Journal of Theoretical and Applied Information Technology
15th February 2023. Vol.101. No 3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN EFFICIENT META-HEURISTIC OPTIMIZATION-BASED COMMUNITY CLUSTERING AND PRIVACY PRESERVING FRAMEWORK FOR LARGE OSN DATA

SHAMILA. M[1], G. REKHA[2], K. VINUTHNA[3]

[1]Research Scholar, Department Of CSE, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur, A.P, India

[2]Associate Professor, Department Of CSE, Koneru Lakshmaiah Education Foundation, Aziz Nagar, Hyderabad, Telangana, India

[3]Associate Professor, Department Of CSE, Neil Gogte Institute Of Technology, Uppal, Hyderabad, Telangana. India

[1]shamila.m@gmail.com , [2]gillala.rekha@klh.edu.in, [3]vinuthna.dwh@gmail.com

ABSTRACT

Privacy preserving plays an essential role to protect the patterns in the large online social networking databases. Most of the conventional community clustering and meta-heuristic models use static optimization functions in order to find the relationship among the local and global graph nodes for privacy preserving process. Also, these approaches use limited number of social networking graph nodes for community clustering and privacy preserving process. In this work, a meta-heuristic optimization-based community clustering framework is proposed to optimize the privacy preserving process. In this work, a homomorphic encryption-based privacy preserving approach is implemented to protect key nodes during the community clustering process. In this research, the cluster error rate of the inter and intra variations are minimized to improve the community detection process. The performance of the proposed community detection approach is tested on different online social networking datasets for community detection. Experimental results prove that the proposed meta-heuristic-based community clustering and privacy preserving framework has better efficiency than the conventional meta-heuristic-based privacy preserving methods on different OSN datasets. Finally, the privacy of the different social networking nodes and its properties are preserved on large number of graph nodes.

Keywords: *Online Social Networking, Privacy Preserving, Local Optimization, Global Optimization, Machine Learning*.

1.INTRODUCTION

PPDM is primarily focused on lowering the privacy risk while amending the data so that sensitive information can be protected when performing data mining operations, according to Wang et al. Data mining with privacy is a dual-pronged strategy. First, sensitive information from the original database should be changed or removed, such as the user's ID, name, contact information, and address, to ensure that the recipient cannot compromise the user's privacy. People are now more at risk of losing both their personal and professional information as they become more accustomed to social networking. As the capacity to collect and store data about users has increased, privacy-preserving issues in the field of data mining have become too critical. Due to its ability to mine a large amount of user-related data while safeguarding their sensitive information, PPDM has gained popularity. Facebook, the most popular social networking site at the moment, has 2.5 billion active users per month [1]. Users can create profiles, make connections, and send messages to other users thanks to it. Status updates, photos, links, videos, blogs, and locations are all maintained for the professionals. The users share information with other users using the tools offered by OSN . For a variety of reasons, the information gathered from such platforms is made available to the public. Sensitive information, such as name, mobile number, email address, social security number, credit card information, photos,

