DOI : https://doi.org/10.56452/7-4-240

ISSN: 0974-5823 Vol. 7 No. 4 April, 2022

International Journal of Mechanical Engineering

DESIGN AND ANALYSIS OF COMMUNITY DETECTION SCHEME BASED ON MACHINE LEARNING STRATEGIES

Lakshita Sejwal

Asst. Professor, Department of Computer Science, Graphic Era Hill University, Dehradun, Uttarakhand India 248002

Abstract—

Now a day's communication becomes so faster and easier due to regular updation in technology and network growth. It is also become very good medium to publicize the any product means any commercial or noncommercial activity to their actual customers, with low cost. Basically communication means to connect with the people of same interest. Social networking apps and websites make this communication out of restrictions. Main challenge in communication through social networking is the community detection. Community is a group of application/people who finds path for their improvement or to avail each other's services. In this article we shows a community detection approach by using deep learning method named as Deep Learning Enabled Community Detection Scheme (DLeCDS). It works on the dynamic dataset selection module, which is outcome based and gives better accuracy then other approaches.

Keywords: Community detection, machine learning, accuracy,communication

1. INTRODUCTION

In modern era, each and every individual are in Social network as well as having a great exposure the field of information technology and its associated developments. This leads a socialnetwork to great pathway with huge development as well as attracts many Government and Non-Government organizations towards it. Generally, Social Network is considered to be the set of people tied up together over different places and share their ideologies in global manner with respect to privacy and access control norms followed by the respective social network medium. In this social network environment, group of peoples are tied up and generally these peoples are called as nodes. The social network speaks to a social structure comprising of a group of peoples or associations that interface with at least one explicit sorts of conditions, for example, family- members, companions, trading associations, business exchanges, thoughts and so forth[1,2,3]. The general way of analyzing social network is finished up with the assistance of charts, so, that the binding of such communities over the social network to be the form of precision vectors. In a simple way, the social network is the group of connected nodes, which can communicate, share and exchange their ideologies and taught over the public portal. These kinds of public medium considerations are bounded with vertices and edges[4,5]. The term called vertices indicates people/elements and the term

called edges represents the relations between the vertices. A wide assortment of edges can exist between the vertices as well as the preferences, decisions and inclinations of people in the communities differs and it lead to the creation of groups/clusters or different categories of networks over a social community[6,7]

Recognition of such social communities can be utilized for some applications, for example, findingan integrated analysis with some clustered networks, finding a group of unmistakable clients for advertising and suggestions or finding energy efficient networks in social network environment onof the critical difficulties in the analysis of social network is the systematic disclosure of communities as well as the communities are viewed as gatherings, bunches, reasonable sub-groups or applications in various areas; community identification over an social network environments distinguishing sets of peoples/nodes, so that the associations of such groups inside a set are more than their association with other connected community nodes. Most importantly, the ideology of this community detection scheme is to discover a group G= {g1, g2, g3, ..., gN} of thenodes in schematic view, in which the G represents the group of edges associated with the social network and the g1 to gN indicates the vertices with connected decisions. This paper introduces anew approach to identify the communities over the social network medium and in which it can be grouped into different significant classifications including community detection accuracy, reduced time consumption for processing of training and testing data, reduced error rate during processingand the designed approach fault level analysis during training and testing phase. All these parameters and metrics are considered to be the processing measure and the designed approach assured to provides the best scenario of outcome in resulting section. The designed approach calledDeep Learning Enabled Community Detection Scheme (DLeCDS), in which it detect the community over the social network environment based on dynamic dataset selection and processing strategies with respect to machine learning strategies. Usually the processing time of any artificial intelligence based procedures is high as well as the complexity need to deal with suchprocess are also maximum. In this paper, a new novel approaches are introduced to eliminate thoseissues and provides a good accuracy level with proper time estimation norms, in which it is achieved by means of Fuzzy Time Estimation Model (FTEM). The detection of communities oversocial network including several processing stages to attain the high level of accuracy with respectto classification attributes. The stages need to process the dataset are as follows: Data Analysis, Stopword Estimation and Removal and Classification, in which all these stages of processing willbe clearly demonstrated over the following section. In general, all the community detection procedures for identifying the communities over the social network are to identify the specific communities based on the training model. But these traditional methodologies focus only for single community detection strategy with classical algorithms.

