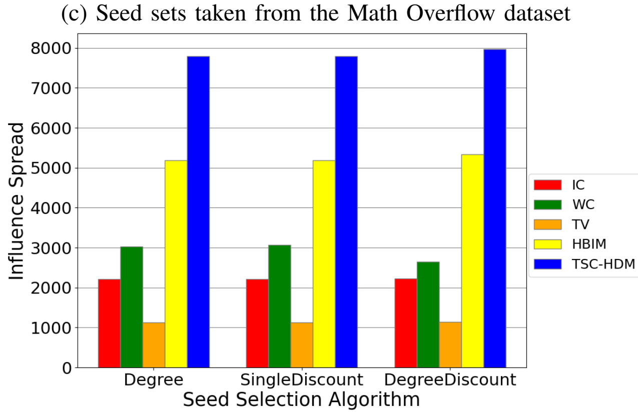
(a) Seed sets taken from the UC Irvine dataset



(b) Seed sets taken from the Email EU-Core dataset



(c) Seed sets taken from the Math Overflow dataset



(d) Seed sets taken from the Email Network dataset

Figure 2: Influence spreads attained by the generated seed sets in the IC, WC, Trivalency and TSC-HDM models

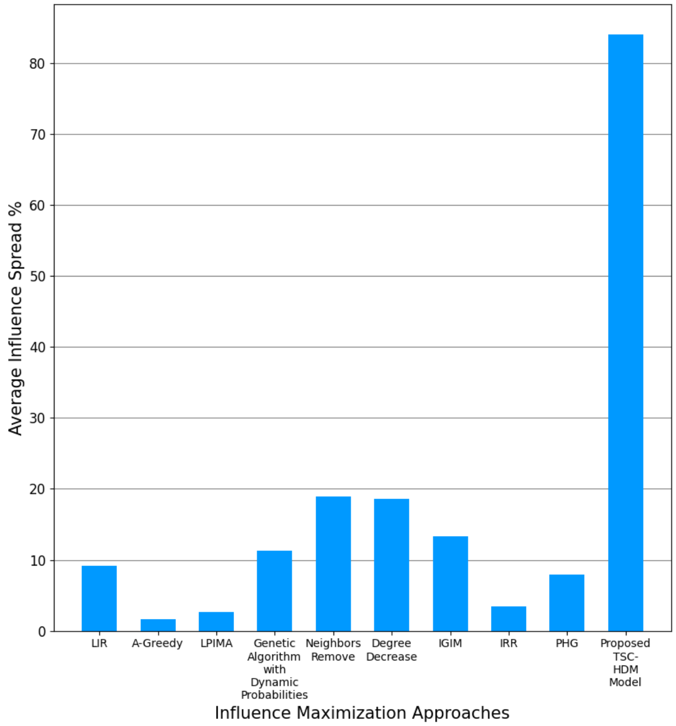


Figure 3: Average % of influence spread under TSC-HDM compared to state of art algorithms

V. Conclusion

OSNs have long been researched upon to better understand online social structures and behaviour. IM aims to maximize influence spread in an OSN and "the influence spread attained by the chosen seed nodes is heavily affected by the underlying diffusion model [22]". In many prevalent diffusion models, diffusion depends either on a pre-defined edge propagation probability, or on the node's degree. However, a node's past interaction pattern can present a realistic view of actual diffusion between the node and its neighbours.

In the proposed work, Time Series Characteristic based Hurst-based Diffusion Model (TSC-HDM) for IM has been presented, which is motivated by the belief that to assess a node's influence potential, its actual past interaction behaviour should be considered as a significant contributor. TSC-HDM draws inspiration from the HBIM model, wherein node activation is done based on the statistical SST (quantified using Hurst exponent (H)) displayed by the node's past interactions [35]. However, in TSC-HDM, the nature of each time series is first explored, and then separate methods have been used for computing H value depending on whether time series is stationary or non-stationary. TSC-HDM has been evaluated using four real-world OSN datasets and results reveal that TSC-HDM outperforms IC, WC, TV, and HBIM models [23, 24, 35]. Hence, taking the node's past interaction details into consideration, and using different methods for computing H value for a time series, whilst considering its stationarity characteristic, remarkably enhances the spread achieved by a seed set.