and videos, is included in the published data. Therefore, maintaining privacy while disseminating social network datasets is a significant challenge. Additionally, the data posted on social networks is available forever. As a result, it produces a tonne of social data. The current privacy preservation approaches modify social dataset to achieve a specific level of privacy before publishing it. The addition and/or deletion of nodes and/or edges are included in this modification operation. A single operation between two nodes modifies the originality of the data subjectively. The methods for protecting privacy put forth by various researchers have some drawbacks in various areas, including: protecting privacy in time-series social network data publication; major deviation in graph properties; adversary access to background knowledge; and poor graph projection method. Social networking has grown in popularity, but it also exposes users to a number of security risks as they reveal personal information like names, ages, genders, contact information, etc. that is readily available to anyone on the network. As a result, understanding data mining algorithms is necessary to secure this data. In social media, there are primarily three types of privacy violations. One of these is identity disclosure, which refers to the disclosure of a user's identity and relationships with other users over a network [2]. Next, sensitive attribute-disclosure arises whenever an attacker gains access to a user's confidential and sensitive information, followed by sensitive link-disclosure, which happens after an association between two users is revealed. Social media platforms give users the setting they need to share their information on open forums while using view options like public, friends, custom, and only me. The user's privacy settings are frequently required by the social media community as well as by the user in order to anticipate problems and take appropriate action. Most users are unaware of the privacy options and services that social networking sites offer. In order to predict users' levels of privacy risk and give them insight into how their data is used by social networking application providers, an analysis of users' privacy is therefore necessary. One of the most significant sources of information for data analysis and the creation of significant findings is social network data. Despite being useful, it is vulnerable to privacy attacks because it contains sensitive data. Online social network patterns that reveal information are studied by [3]. Information about user names and profile pictures on various social networks, details about hobbies and interests, and a variety of user information available on various social network platforms are among the patterns that have been observed [4]. On the social networking site Facebook, the authors [5] conducted extensive experiments on more than 4000 students (users) from Carnegie Mellon University. Users frequently share personal information that could lead to privacy implications [6] like stalking, re-identification, or the creation of a digital dossier, according to their observations. These days, people [7] share geo-tagged media that includes an image of multiple people (friends) as well as embedded location data. The person who shares the media has control over their privacy, whereas that person's friends may be subject to privacy implications. Particularly, privacy implications appear when the media or data is unrelated to a person or friend and is susceptible to privacy problems. In the context of the current social networks, the authors [8] have examined and discussed privacy implications. A set of data objects is grouped into multiple groups or clusters through the process of clustering, which ensures that objects within a cluster have a high degree of similarity but are highly dissimilar to those in other clusters. There are many applications for clustering in data mining, including biology, security, business intelligence, and web search. Every day, a lot of data is gathered from automatic devices like geospatial, bio-medical, security, marketing, and satellite images. A general example of this is the scientific community, where co-authorship or citation data is processed to create bibliographic networks. In order to derive intriguing patterns and trends about the underlying papers, it is used in conjunction with publications' contents. It is indicated that this analysis may be the first instance from which the criteria from above would apply. Although there is a tonne of information and content online because some documents are archived in networks, Due to the content that is available with these networks, a number of bibliographic networks and document collections are explicitly archived and used in conjunction with fundamental data-centric models. Standard studies on social network analysis, which historically have shown the adoption and growth of the internet or computers, do not focus on online interactions. Social media platforms are incredibly dynamic tools that develop quickly over time as a result of newly added edges that signify emerging forms of social interaction. The social network is a web-based platform for user communication and idea sharing. It is a structure made up of various types of entities, including people, organisations, and groups, along with their connections and associations. Through social networks, a group of people can interact and share interests with one another in many different ways

[9]. In determining the secure-multiparty computations, the distributed PPDM problem closely resembles the field of cryptography. The areas where the fields of PPDM and cryptography intersect are viewed. The functions are calculated over the inputs that are provided to multiple recipients using a broad approach using cryptographic techniques, without sharing inputs among recipients. Additionally, the accuracy of computation for the approaches is more crucial. The probabilistic function's computation is defined in the context of complex problems involving a variety of multiparty inputs. These potent methods abstract the primitives from the computationally demanding data-mining problems. Even though there are generic solutions for multi-party situations, the majority of these techniques are described as the 2-party case. The census data is one significant example of how the government gathers private information about the populace and uses it for economic planning and research to create a tool. Furthermore, no one's personal information should be revealed or retrieved from information that is released. Additionally, companies whose databases are made available to the public on their own websites for the use of their customers shouldn't have the ability to match records. The process of randomization makes it simple to implement during the data collection process because each record is supplemented with noise that is unrelated to the possibility of any other data-records. Due to the difficulty of masking outlier records, this technique is much less effective. Privacy protection is not necessary at the time of data collection, but the technique's accuracy depends on the local record behaviour, which is required. The randomization framework's failure to take into account the likelihood of locating the owner of the available records is yet another flaw. The fundamental technique suggested for group anonymization is the K-Anonymity model. Each individual in the transformed dataset cannot be identified, at least from k-1 other individuals in the dataset, thanks to the anonymity model. The k-anonymous table will therefore contain at least k records with identical values. Utilizing concepts of generalisation and suppression, this is accomplished. The original values are swapped out for generalised values when using the generalisation approach. For a particular interval of values, a general value is used in place of the numeric attribute values. Using the taxonomy tree is another generalisation technique strategy. By substituting the value of the child node with either the root node or the node located along the path, generalisation is achieved. Before sharing the attribute's value for analysis, the suppression method deletes it. These methods are designed to stop information from leaking out while the data mining model is being computed. Restricted access or policies are used in non-cryptographic approaches, while encryption is typically applied using cryptographic techniques. Sensitive data can be encrypted and preserved using cryptography. This has been introduced by the researchers as a well-liked security mechanism for sensitive attributes. A decision tree classification-based cryptographic protocol on horizontally partitioned databases makes the assumption of two data sources and permits each to compute missing values without disclosing any information about the other, preserving complete privacy. The mutual trust of all parties involved in the data mining is a requirement for the robustness of these cryptographic techniques. Participants' actions can be divided into two categories: semi-honest and malicious. While performing the intermediate and final computations, parties engaging in semi-honest behaviour are curious to learn about the private information of other parties, but they never deviate from the protocol. Evil actors work together with others and veer from the rules. Secure Multiparty Computations (SMC) are used by cryptographic techniques to perform distributed data mining operations. Decryption may or may not be used in the encryption techniques used in privacy-preserving data mining to perform computations over encrypted data. Numerous cryptographic protocols, including Elgamal, RSA, Hillcipher, etc., allow the parties to perform any mining operation on their inputs without disclosing the specifics of those inputs [10].