The designed model is intended to design a new approach called DLeCDS in association with Fuzzy Time Estimation Model to provide the dynamic dataset selection and processing model with best time perceptions to identify the community over social network medium in an efficient way with respect to the mentioned processing stages. The classical models are used to estimate the communities in the social network for advertising category election and promotional uses alone, but now-a-days many social activities are conducted based on such innovative designed workflow model. The major concentration of such model is time consumption estimation and the performance measures, in which the designed workflow will eliminate all the issues present over the past approach and provide an efficient solution for such issues.

2. RELATED STUDY

Ramalakshmi. R'et al., 2020, designed a paper related to hybrid machine learning model to identify the communities over social network platform with respect to several analysis and estimation strategies. In this paper the authors applied a hybrid approach of Social-Network-Analysis model, in which the community detection accuracy is good enough on this work. This paper is based on accumulation of twitter dataset with real time attributes and values presented on it, in which the illustration of dataset is also clearly mentioned over the paper. The process of overlapping communities estimation and detection principles are also mentioned over the approach in clear manner. The estimation of such communities are identified based on several unique parameters such as user-tag, in which it is shared by the respective individual to group or connectedpeople as well. The major way of identifying such communities are analyzed by means of posts, in which it is considered as a major medium to evaluate the group of people who are all have the same taughts and opinions. In this paper (Ramalakshmi.R'et al., 2020), the structural analysis are conducted by means of Particle Swarm Optimization principles as well as the Naive Bayes-FFA analyzes the link which is having low impact and the covariant structure for detecting communities. The authors mentioned the designed model achieved 93% accuracy in results and the resulting section guarantees those points in clear manner. As well as the fuzzy rules are associated over the designed approach to improve the accuracy levels in good manner with proper time complexities.

The major advantage of this paper is to anlayze the community detection principles based on two intelligent machine learning algorithms such as Naive Bayes based FFA and the Fuzzy associationswith PSO, in which these algorithms are working together in hybrid manner and provides 93% accuracy in results. The limitation identified from this paper is to the concentration of limited system information metrics concentrated on the designed approach instead of analyzing multiple metrics, in which these things are mentioned clearly over the paper future scope area.

VictorChang., 2018, designed a paper related to Big Data associated Social Network manipulations with respect to six most important metrics. This paper identifies six different social network hanlding issues as well as provides the detailed solution for those issues in clear manner. A Social-Network based Application Programming Interface (API) is designed over the paper and manipulate the social network information from the most important social media called FaceBook. The designed paper outcome logics and the result estimations are cross checked with six variant methdologies and the resulting summaries are given on the paper in clear manner. The resilient abilities of the designed Social-Network based Application Programming Interface are estimated clearly with proper analysis and the tested results are given in clear manner over the designed paper. The designed paper research parameters are accumulated based on Big Data logics and the social network analysis are compared with classical processing logics and provide the accuracy levels accordingly. The designed approach guarantees the economic savings over the practical real time implementations and the cost reduction will provides multiple client associations in the social network community.

The major advantage of this paper is it provides the cost efficient Big Data associated social network analysis with high end processing strategies. The designed work of this paper cross-check the results with several classic approaches and provides the estimated accuracy range and the robust resulting level in outcome. The limitation identified from this paper is the time complexities, in which the overall processing time is little more compare to the classical algorithms due to the processing of overall data with designed logics, so that the future work of this approach need to concentrate more on it.

