A Comprehensive Insight towards Research Direction in Information Propagation

Selva Kumar S
Assistant Professor
Department of Computer Science & Engg.
BMS College of Engineering
Bangalore, India

Dr. Kayarvizhy N
Associate Professor
Department of Computer Science & Engg.
BMS College of Engineering
Bangalore, India

Abstract—The concept of Information Propagation has been studied to illustrate the particular, discrete, and explicit behavior of the nodes in a complex and highly distributed and connected networks. The complex network structure exhibits various challenges towards information propagation due to the usage of diversified communication protocol and dynamic behavior in the context of uncertainty. This paper is first one of its kind, which reviews frequently addressed problems, the most significant research techniques, for addressing various research problems associated with the information propagation concerning to social network analysis, data routing behavior in the multi-path wireless networks, multimedia transmission, and security. This paper is useful for the various researchers, academicians, and industry having research interest into social network analysis, predictive modeling and information propagation analysis.

Keywords—Information Propagation Prediction; Social Network Analysis; Predictive Modelling; Information Propagation

I. INTRODUCTION

With the advent of modernization in telecommunication and pervasive computing, there has been an enormous growth of data generation [1] [2]. Information is being generated using both voluntary and involuntary process. Voluntary process of information generation will be that generated by a particular system or process or an individual, e.g. credit/deposit update of user’s account in a bank, call logs generated by making calls, chat script of two individual over a network etc. Involuntary data is being generated by sensor or any particular data capturing unit. Normally, the process of information generation using involuntary process uses a particular schedule or duration during which the information is being recorded after being generated. However, for the voluntary process of information generation, there is no such schedule and the process could be extremely dynamic in nature [3] [4]. Hence, maximum complication and research problems could definitely arrive in the process of voluntary information generation. The challenges are i) predicting the time and node for information generation is not possible, ii) there exist uncertainty of applying certain storage models or applying analytical operation over the generated data, iii) it also effects the security features, iv) identifying or analyzing relationship among nodes in the network [5][6]. The existing size of information in the era of cloud is quite high dimensional and thereby performing sub-clustering on such higher dimensional data is still an open research problem. Although it has become two decades of using the internet, it is quite necessary to understand the true picture of information propagation over the complex networking structures [7] [8] [9]. There has been various break-through research works towards mining operation but in reality performing complex analytical operation towards complex data generated by the voluntary process in a social network is still an NP-hard problem [10]. At present, there are a various offering of multiple technologies with different communications protocols, which is striking factor influencing information propagation. The most difficult are to understand the traffic model or the underlying topology that governs information propagation of recent time.

Another challenging issue in the area of the information propagation is to develop a generalized protocol that can perform exploration of the user-defined queries of the complex network. Although, the majority of the information extracted is in the form of the text, but all the information are not in textual format and may possibly differ. This could lead to another impediment towards information processing in complex networks. At present, graph and contents are used for performing analysis of the social network data but for the large and massive size of the data, such techniques are also becoming inapplicable [11][12]. This voluntary process leads to a generation of massive data which is uncertain in nature, is voluminous in size, and highly heterogeneous in its type. Such problem would definitely pose a significant level of impediment towards working on the upcoming complex forms of the network with more pervasiveness.

This research paper discusses the theoretical aspects of information propagation and attempts to highlight the progressive picture of its research work and its possible effectiveness. The information propagation is still in a nascent stage with respect to a voluntary process of data generation which requires further attention. This paper discusses various research problems being addressed by the existing researcher towards the domain of information propagation.

Section II discusses the information propagation followed by existing trends of research work in Section III. Various significant research techniques towards information propagation are discussed in Section IV followed by a discussion of the research gap in Section V. Finally, Section VI summarizes the contribution paper.
II. INFORMATION PROPAGATION