2.RELATED WORKS

The main aim is to distribute the most allocated method for preserving the privacy in data-mining which allows computation of functional amount of statistics for the whole dataset without revealing the privacy of the user's datasets. There are several efficient methods used for PPDM that proposed later with a large-scale study of data-mining in current decades. For ensuring the privacy reservation, most techniques are using some transformation form on the original-data. Here, this case is more important for maintaining the advantage of privacy preservation even after dataset is transformed that made usage for mining. According to [12], a serious privacy issue can arise even with a high-quality k-anonymized dataset because of a lack of key attributes. In particular, the degree of privacy protection is independent of the number of quasi-identifiers in the attribute set. Additionally, the

degree is established using the specific sensitive values connected to the attribute set. This observation is used to process the idea of l-diversity. The authors recommended applying the suggested method to numerous sensitive attributes.

[13] used random graph theory to describe a class of attacks in which the attacker uses data gathered from social networks to identify the structured planted data and looks for links between the target vertices. Neighbourhood attacks diverge from the suggested model, and a deliberate structure set is erected before social network data is anonymized. For future improvements, the non-interactive mechanism guarantees the privacy of logical ideas in social networks. By using description logic and some anonymity metrics, [14] represented the formalism of the underlying knowledge and assessed the risk of confidentiality breaches caused by disclosing social network data. For data from social networks, anonymization algorithms are not provided. The proposed system should be examined for the tabulated data, according to the authors. By arbitrarily adding or removing edges, the perturbation method based on random graph is created in order to anonymize social networks. The model, which is not labelled, is assumed to represent the edges and nodes in the social network. The author also suggests using random perturbation to increase anonymity and reduce information loss. [15] primarily focused on social networks with labelled edges but unlabelled nodes. Some edge types should be concealed because they are more delicate. For the purpose of preventing link re-identification, edge-anonymization methods employ edge clustering and removal. The author offers additional suggestions to boost the efficiency of the suggested technique. The method based on simplified windowing for "encouraging decision trees in Distributed Data-Mining Situations" was discussed by [16]. The specimen that is present in the accessible preparation cases is chosen by incorporating the Windowing technique to support by conventional choice tree calculation. To degrade the situation gracefully, the author suggested setting up the bias experiment on Parallel Counter GPU. [17] dealt with the Kernel K-means model algorithm to analyse enormous datasets by carrying out significant clustering and countless applications. In this method, the three phases of the MapReduce programming model—bit-grid calculation, network-trimming technique, and k-means clustering model—are followed. For the proposed research, the author suggested analysing the static kernel-matrix trimming. By combining neighbour requests for a single-inquiry, the set of questions is asked among his or her conceal. On the basis of the classification's sub-features, the authors have provided suggestions for new concepts. Fuzzy techniques were used by [18] to choose each element's sub-feature and maintain privacy while collecting information from parties taking part in a distributed environment. This proposed work's main objective is to identify the class that can be identified by the choice and privacy of a sub-feature. This work yields better results when choosing the sub-features for various situations. According to [19], developing perturbation maintains the confidentiality of the sensitive records. This method is only applied to protect privacy in cases where the data has already been tampered with. This common approach is used for data distortion and privacy maintenance. The original data cannot be successfully recreated and cannot be used for the vast amount of online data. [20] examined how privacy was preserved when information was shared for this composition on Facebook and other social media platforms. The elders who observed the behaviour are studying the information that was shared about it. The public sharing of information about themselves on their profile pages is chosen by them.[21] examined the emergence of multimedia-oriented mobile social networks to analyse the concerns regarding the privacy of multimedia services (MMSN). The users of this method receive the multimedia services from nearby social networks and online social communities. When information sources are used in this manner, the culture has a significant impact on the decision to make an online purchase. The private matching protocol was created with fine-grained functionality to enable the execution of two users' matching profiles without disclosing any kind of profile-related data. The currently used method known as coarse-grained special matching for PMSN protocols allows for the identification of user differences and supports a number of matching metrics at different privacy levels.