VI. Future Scope

A limitation of the work presented in this paper is that in alignment with much of the research works carried pertaining to IM in OSNs, a static snapshot of the data pertaining to the node's past interactions is being used to assess its temporal behaviour. A node's past static behaviour forms the base of the proposed TSC-HDM diffusion model, with the assumption that the node can be expected to continue displaying similar behaviour in the future. Approaches developed by taking the other aspects of node's structure and behaviour into consideration can be further explored.

References

[1] N. Magaia, P. Ferreira, P. R. Pereira, K. Muhammad, J. Del Ser, and V. H. C. de Albuquerque, "Group'n route: An edge learning-based clustering and efficient routing scheme leveraging social strength for the internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 19 589–19 601, 2022.

[2] X. Zhu, D. He, Z. Bao, M. Luo, and C. Peng, "An efficient decentralized identity management system based on range proof for social networks," *IEEE Open Journal of the Computer Society*, vol. 4, pp. 84–96, 2023.

[3] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining*, 2001, pp. 57–66.

[4] M. Han and Y. Li, "Influence analysis: A survey of the state-of-the-art," *Mathematical Foundations of Computing*, vol. 1, p. 201, 2018.

[5] K. Li, L. Zhang, and H. Huang, "Social influence analysis: Models, methods, and evaluation," *Engineering*, vol. 4, pp. 40–46, 02 2018.

[6] S. Peng, G. Wang, and D. Xie, "Social influence analysis in social networking big data: Opportunities and challenges," *IEEE Network*, vol. 31, no. 1, pp. 11–17, 2017.

[7] J. Sun and J. Tang, *A Survey of Models and Algorithms for Social Influence Analysis*. Springer, Boston, MA, 03 2011, pp. 177–214.

[8] J. Del Ser, M. N. Bilbao, I. Laña, K. Muhammad, and D. Camacho, "Efficient fake news detection using bagging ensembles of bidirectional echo state networks," in *2022 International Joint Conference on Neural Networks (IJCNN)*, 2022, pp. 1–7.

[9] S. Peng, Y. Zhou, L. Cao, S. Yu, J. Niu, and W. Jia, "Influence analysis in social networks: A survey," *Journal of Network and Computer Applications*, vol. 106, pp. 17–32, 03 2018.

[10] E. Stevenson, V. Rodriguez-Fernandez, H. Urrutxua, and D. Camacho, "Benchmarking deep learning approaches for all-vs-all conjunction screening," *Advances in Space Research*, 2023.

[11] G. Zhang, J. Wu, G. Jeon, Y. Chen, Y. Wang, and M. Tan, "Towards understanding metaverse engagement via social patterns and reward mechanism: A case study

of nova empire," *IEEE Transactions on Computational Social Systems*, 2022.

[12] M. Li, X. Wang, K. Gao, and S. Zhang, "A survey on information diffusion in online social networks: Models and methods," *Information*, vol. 8, no. 4, 2017.

[13] J. Huertas-Tato, A. Martín, J. Fierrez, and D. Camacho, "Fusing cnns and statistical indicators to improve image classification," *Information Fusion*, vol. 79, pp. 174–187, 2022.

[14] M. Imran, S. Khan, H. Hlavacs, F. A. Khan, and S. Anwar, "Intrusion detection in networks using cuckoo search optimization," *Soft Computing*, vol. 26, no. 20, pp. 10 651–10 663, 2022.

[15] D. B. Kurka, A. Godoy, and F. J. V. Zuben, "Online social network analysis: A survey of research applications in computer science," *CoRR*, vol. abs/1504.05655, 04 2016.

[16] M. Rehman, I. T. Javed, K. N. Qureshi, T. Margaria, and G. Jeon, "A cyber secure medical management system by using blockchain," *IEEE Transactions on Computational Social Systems*, 2022.

[17] B. Chen, L. Wu, H. Wang, L. Zhou, and D. He, "A blockchain-based searchable public-key encryption with forward and backward privacy for cloud-assisted vehicular social networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 5813–5825, 2019.