There is a massive and infinite amount of information that flows over the World Wide Web (WWW) almost every second. Information propagation is one of the essential theories that act as a backbone for a maximum number of networks. Hence, exploring such latent links from the generalization links are the agenda of information propagation. The model of information propagation can be represented as the transmission of specific information from one node i to another node j for a given communicating environment and it is usually illustrated using abstract algebra using lattice. It was also observed that the theory of the information propagation is interpreted in the form of security policies in the communication channel. The information propagation represents the mechanism by which a data propagates throughout the networks and system. Maximum of the existing research techniques considers the challenges of enhancing the communication among the nodes and identifying the potential node for a specific job in a very dynamic environment. The message to be propagated would be in the form of a static web-page, a textual message shared in a social network, an address of a particular site, etc. The biggest challenging factor in this regard is to increase the influence of any particular node in a positive sense. The process of increasing the influence in information propagation targets to choose a certain quantity of nodes that has higher influence over other nodes for a given network [16] [17]. The another research challenge in information propagation is information diffusion [18]. These problems equally affect the structure of the link for a given network. However, it doesn’t leverage any effects towards the information that are being exchanged among the nodes, which negates or weakens the present communication channel.

In the conventional network, it has been seen that the information diffusion considers the spreading of the influence among its adjacent nodes and grows exponentially over a period of time. However, as per certain studies, it was seen that probability has no dependencies towards neighborhood size [19]. It is the quantity of the inter-connected components with the nodes that affect the node’s influential factor. The theory of information propagation has introduced certain algorithm e.g. PageRank algorithm [20], which is used at present by Google. Basically, Page Rank algorithm performs search optimization over a given link. It allocates a numerical weight to all the hyperlinked contents over the internet. The work carried out by Sargolzaei et al. [21] has elaborated discussed the problems pertaining to presently used technique to understand the node influence in information propagation. The authors have also discussed certain equivalent techniques related to PageRank and its possible enhancement in future. The implication of information propagation is quite high and is potentially needed in order to extract a lot of information within the given network. The next section will discuss the existing research trends towards information propagation.

III. RESEARCH TRENDS

This section discusses the existing research trends with an aid of certain statistics of research papers published and archived in last six years. It is explored that information propagation has various forms of associated implication towards different forms of research problems.

Fig. 1. Implication of Information Flow

Fig.1. (a) shows that there is a possible network communication existing among all the different types of users. However, just by observing this would mean the existing of links but doesn’t give much discrete information about the type of the links [13]. The patterns of behavior among the communicating nodes are very different in a centralized system as well as in distributed system. To meet the demand of the existing customers, the majority of the enterprises are turning up their application from centralized to distributive so that it can be reached to maximum customers [14]. The presence of such links among the communicating nodes in Fig.1 (a) doesn’t show any form of distinguishing behavior among the nodes. Therefore, information propagation assists in discretizing such forms of the communication links. Fig.1 (b) shows an unambiguous form of the links that were latent in nature. Hence more the dimensionality of data size is more and the probability that such links become more latent in nature [15].

www.ijacsa.thesai.org
It is noticed that there are 9285 conference papers and 3319 research journals published from 2010 to 2016 pertaining to various problems in information propagation. A closer look into this statistics shows that less number of standard implementation work being focused as visible by a lower number of journals as compared to the conference. Apart from this, there are 200 easy access articles, 5 standards, and 2 course related articles were published in this duration. In this, also did in-depth investigation towards understanding these trends to infer the various forms of research work that has been evolved over such period of time.

Here find that there are 71 journals published for social network problems (Fig.3), 97 journals towards vehicular network problems (Fig.5), 19 journals towards addressing multimedia problems (Fig.6), and 124 journals towards security problems (Fig.7). This statistics easily shows that although there are some numbers of studies towards wireless network but there are no journals stating implementation of dynamic web-based applications. From the security viewpoint also, there is quite a less number of implementation as compared to the massive progress being done on security algorithm in recent times. Information flow is normally studied with respect to the social network, but an implication of social network problems are less to be seen in existing research work. A similar trend can be also seen in other reputed publishers e.g. Springer, Elsevier, InderScience etc. By looking into the statistics of existing techniques, it can be inferred that Information Propagation is still theoretical and will need a further amount of investigation to explore better possibilities of enhancing the communication system in every respect. The next section discusses the recent research implementation techniques.