Bingsheng Liu'et al., 2019, designed a paper related to social network information analysis with respect to decision making principles for huge scale network with trustworthy procedures. This paper (Bingsheng Liu'et al., 2019) designed a trustworthy data analysis model for processing the huge data on social network medium based on the designed methodology called Trust- Relationship based Conflict-Detection and Elimination-Decision Making model, in which it is applicable for the large scale connected community people with big data to process and attain thedecision with high precision. In this paper, the authors utilize social networking data estimations and non linear optimization logics to recognize and remove clashes among respective individuals. By finding the ideal answer for the designed non linear optimization logic, an advanced evaluations need to be constructed to evaluate the huge data from the decision makers, who shows the most extensive level of contention simultaneously, just as ensuring that an adequate decrease of the gathering strife degree is accomplished. In the third and last cycle, another decision making principle is designed for Large scale Group-Decision-Making that decides DM's capacity dependent on their contention degree. A numeric model is shown and a reasonable situation is executed to show the possibility of the designed Trust Relationship-based Conflict Detection and Elimination decision making model.

The major advantage of this paper is it provides the multi-disciplinary and efficient social networkdata processing model and the designed data processing logics are time efficient and more accurate in results, in which the resulting portion proved these points clearly. The limitation identified from this paper is the group propagation estimation accuracy, in which the designed model handles more than lakhs of people in a group, so that the timing and processing accuracy isprobable, so the future scope of the paper needs to concentrate more on it.

3. METHODOLOGIES

This paper shows an approach of enhancing the social network community detection model with respect to dynamic dataset selection and processing principles. The designed model introduced two efficient social network analysis algorithms called Deep Learning Enabled Community Detection Scheme (DLeCDS) and Fuzzy Time Estimation Model (FTEM). In which the designed DLeCDS algorithm analyze the social network with dynamic data model, that illustrates the respective user can get any social community from the model with respect to the selected dataset for processing. For example, the user selects the data model for Covid-19, in whichthe attribute and the required parameter specifications of the selected dataset is described to the user in detail with system specification norms. Once the user modify or apply the testing data according to the given instructions provided by the designed model will results accurately withoutany intervention.

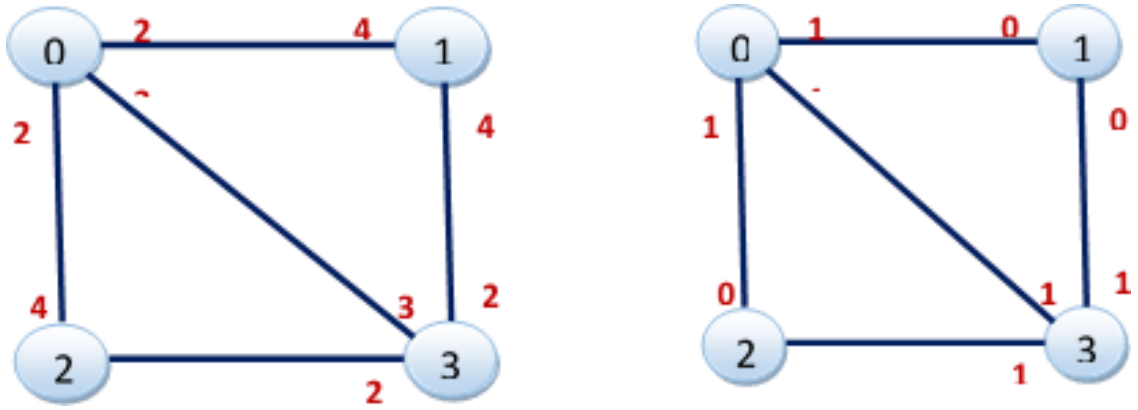
The designed model of DLeCDS concerns more on time estimations, in which the dynamic data selection and processing model cannot specify the exact time constrain for processing the information available into the social network, so that a proper time analysis process is required to process such things in an efficient manner. The designed approach of DLeCDS adapts Fuzzy TimeEstimation Model (FTEM) to process the time for manipulating the overall dataset and provides the resulting accuracy properly without any time delay. The process of fuzzy and its associations related to the paper is described in detail over the following summary. The estimation of the designed approach is analyzed in terms of edges and vertices, in which the individuals or any entities are represented as vertices and the edges are represented by means of connections betweenthose vertices. The general norms of connected community procedures are analyzed by means of edge and vertices estimation principles. The