[18] J. Shepherd. 28 essential influencer marketing statistics you need to know in 2023. Accessed on 23rd April 2023. [Online]. Available: https://thesocialshepherd.com/blog/influencer-marketing-statistics

[19] E. Rodrigue. 31 influencer marketing stats to know in 2023. Accessed on 23rd April 2023. [Online]. Available: https://blog.hubspot.com/marketing/influencer-marketing-stats#:~:text=Influencer%20Marketing%20Effectiveness,-7.&text=50%25%20of%20Millennials%20trust%20product,an%20effective%20form%20of%20marketing

[20] D. M. Institute. 20 surprising influencer marketing statistics. Accessed on 23rd April 2023. [Online]. Available: https://digitalmarketinginstitute.com/blog/20-influencer-marketing-statistics-that-will-surprise-you

[21] K. Muhammad, M. Sajjad, I. Mehmood, S. Rho, and S. W. Baik, "Image steganography using uncorrelated color space and its application for security of visual contents in online social networks," *Future Generation Computer Systems*, vol. 86, pp. 951–960, 2018.

[22] Y. Li, J. Fan, Y. Wang, and K.-L. Tan, "Influence maximization on social graphs: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, pp. 1852–1872, 02 2018.

[23] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003, pp. 137–146.

[24] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'10)*, 2010, pp. 1029–1038.

[25] G. Tong, W. Wu, S. Tang, and D. Du, "Adaptive influence maximization in dynamic social networks," *IEEE/ACM Transactions on Networking*, vol. 25, no. 1, pp. 112–125, 02 2017.

[26] X. Deng, Y. Pan, H. Shen, and J. Gui, "Credit distribution for influence maximization in online social networks with node features," *Journal of Intelligent and Fuzzy Systems*, vol. 31, pp. 979–990, 2016.

[27] E. Even-Dar and A. Shapira, "A note on maximizing the spread of influence in social networks," in *WINE'07: Proceedings of the 3rd international conference on Internet and network economics*, 2007, pp. 281–286.

[28] P. Clifford and A. Sudbury, "A model for spatial conflict," *Biometrika*, vol. 60, no. 3, pp. 581–588, 1973.

[29] A. Ullah and S. Anwar, "The effective use of information technology and interactive activities to improve learner engagement," *Education Sciences*, vol. 10, no. 12, p. 349, 2020.

[30] Y. Wang, W. Wang, Q. Chen, K. Huang, A. Nguyen, S. De, and A. Hussain, "Fusing external knowledge resources for natural language understanding techniques: A survey," *Information Fusion*, vol. 92, pp. 190–204, 2023.

[31] B. Saxena and P. Kumar, "A node activity and connectivity-based model for influence maximization in social networks," *Social Network Analysis and Mining*, vol. 9, 2019.

[32] M. Amin, D. Shehwar, A. Ullah, T. Guarda, T. A. Tanveer, and S. Anwar, "A deep learning system for health care iot and smartphone malware detection," *Neural Computing and Applications*, pp. 1–12, 2020.

[33] C. Fan, J.-L. Guo, and Y.-L. Zha, "Fractal analysis on human behaviors dynamics," *Physica*, vol. abs/1012.4088, pp. 6617–6625, 2012.

[34] Q. Liu, X. Zhao, W. Willinger, X. Wang, B. Zhao, and H. Zheng, "Self-similarity in social network dynamics," *ACM Transactions on Modeling and Performance Evaluation of Computing Systems (TOMPECS)*, vol. 2, pp. 1–26, 10 2016.

[35] B. Saxena and V. Saxena, "Hurst exponent based approach for influence maximization in social networks," *Journal of King Saud University - Computer and Information Sciences*, 2019.

[36] L. Kirichenko, T. Radivilova, and Z. Deineko, "Comparative analysis for estimating of the Hurst exponent for stationary and nonstationary time series," *Information Technologies and Knowledge*, vol. 5, p. 371–388, 2011.

[37] M. J. Wairimu, "Features affecting Hurst exponent estimation on time series," Master's thesis, Jomo Kenyatta University of Agriculture and Technology, 2013, [Online accessed 2 December 2019].

[38] H. Zhang, T. N. Dinh, and M. T. Thai, "Maximizing the spread of positive influence in online social networks," in *2013 IEEE 33rd International Conference on Distributed*

Computing Systems, 2013, pp. 317–326.

[39] S. Shrivastava, D. Chaudhary, Y. Gupta, and S. Sukhija, “Cost effective influence maximisation,” 1 2021, pp. 90–93.

[40] Y. Li, W. Chen, Y. Wang, and Z. Zhang, “Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships,” *WSDM 2013 - Proceedings of the 6th ACM International Conference on Web Search and Data Mining*, pp. 657–666, 11 2011.