Connecting Heterogeneous Social Networks with Local and Global Consistency (COSNET), a model based on energy, was presented by [35] to address the issue between various networks. In decentralised multi-hop mobile-social networks, [22] developed a design mechanism that prioritises user-submitted profiles when searching for people based on profile-matching. The intended mechanism, which prevents participant profiles and preferences from being disclosed, is called privacy preservation. The attack is also known for enabling the outside enemy to summarise socioeconomic factors like sexual orientation, age, and education in the clients who are watching for the exposed area profiles. For roles like

duration and time of the time triggering and role-activation of other ones, temporal restrictions are imposed. The model is typically related to web-based applications and addresses security requirements in a desirable way. The DAC and MAC models for businesses using web-based applications do not support the evolving security requirements. The RBAC model does not include fine-grained access controls because it is designed for large businesses and does not have any central authority over user roles. To solve this issue, attribute-based encryption techniques were developed. In order to process data provision access based on a set of attributes, [23] proposed the Attribute-based Encryption (ABE) model. This ABE method depends on the encrypted public key, which handles encryption and decryption tasks according to user attributes. Additionally, the attributes that belong to the set of users are where the private keys of users are gathered. The data owner who uses ABE systems and encrypts it imposes specific access policies. The user can decrypt the data by being given a private key when the owner has defined an access policy for them. ABE model applications are what create multi-point-to-point communication systems because they are more common. With its fine-grained provisions, access abstractions are also provided at various levels.

[24] examined the access structure of the attribute based on encryption that is non-monotonic. The main benefit of this suggested model is that it has non-monotonic access, whereas the ABE technique does not. In this work, the terms of an access-formula based on attributes are used to express the users' access policies. When compared to other methods, this technique is performing better. In [25] introduced the secure MA-ABE scheme, which is decentralised. This scheme, which relies on the identity of the fuzzy attribute-based encryption (MA-FIBE) model, offers multiple authorities without a central authority. In this method, the messages are encrypted by the encryptor, and if they contain distributed keys for the specified attributes, the decryptor is used to decrypt them. This technique offers the security proof based on the distributed-key generation protocol, the joint zero-secret sharing protocol, and the standard decisional-bilinear Diffie-Hellman assumption. This method uses two cipher-text and key policies from attribute-based encryption techniques. [26] examined the MA-ABE scheme using the accountability process, which uses this technique to provide wildcards and AND gates. The attribute or user authority in the current MA-ABE technique leaks the decrypted key that is linked to user results in security threats. To address the issue pursued in the current work, the MA-ABE scheme offers user accountability. For the purpose of confirming the users' accountability, the users who reveal the decryption key to others are identified. Users and attribute authority both have less trust assumed in them. The main benefit of this proposed work is that, when compared to other techniques, it reduces overheads. There are numerous methods built on MA-ABE that people have developed to access secured provisions in their own unique ways. Generally speaking, two types of ABE-based techniques are offered: the Key-Policy Dependent Attribute-based Encryption (KA-ABE) model and the Ciphertext Policy-dependent Attribute-based Encryption Technique (CP-ABE). An effective model based on attribute encryption of key policy was proposed by [27]. KP-ABE is attached with the data access policy for users' private keys. The data owner encrypts the data content and specifies a set of attributes in ciphertext. The private key of the user is accessing the system's trusted authority's specified data access policy. The main weakness of the KP-ABE approach that leads to unauthorised access to data is tampering with the private key. Leakage of the secret key leads to threats like insider attacks and compromise key attacks of various security. Due to this flaw, the ciphertext policy was developed using the ABE technique.

PROBLEM STATEMENT:

Multi-party social networking data still face some difficult situations for community detection. It is based on privacy protection strategies that are frequently employed in collaborative filtering systems to protect users' privacy. By using filtering strategies and incorporating high-quality original data, the required level of privacy and accuracy is achieved. The randomization algorithm was chosen to enable the accuracy of the data properties with sufficient precision in which the individual entries are noticeably twisted. The distortion required for privacy protection is determined using the privacy measure. Also, as the size of the noisy node properties increases, it is difficult to find the weighted nodes with highest contextual community clusters .