IV. EXISTING RESEARCH IN INFORMATION PROPAGATION

This section discusses the existing research work being carried out in information propagation. There are various forms of studies being carried out in this regard addressing to various problems. The easier way to understand information propagation in real-life is that of a transaction of email communication system. Studies on information propagation towards email-based communication system were carried out by Lejun et al. [22] using mathematical formulations with an aid of network-based attributes e.g. degree, density,
betweeness, clustering coefficient, Eigenvector, etc. Similar attributes were used to find present work is better than existing mechanism. Various conceptualized model has also been presented in order to model in better shape. Usage of survival theory was seen in work done by Rodriguez et al. [23] in order to develop a new multiplicative model of information propagation. Usage of software component model for constructing architecture towards forecasting the reliability of the message flow was seen in the work of Brosh et al. [24]. Consideration of inter-contact time was adopted by Boldrini et al. [25] in order to investigate the instances of interaction among the nodes during information flow. Adoption of the neural network was found to have better classification usage in the proposed work of Latorre et al. [26]. Analysis of Information Propagation in Academic social network is done to identify the trend topics by S S Kumar et al. [27]. The model developed to capture the popularity of the video being propagated in Online Social Networks based on the sharing and viewing by the users of Subhankar et al. [28]. There are various such techniques that have been implemented during 2010-2016. Hence, for better illustration, this section will address the frequent area on which there are a significant amount of research work has been carried out toward information propagation.

A. Studies on Social Network

The information propagation is much studied in the context of a social network. The general concept of information theory finds its direct utility in analyzing social networks. Zhuang and Yagan [29] have investigated the impact of the physical network with an online social network in order to form a multilayered network. A clustering concept was introduced with arbitrary graph and the study outcome was evaluated with respect to epidemic size using varies cases of clustering. A similar idea of using physical and social network was also found to influence the cost of propagation. The concept of multilayer network was also seen in the work of Salehi et al. [30] where a modeling aspect of propagating the process is discussed. Reduction of such cost of propagation was found in the work of Guler et al. [31]. The uniqueness in the study was that the technique can identify the negative nodes existing in the path of information propagation and perform elimination of them. The study outcome was found to significantly minimize propagation complexity compared to existing techniques. Wong et al. [32] have presented a stochastic modeling that performs recommendation in the social network. The mechanism assists in giving recommendation of friend to the user. A unique diffusion model was presented by Niu et al. [33] where an analytical modeling was used to investigate the effect of multiple sources of influencing the node. The design completely uses Expectation Maximization approach in order to enhance the performance of information propagation. Zhang et al. [34] have presented a model that models the interactivity among the contagions. A co-competition modeling is presented in this work. Study towards the significant index of performance for monitoring propagation speed was carried out by Chou et al. [35]. The presented technique is essentially meant to be acted as a reference model in order to extract certain specific information pertaining to social network applications e.g. topological degree, rates of contact, the strength of social dynamics, etc. Cha et al. [36] have introduced a technique of investigating the characteristics of information propagators. Using twitter database, the authors have discussed the possible relationship among. There are various dynamics of the study concerned with the applicability of research work towards social network and each study has different forms of approaches. Also, will further classify the applicability of the existing techniques of information propagation towards various other implications of social networks:

- **Emotion Analysis**: The theory of information propagation also plays a significant when it comes to application performing emotional analysis. Normally, such application uses different forms of dictionaries and lexical database in order to extract certain degree of semantic information from the posts in social network. A unique cascading model is presented by Wang et al. [37] on the basis of emotional analysis in order to analyze the propagation probability of data pertaining to massive number of users. Dragoni et al. [38] have presented a framework for concept polarity as well as its linked uncertainty using fuzzy logic. A knowledge graph is developed using existing lexical resources, where the technique is used for propagating information pertaining to sentiments. Exactly similar approach of using fuzzy logic was also seen in the work carried out by Trung and Jung [39]. This technique performs analysis of sentiments from the social network with an aid of propagation modeling using fuzzy logic. Overgoor and Rosen [40] have also presented a simple modeling of sentiment analysis on the basis of online news in order to extract the information related to popularity.

- **Financial Analysis**: There has been a study related to information propagation also in the applications of finance-based. Normally, such application requires a highly safer and fault tolerant mechanism of propagating information. Toriumi and Komura [41] have presented a framework where an investment index has been designed on the basis of transfer entropy in information propagation. For this purpose, a cross-correlation technique is used to the stock prices and hence it generates a stock network. Financial analysis can be also carried out considering data from the financial news that bears various economics-related technicalities. Chang et al. [42] have extracted a significant feature of semantics from financial news using a neural network. Most recently Kaminski et al. [43] have presented a model that extracts correlation factor between posts of twitter and open source peer-to-peer money. The study basically computes the emotion from the post regarding bitcoin market. Similar work considering bitcoin network was also carried out by Decker and Wattenhofer [44]. Sprenger and Welpe [45] have investigated sentiments associated with stock-related tweets using a computational linguistics-based approach in financial information propagation.