DOI : https://doi.org/10.56452/7-4-240

general strength estimation strategy for community detection over the social network is defined based on the following metric and the syntatic representation is given as

Str(E,V)= (Energy(e,v) - T(e,v)) **(1)**

Where the Str(E,V) represents the strength level of the estimated edges and vertices and the T represents the time duration for processing such metrics. The general metric of notation associatedwith vertices are (V) and the edges are (E). The following figure, Fig-1 illustrates the energy level associations of the designed community detection principles with respect to vertices and edges. In which the figure shows the principle strength ratio and associated weights of the designed model.



(a) (b)

Fig.1 (a) Strength and (b) Weight Value estimations of Edges and Vertices

A. Data Analysis

The designed model of DLeCDS utilizes the dynamic dataset selection and processing principle, in which the data analysis process purely depends on the type of dataset selected and the dynamicity provides various features associated in it. The designed model allows the user to input the dynamic dataset according to their requirements based on the defined principles. The data analysis part concentrates on given input data and analyzes the structure associated with it and theraw content extraction as well as process the content presented into it. The general concern of dataset and the associated bindings are Covid-19 dataset, in which this is specifically useful for identifying the community of people bounded in social network in covid-19 affection cases. This particular dataset is accumulated from World Health Organization's official site and the case mentioning on the dataset are 76 lakhs and above as well as the death ratios are marked as 17 lakhs and above. The next category of dataset specification is Blood Donor dataset, in which it can be extracted from datacamp.comn site and the dataset provides around 748 sample donor details with respect to prediction laws. Similarly the user can input n number of datasets according to their convenience and train a model to identify the particular community from the social network medium. The following table, Table-1 illustrates the required dataset parameters and the value ranges in detail.

DOI : https://doi.org/10.56452/7-4-240

Table-1 Customized Dataset Parameters

Parameters	**Range**
Minimum Number of Columns	3
Maximum Number of Columns	10
Minimum number of Data Rows	100
Maximum number of Data Rows	1Lakh
Weight Factor	1-4
Minimum Prediction Label Column	1
Maximum Prediction Label Column	3
Minimum number of Attribute(s) per Column	1
Maximum number of Attributes perColumn	5

B. Stopword Estimation and Removal

The designed algorithm called DLeCDS associates the Stopword Estimation and Removal function to maintain the designed model robustness and assures the resulting accuracy with best time interval. The Stopword analysis and removal process identifies the unwanted stop words in the given phrase or dataset content and removes the respective words from the content and provides the novel data for processing. The purpose of associating this function into the designed logic will enhance the possibilities to attain better resulting accuracy level and outcome retrieval process willget improve based on this scheme.

C. Classification

All the machine learning principles are highly confident in its classification schema, but the extraction of data content and the training phases gives the confident threshold level to such classification algorithms to attain the maximum accuracy level. This paper shows a machine learning algorithm called Deep Learning Enabled Community Detection Scheme (DLeCDS), in which it accumulates the processed dataset for processing as well as the user can provide dynamic dataset with respect to the parameter specification ranges. This kind of dynamic dataset processing scheme consumes more time to analyze and manipulate the given dataset, but the designed model utilizes the power of fuzzy associations with this approach of DLeCDS and provide a new algorithm called Fuzzy Time Estimation Model (FTEM). In which the FTEM provides accurate time estimations with respect to probabilistic principles. The fuzzy logic usually follows the MinMax logic, in which the output assures the maximized output gain with minimized loss ratio. This concept of fuzzy is attained over the designed approach of DLeCDS to attain highaccuracy with estimated time limit without any processing delay. With the help of this designed approach, user can dynamically select any dataset and test the data accordingly to attain the properdetected communities respective to their needs.