RESEARCH GAP:

The main research gap identified in the traditional graph based social networking models include:

a) Difficult to find strong node to node relationships among complex networks.

b) Difficult to process a complex network for noise detection.

3.PROPOSED MODEL

In the proposed work, a hybrid meta-heuristic-based privacy preserving approach is implemented on the different online social networking datasets. Initially, OSN datasets are initialized for graph based local and global optimization process. In this work, a new local and global optimization measures are used to find the weak and strong nodes for the community clustering and privacy preserving process. A probabilistic weighted based community clustering approach is used to improve the privacy preserving process on large graph nodes. Finally, k-clusters are detected using the graph clustering approach.

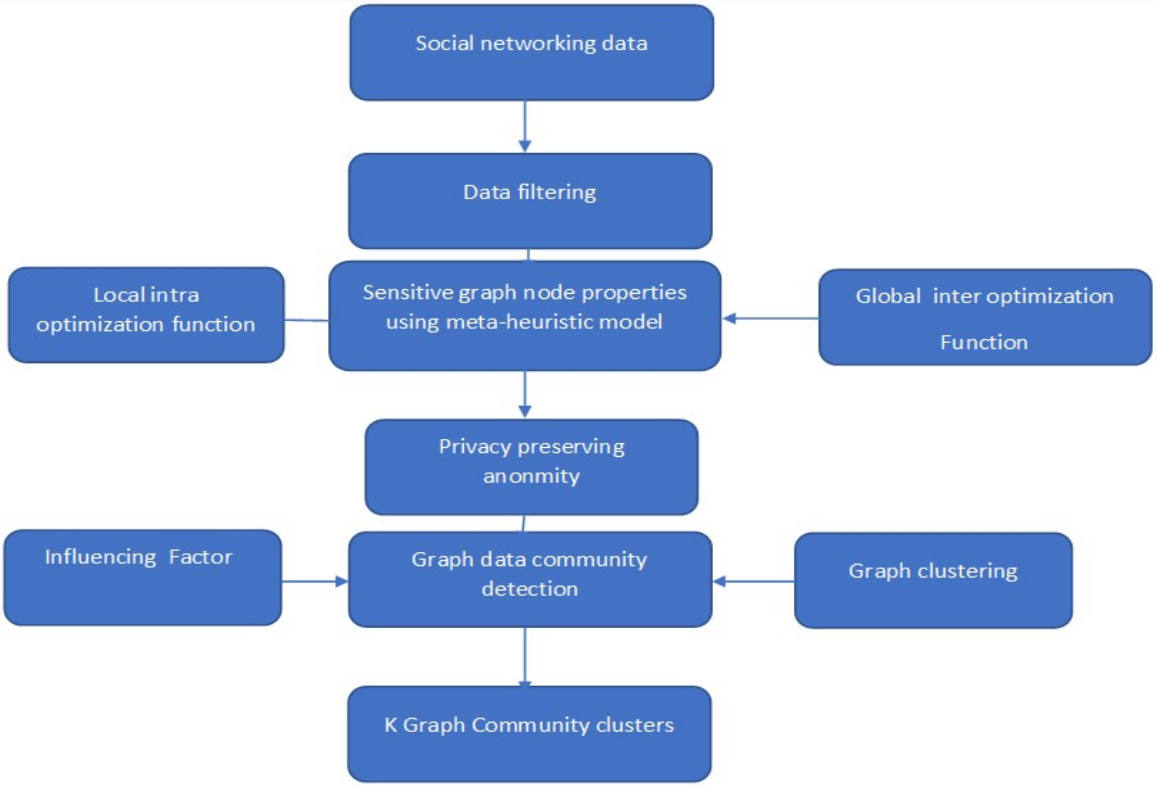

Figure 1: Proposed Framework

Algorithm1: Graph Data filtering

Input Graph data G(V,E)

Output : Filtered Graph Data.

Procedure:

1. Load Graph G(V,E).

2. Compute Adjacency matrix A(G) and Node degree matrix N(G).

3. To each node V(i) in the G.

4. To each node V(j) in the G.

5. if(D(V(i))==0|| D(V(j)==0)

6. then

7. Remove node V(i) or V(j) from Graph G.

end if

8. end loop

In this algorithm, input graph nodes and edges are initialized for the adjacent matrix computation and node degree matrix. In this filtering, isolated or unique nodes are detected and removed from the

graph G. Different node properties are proposed in order to find the weak, neutral and strong nodes.