- **Rumor Analysis**: Rumor detection and analysis is another frequently exploit research topic pertaining to social network analysis and information propagation.
One of the bigger challenges in this is to identify the extent of legitimate contents in the social post when the information propagation takes place in highly dynamic manner. Hamidian and Diab [46] have presented a technique to identify rumor from the twitter data using the feature of a latent vector. The authors have also used machine learning for this purpose. Tan et al. [47] have presented a propagation model for rumor where the emphasis is more on understanding its underlying patterns during information propagation. A mathematical modeling-based approach is used for this purpose for developing attributes for node interaction as well as for investigating the dispersion over the feasibility towards multiple states. Study towards rumor detection has been also carried out recently by Liu and Xu [48] where the authors have modeled information propagation in order to differentiate rumor-based messages. Wen et al. [49] have presented an analytical technique that can distinguish between positive and negative forms of information in social networks. Amoozgar and Ramezanian [50] have introduced a computational model for investigation the association between convergence of rumor and societal homogeneity in the process of information propagation. A similar form of study considering twitter data has been discussed by Jin et al. [51]. Krishna Kumar and Geetha Kumari [52] have used cognitive psychology in order to investigate the dispersion of rumor. With an aid of collaborative filtering property, the technique is also able to find the speed of rumor propagation. Luo et al. [53] have presented a technique to explore the source of rumor origination in social networks for a given observations that are quite limited. Wen et al. [54] have implemented a process of clarifying the rumor in order to discourage the spreaders of rumors. A similar cadre of research work towards eliminating the rumor is also carried out Krishnamurthy and Hamdi [55] where the authors have used acyclic graphs. Nel et al. [56] have discussed a model of rumor detection that uses clustering for detecting the similar publishing features.

B. Studies on Wireless Network

The studies on the information propagation are also carried out towards wireless network in different forms. The nature and characteristics of the network are quite essential when it comes to transmitting sensitive information. Most recently, Kim and Peeta [57] have used graph theory to model a communication system among the vehicles. A conceptual model is designed using integration of physical network for traffic flow, network layer, and virtual layer for intercommunication among the vehicles. A simulation-based approach has been used for this purpose. Liu et al. [58] have adopted a probabilistic approach for resisting the relay for message propagation in a mobile adhoc network. The authors have used Markov chain for message characterization and used simulation-based approach for evaluating their idea. Zhang et al. [59] have presented a technique for constructing an array of the virtual antenna using MIMO for performing information propagation. The investigation shows that speed of information propagation increases with increase in the density of the vehicle. Kim et al. [60] have presented a technique for communication among the vehicles in order to assists in the flow model of the traffic. The authors have developed an integrated traffic model using analytical modeling approach. Du et al. [61] have presented an investigation that computes delay of information propagation considered for vehicle-to-vehicle communication. The analysis of the study was carried out on multiple-way segments of the road in order to compute propagation delay on multiple scenarios. Study towards the delay of the information propagation is also carried out by Iyer and Vaze [62]. The author defines the speed of information propagation as a limit of the proportion of the spatial attribute from source to destination with respect to the duration consumed for successful delivery of information using multiple hops. The studies carried out by Wang et al. [63] have investigated on the speed of the information velocity when the data transmission takes place between two moving vehicles. The investigated have also focused on studying the underlying relationship among various traffic-based attributes e.g. direction, velocity, density, etc. Both dense and sparse network has been investigated considering bi-directional highway model using a traffic simulation tool. A similar study of a velocity of information speed during message transmission was studied by Han and Yang [64] considering the case study of cognitive radio networks of multihop type. The technique mainly used the maximum network for increasing the velocity of data transmission using the infinite plane. It also increases the velocity of message forwarding between the transmitter and receiver node. The problems associated with the dissemination of information are addressed by Li et al. [65] in connection with device-to-device communication. A constraint-based approach was formulated for constructing time and distance based limits. Pereira et al. [66] have introduced a diffusion-based data propagation model using the expectation-maximization algorithm. Baccelli et al. [67] have investigated the case study of delay tolerant network in order to further assess the characteristics of speed associated with information propagation among the vehicles. The authors find that speed of information propagation behaves in a different way when compared with certain thresholds. Using simulation-based study on multiple speed of a vehicle, the study characterizes the speed of information propagation. Similar study of speed of information propagation has been carried out by Jacque et al. [68]. Wang et al. [69] have used a specific routing technique called as most forward within radius used in the vehicular adhoc network for assessing the performance of speed of information propagation. The presented study also computes anticipated distance of transmission, the rate of packet loss, coverage etc. Xu et al. [70] have discussed the limits associated with the propagation of information in larger deployment scenario of the wireless network. A presence of higher limits of broadcast and unicast speed is found in presence of network noise. Zhang et al. [71] have presented a model where the speed of the vehicle is mapped with the constructed time-slots in order to derive an analytical modeling of information propagation associated with a mobile adhoc network. A simulation-based approach is used for assessing the presented study with respect to delay/gap as well as the logarithm of information speed using Poisson arrival model.
C. Studies on Multimedia Propagation