The algorithm efficiency is estimated based on the factors followed such as speed and performance. In this concern, the characteristic of speed is estimated by means of noise ratio

detection and delay analysis norms as well as the measures can be processed by means of estimating the following parameters such as weight and strength, group size, maximum edge ratio and the minimum vertices range and finally the taught to organize the community, in which all these mentioned parameters are combine worked together to estimate the algorithm efficiency by means of the following equation,

$$\mathbf{Time\_Spe}(\alpha) \rightarrow \mathbf{G(n)Q} = \mathbf{2(Vmin, Emax)} + \frac{1}{} (\textit{Stren}(\alpha) \pm \textit{Weight}(\mathrm{Q})) \quad \mathbf{(3)2}$$

Where the sign (α) indicates the maximum time and speed estimation, G(n) indicates the group of people over the social network community with respect to specific characteristic (β) and the V and E indicates the edges and vertices of the community as well as the strength and weight measures are properly indicated with α and β signs.

4. RESULTS

In this paper, a machine learning approach called Deep Learning Enabled Community Detection Scheme (DLeCDS) is described, in which it process the dataset dynamically according to the specific parameters and provide the high accuracy ranges on results. The outcome and its accuracy ranges are given in the figure-2. The designed system DLeCDS time efficiency is taken care by the new fuzzy based methodology called Fuzzy Time Estimation Model (FTEM), in which it estimates the processing time and frequency based on the quantity of data to be processed. This FTEM methodology associates the MinMax approach to provide the best probabilistic results, by means of reducing the fault ratio and improving the success ratio over an outcome as shown in figure-3. The designed approach fault estimations are clearly analyzed by means of identifying the bit error rate as well as the designed approach of DLeCDS provides least bit error rate during processing of heavy data; shown in figure-4.