[41] W. Tang, G. Luo, Y. Wu, L. Tian, X. Zheng, and Z. Cai, “A second-order diffusion model for influence maximization in social networks,” *IEEE Transactions on Computational Social Systems*, vol. 6, no. 4, pp. 702–714, 2019.

[42] S. Yu and Y. Li, “Maximizing social network influences based on user preferences,” in *2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS)*, 7 2020, pp. 987–993.

[43] N. Hudson and H. Khamfroush, “Behavioral information diffusion for opinion maximization in online social networks,” *IEEE Transactions on Network Science and Engineering*, 10 2020.

[44] P. Li, H. Nie, F. Yin, J. Liu, and D. Zhou.

[45] X. Kong, Z. Gu, and L. Yin, “A unified information diffusion model for social networks,” 7 2020, pp. 38–44.

[46] F. Kazemzadeh, A. A. Safaei, and M. Mirzarezaee, “Influence maximization in social networks using effective community detection,” *Physica A: Statistical Mechanics and its Applications*, vol. 598, p. 127314, 2022.

[47] B. Fu, J. Zhang, H. Bai, Y. Yang, and Y. He, “An influence maximization algorithm for dynamic social networks based on effective links,” *Entropy*, vol. 24, no. 7, 2022.

[48] Q. Li, L. Cheng, W. Wang, X. Li, S. Li, and P. Zhu, “Influence maximization through exploring structural information,” *Applied Mathematics and Computation*, vol. 442, p. 127721, 2023.

[49] A. Rezvanian, S. M. Vahidipour, and M. R. Meybodi, “A new stochastic diffusion model for influence maximization in social networks,” *Scientific Reports*, vol. 13, no. 1, p. 6122, 2023.

[50] H. E. Hurst, “Long-term storage capacity of reservoirs,” *Transactions of the American Society of Civil Engineers*, vol. 116, no. 1, pp. 770–799, 1951.

[51] B. Qian and K. Rasheed, “Hurst exponent and financial market predictability,” in *Proceedings of the Second IASTED International Conference on Financial Engineering and Applications (FEA 2004)*, 2004, pp. 203–209.

[52] C.-K. Peng, S. Buldyrev, S. Havlin, M. Simons, H. Stanley, and A. Goldberger, “Mosaic organization of DNA nucleotides,” *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics*, vol. 49, no. 2, pp. 1685–1689, 2 1994.

[53] M. Resta, “Hurst exponent and its applications in time-series analysis,” *Recent Patents on Computer Science*, vol. 5, no. 3, pp. 211–219, 2012.

[54] R. Bryce and K. Sprague, “Revisiting detrended fluctuation analysis,” *Scientific reports*, vol. 2, no. 1, pp. 1–6, 2012.

[55] T. Kleinow, “Testing continuous time models in financial markets,” Ph.D. dissertation, 7 2002.

[56] W. Chen, Y. Wang, and S. Wang, “Efficient influence maximization in social networks,” in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD’09)*, 06 2009, pp. 199–208.

[57] D. Liu, Y. Jing, J. Zhao, W. Wang, and G. Song, “A fast and efficient algorithm for mining top-k nodes in complex networks,” in *Scientific Reports*, vol. 7, 02 2017.

[58] K. Sheng and Z. Zhang, “Research on the influence maximization based on community detection,” in *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 05 2018, pp. 2797–2801.

[59] S. Agarwal and S. Mehta, “Social influence maximization using genetic algorithm with dynamic probabilities,” in *2018 Eleventh International Conference on Contemporary Computing (IC3)*, 08 2018, pp. 1–6.

[60] M. Adineh and M. N. Baygi, “Maximum degree based heuristics for influence maximization,” in *2018 8th International Conference on Computer and Knowledge Engineering (ICCKE)*, 10 2018, pp. 256–261.

[61] D. Huang and L. Pan, “An improved group-based influence maximization method in social networks,” in *2018 Third International Conference on Security of Smart Cities, Industrial Control System and Communications (SSIC)*, 10 2018, pp. 1–6.

[62] G. Xie, Y. Chen, H. Zhang, and Y. Liu, “MBIC: A novel influence propagation model for membership-based influence maximization in social networks,” *IEEE Access*, vol. 7, pp. 75 696–75 707, 2019.

[63] L. Qiu, W. Jia, J. Yu, X. Fan, and W. Gao, “PHG: A three-phase algorithm for influence maximization based on community structure,” *IEEE Access*, vol. 7, pp. 62 511–62 522, 2019.