There has been a significant amount of research work towards multimedia transmission in last decades [72] [73] [74]. Such forms of research attempts deal with the retention of high data quality at the receiving side, reduces the complexities associated with information propagation. Usually, multimedia contents e.g. image, video, audio etc are quite heavier files when compared with other conventional files during information propagation. Transmitting such forms of heavy data consumes maximum network resources and hence the information propagation is likely to be affected. Existing studies normally uses the performance parameters that centers around the quality of the multimedia contents e.g. peak signal-to-noise ratio, sensitivity, root mean square error, etc. Although, there are many such studies, but will discuss some of them with higher relevancy. An investigation was carried out towards sharing of the video by Li et al. [75] over social networks. The authors have investigated about the pattern of information propagation. Various forms of user’s behavior, structures pertaining to typical propagation, etc were investigated in this study. Yang et al. [76] have presented a unique discussion of the unreliability problems associated with pairwise similarities. The authors have solved this problem using tensor product graph in order to regain the better amount of reliability using superiorly compressed video dataset. Yao et al. [77] discuss the problem associated with the re-ranking algorithm during retrieval process of image and video. The author uses the mutual exchange of the information in order to enhance the performance of visual search operation. Bracamonte and Poblete [78] have addressed the problem associated with image tagging which is one of the frequently used operations in the social network. This operation is mainly used for retrieving significant data. This technique introduces a graph-based mechanism that performs automatic tagging of the contents of multimedia. Apart from the above-mentioned studies, the majority of the studies are related to simple transmission of multimedia over certain forms of networks.

D. Studies on Security

Security is one of the potential demands for any forms of application and it is highly important in propagating information using secured channel. In past, there has been various significant studied being carried out towards evolving security systems using sophisticated cryptographic and encryption approaches [79]. Usage of public key cryptography is frequently used in wireless networks, which requires the secret key or certain level of passphrases to be propagated in a specific manner. Hence in this regard, this section will discuss some of the unique research attempts towards security pertaining to information propagation. Most recently, Haghighi et al. [80] have developed a novel model that performs the analytical operation in order to understand the dynamics related to data transmission in wireless sensor network. The technique uses both spatial as well as a geometrical constraint in order to have a better mapping of malicious objects e.g. worms in a sensor network. Gervais et al. [81] have investigated about proof of work blockchain associated with the usage of bitcoin and presented a quantitative modeling of assessing its security effectiveness. A new security model has been designed that using mining-based attributes for assessing the ability of secured information propagation. There is also a research work being carried out based on the transmission of malicious codes through short messaging services in a cellular network by Yun et al. [82]. A stochastic process is presented using analytical modeling using reputation, trust, and user-relationship in order to infer the malicious behavior in information propagation [83]. Patil et al. [84] and Teng et al. [85] have discussed a possible relationship between location and privacy. The authors have presented a collaborative modeling in order to protect privacy during the information propagation considering Bayesian attacks. The prime motive is to ensure location privacy during information propagation.