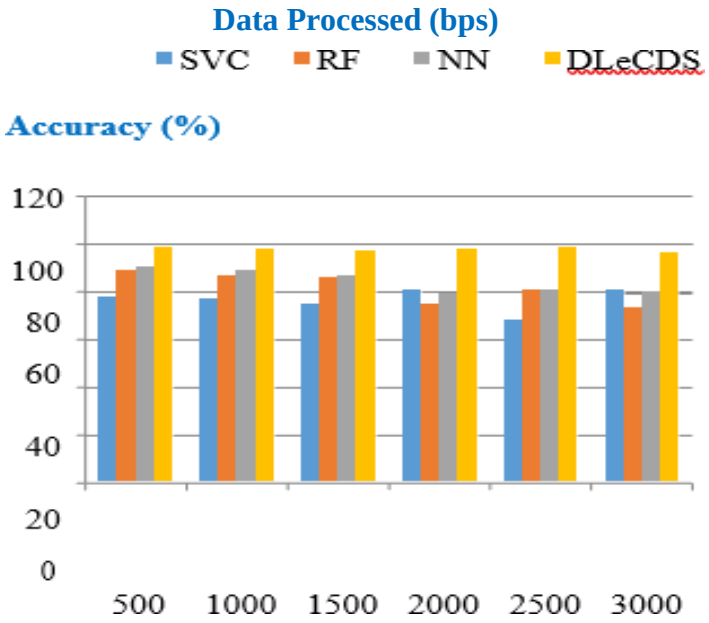


Fig.2 Analysis of Community Detection Accuracy

DOI : https://doi.org/10.56452/7-4-240

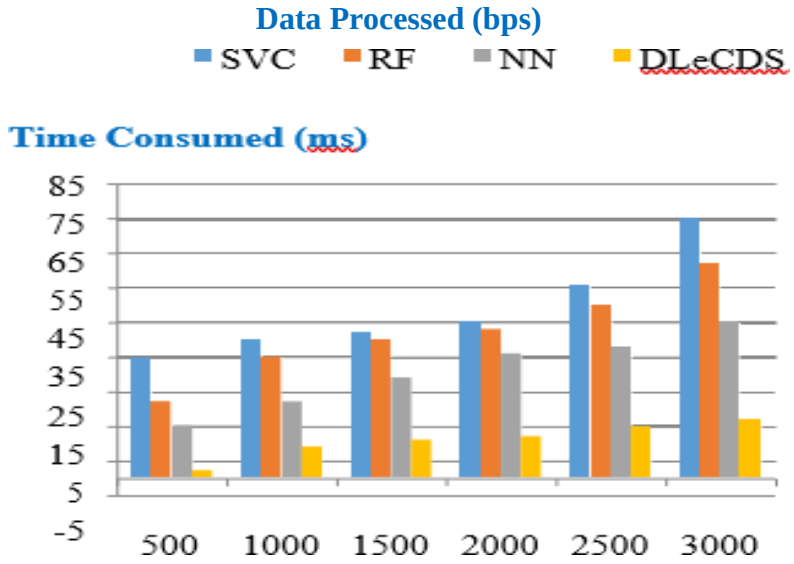


Fig.3 Analysis of Time Consumed for processing the data to identify the community over social network platform

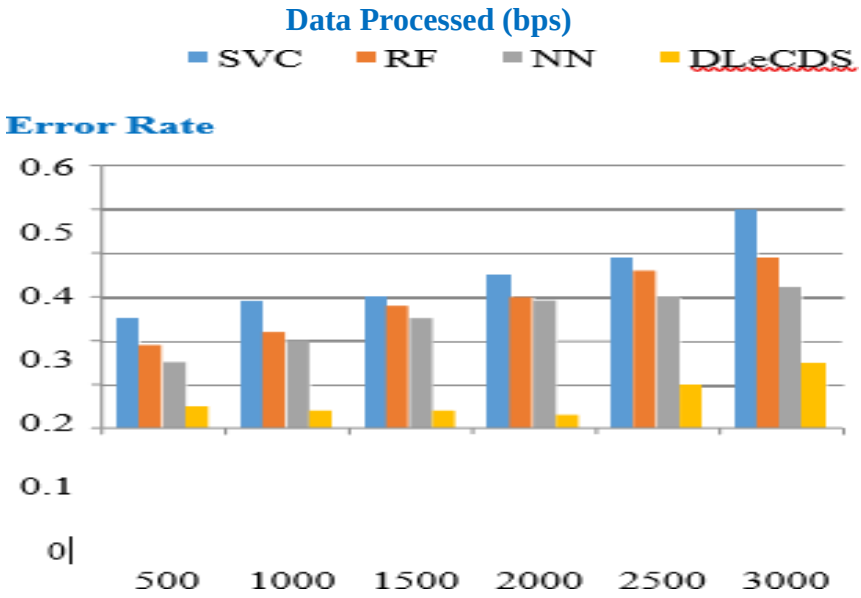


Fig.4 Error Rate Estimation of the designed model DLeCDS

CONCLUSION

the analyzed and designed model is good enough for identifying the required communities from the social network environment without any time issues and intervention.

REFERENCES

1. Bingsheng Liu, Qi Zhou, Ru-Xi Ding, Iván Palomares, Francisco Herrera, “Large-scale groupdecision making model based on social network analysis: trust relationship-based conflict detection and elimination”, European Journal of Operational Research, Elsevier, Vol. 275, Issue. 2, pp. 737-754, 2019.