E. Taxonomies of Existing Problems

Fig. 8 shows the taxonomies of the various research problems addressed in the most recent times pertaining to the information propagation. A closer look into all these problems will show a common link i.e. the social network that offers a comprehensive platform for sophisticated exchanging of the significant information among the users in the most dynamic environment. The majority of the techniques offered using the backbone of the social network has also offered a potential structure representing the interactions among the users over network along with consideration of various network-based dynamics. The various phases of the content generation is the process of information propagation cannot be denied and it significantly affects the processing, production followed by distribution of the data from one to another end of the network. There is already various forms of the applications that area affected by such new advent of technologies e.g. online data sharing process among the communities, video-based information propagation, various dynamic services of multiple remote desktop connectivity, various virtualized-based applications over the cloud, patients details sharing using specific healthcare management etc. Hence, there are various applications that have been evolved due to the process of information sharing in most recent time. The next section will elaborate on the open-end research problems.
V. RESEARCH GAP

This section discusses the potential research gap explored in the area of information propagation by referring the discussion on the prior section. As it has been observed that:

- **Scattered Researched Form**: The concept of information propagation was not much focused on any particular problems. It can be seen that there are large number of studies being focused on adhoc network e.g. vehicle-to-vehicle, device-to-device, mobile adhoc network etc. Although, there are also studies carried out in other forms of wireless networks, but still the inclination of the problem is more on social network analysis and not on networking much. This creates an uneven balance in the research attempts for addressing the problems in information flow. Moreover, there are quite a few research implementations towards dynamic web-based networking system that may cause a serious impediment towards upcoming pervasive network. Moreover, from the viewpoint of problems associated with the social network, the existing studies towards information propagation doesn’t deal with its inherent problems (e.g. computing the influential node, optimizing performance of social network) etc. Therefore, on that context, it can be said that there is less work towards social network problems using information propagation in recent times.

- **Studies towards Computing Potential Node**: Basically, the structure of the social network is highly dynamic in nature and finding one node or set of some nodes based on certain user-defined criterion is basically an NP-hard problem in information propagation. This pilot study toward existing approaches of exploring potential node in dynamic network finds that majority of the research techniques considers that significance value increases with maximization of degree of vertices. This leads to significance performance declination in large heterogeneous scale network. The existing approaches have also found usage of unconstrained and bound-constrained optimization problems towards exploring potential nodes in social network. However, such approaches are never stable as their fitness functions are more dependent on anticipated value of information diffusion. Finally, there is not much work towards aiming for reducing computational complexities in this regard with less evidence of any benchmarked model or research work till date.

- **Studies toward Optimization**: The problem considered in this part of the study is basically related to the enhancing the information dissemination performance in any dynamic networks. In an existing mechanism, various forms of business strategies are formed using ranking algorithms (e.g. Page Rank, etc). However, the biggest problems in using such applications are that they don’t offer contextual information that is correlated with specific business need. This is basically because increasing the level of information diffusion is again an NP-hard problem on any dynamic network and the dissemination process can be hardly controlled or fine-tuned according to the business needs. Also, such problems are yet to receive attention and are worth to be considered as a novel optimization problem in information propagation.

VI. CONCLUSION

In the process of information propagation, the processes as well as all the communication factors are equally responsible for a node to be affected by its influential properties. However, normally presence of such massive and complex data gives rise to high-dimensional data where it is quite a challenging task to extract a dynamic pattern from a given set of multi-dimensional data during information propagation. The manuscript also discusses the various scenarios where information propagation is studied intensively by various researchers till date on the social network, wireless networks, multimedia transmission, and security. It is also seen that majority of the work carried out towards information propagation was on the social network and in wireless networks. The existing studies towards solving issues of the social network are also seen mostly focusing on rumor propagation, emotional analysis, and financial analysis. In reality, there is quite a less significant work carried out to extract the most potential node in a complex network. It was also studied that research techniques towards optimization are few to find in the line of information propagation. Therefore, future work will be carried out in the direction to address such problem.

ACKNOWLEDGMENT

The work reported in this paper is supported by the college through the TECHNICAL EDUCATION QUALITY IMPROVEMENT PROGRAMME [TEQIP-II] of the MHRD, Government of India.

REFERENCES


[52] KP Krishna Kumar* and G Geethakumari, “Detecting misinformation in online social networks using cognitive psychology”, Human-centric Computing and Information Sciences 2014, 4:14