2. Jooho Kim and Makarand Hastak, “Social network analysis: Characteristics of online social networks after a disaster”, International Journal of Information Management, Elsevier, Vol. 38, pp. 86-96, 2018.
3. Mohamed Moussaoui, Montaceur Zaghdoud, and Jalel Akaichi, “Clustering Social Network Profiles Using Possibilistic C-means Algorithm”, International conference on IntelligentInteractive Multimedia Systems and Services, Springer, pp. 419-428, 2018.
4. Deverajan, G. G., Muthukumaran, V., Hsu, C. H., Karuppiah, M., Chung, Y. C., & Chen, Y. H. (2021). Public key encryption with equality test for Industrial Internet of Things system in cloud computing. Transactions on Emerging Telecommunications Technologies, e4202.
5. Muthukumaran, V., Manimozhi, I., PV, P. S., Karthikeyan, T., & Gopu, M. (2021). Public KeyEncryption With Equality Test for Industrial Internet of Things Based on Near- Ring. International Journal of e-Collaboration (IJeC), 17(3), 25-45.
6. Muthukumaran, V., Joseph, R. B., MUNIRATHNAM, M., Kumar, S. S., & Kumta, S. D. (2021). Efficient Proxy Re-encryption for Internet of Things based on Seminear-ring. Design Engineering, 905-914.
7. Nagarajan, S. M., Muthukumaran, V., Murugesan, R., Joseph, R. B., & Munirathanam, M. (2021). Feature selection model for healthcare analysis and classification using classifier ensemble technique. International Journal of System Assurance Engineering and Management,1-12.
8. Louise Ryan, Alessio D’Angelo, “Changing times: Migrants’ social network analysis and thechallenges of longitudinal research”, Social networks, Elsevier, Vol. 53, pp. 148-458, 2018.
9. A. Singh, S. Garg, S. Batra et al., “Probabilistic data structure-based community detection and storage scheme in online social networks”, Future Generation Computer Systems, Elsevier, Vol. 94, pp. 173-184, 2019.
10. F. Strozzi, E. Garagiola and P. Trucco, “Analysing the attractiveness, availability and accessibility of healthcare providers via social network analysis (SNA)”, Decision Support Systems, Elsevier, Vol. 120, pp. 25-37, 2019.
11. Muthukumaran, V., & Ezhilmaran, D. (2020). A Cloud-Assisted Proxy Re-Encryption Scheme for Efficient Data Sharing Across IoT Systems. International Journal of Information Technology and Web Engineering (IJITWE), 15(4), 18-36.
12. Gyoung-Ju Lee, Son-Il Pak, Kwang-Nyeong Lee and Sungjo Hong, “Movement-Based Biosecurity Zones for Control of Highly Infectious Animal Diseases: Application of Community Detection Analysis to a Livestock Vehicle Movement Network”, Sustainability, 2019.
13. Vinoth Kumar V, Ramamoorthy S, Dhilip Kumar V, Prabu M, Balajee J.M., “Design and Evaluation of Wi-Fi Offloading Mechanism in Heterogeneous Network”, International Journal of e-Collaboration (IJeC), IGI Global, 17(1), 2021.
14. Divyapushpalakshmi M Ramalakshmi R, "Hybrid machine learning approach for community and overlapping community detection in social network", Emerging Telecommunications Technologies, 2020.
15. Jaimie McGlashan et al., “Social network analysis of stakeholder networks from two community-based obesity prevention interventions”, PLOS one, 2018.

16. Vinoth Kumar, V., Arvind, K.S., Umamaheswaran, S., Suganya, K.S (2019), "Hierarchal TrustCertificate Distribution using Distributed CA in MANET", International Journal of InnovativeTechnology and Exploring Engineering, 8(10), pp. 2521-2524
17. Karthikeyan, T., Sekaran, K., Ranjith, D., Vinoth kumar, V., Balajee, J.M. (2019) "Personalized Content Extraction and Text Classification Using Effective Web Scraping Techniques", International Journal of Web Portals (IJWP), 11(2), pp.41-52
18. Vinoth Kumar V., Ramamoorthy S. (2018) Secure Adhoc On-Demand Multipath Distance Vector Routing in MANET. In: Mandal J., Saha G., Kandar D., Maji A. (eds) Proceedings of the International Conference on Computing and Communication Systems. Lecture Notes in Networks and Systems, Springer, Singapore, 24, pp. 49-63
19. Dhilip Kumar V, Vinoth Kumar V, Kandar D (2018), "Data Transmission Between Dedicated Short-Range Communication and WiMAX for Efficient Vehicular Communication" Journal of Computational and Theoretical Nanoscience, Vol.15, No.8, pp.2649-2654