Graph node properties selection using meta-heuristic model

$$G_{intra-neig}(C_i(G)) := \sum_{j \in C_i} A(i,j); \frac{}{Total \text{ number of intra edges of node i.}}$$

$$G_{inter-neig}(C_i(G)) := \sum_{j \notin C_i} A(i,j); \frac{}{Total \text{ number of inter edges of node i.}}$$

$$Weak_{C_i(G)}(v) = 1; if \exists C_j \in C \; G_{intra-neig}(C_i(G) < G_{intra-neig}(C_j(G) --- (1)$$

$$Weak_{C_i(G)}(v) = 0; else$$

$$Neut_{C_i(G)}(v) = 1; if \exists C_j \in C \; G_{intra-neig}(C_i(G) = G_{intra-neig}(C_j(G) ----(2)$$

$$Neut_{C_i(G)}(v) = 0; else$$

$$Strong_{C_i(G)}(v) = 1; if \exists C_j \in C \; G_{intra-neig}(C_i(G) > G_{intra-neig}(C_j(G) --- (3)$$

$$Strong_{C_i(G)}(v) = 0; else$$

$N(C_i)$: $Cardinality$ of C_i

$Local$ node search based intra-neighbor score is computed to each partition C_i as

$$\phi_1\left(Obj_{intra-neig}(C_i(G))\right) := \frac{(G_{intra-nei}(C_i(G)) + \sum_{\forall v \in C_i} Weak_{C_i(G)}(v))}{N(C_i)}; --(4)$$

Global node search based inter-neighbor score is computed to each partition C_i as

$$\phi_2(Obj_{inter-neig}(C_i(G))): = \frac{(G_{inter-neig}(C_i(G)) + \sum_{\forall v \in C_i} Strong_{C_i(G)}(v))}{\sum_{v \in C_i} G_{inter-nei}(C_i(G))} --(5)$$

$Multi - Objective$ Community Detection Problem is given as

$$\text{Min}\{\phi_1(Obj_{inter-neig}(C_i(G))), \phi_2(Obj_{intra-neig}(C_i(G)))\} ----(6)$$

In the above computational measures, eq (1) is used to find the weak nodes using the intra-weighted graph nodes. Eq(2) is used to compute the neutral nodes using the intra-weighted graph nodes. Eq (3) represents the computational of strong nodes using the intra-neighbour graph nodes.

Algorithm-2

(K,D) anonymity based node privacy preserving(Proposed privacy preserving anonymity approach)

Input: Construct Degree Graph G

Output: Graph G with tree structure

1. Compute degree sequence(DS) from G.

2. Find the summative count of each node in the G using the degree sequence using vector A.

3. Construct the tree using the vector A.

4. Combine the node with lower counts in the tree.

5. Find new node count and its degree in the tree.

6. if the number of nodes in new level is greater than 1 then repeat step 4.

8. Apply Homomorphic encoding on each node centrality and key node selection

$$\alpha = \text{NodeCent} = \frac{|N(v_i) \cup N(v_j)| - |N(v_i) \cap N(v_j)|}{2.\max\{\deg(v_i), \deg(v_j)\}}$$

$$\beta = |N(v_i) \cap N(v_j)|$$

Algorithm 3: Graph Community detection using hybrid clustering algorithms

In this algorithm, a hybrid graph community clustering approach is implemented on the selected essential key nodes using the algorithm 2. In this approach, a hybrid KNN approach is designed and implemented on the key strong graph nodes.

Find the nearest density objects using the proposed probabilistic KNN method.

$$Dist_c = mean^K + \kappa . \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\phi_i^K - mean^K)^2}$$

Here, N is total number of current cluster objects ϕ_i^K is the average of the kth nearest neighbour to ith node. & $\phi_t^K = max_j\{KNN_i(Dist_{ij})\}$, and $mean^K$ is the average of ϕ_i^K ,

computed as $\text{mean}^K = \frac{1}{N}\sum_{i=1}^{N}\phi_i^K$

Influencing Factor =κ=Max{D(O_j),C_k : k-nearest objects}

Pr o posed local density estimation is given as

$$PLDE(v_i) = \frac{1}{\eta}\, e\left(-\frac{\|log(CD(v_{ij})) - Dist_c\|^2}{2{\sigma_c}^2}\right) \sum_{j\in KNN_i}^{\Sigma\frac{CD(v_{ij})^2}{Dist_c}} T.e$$

- Filter all the k-nearest neighbour objects using the local kernel density estimation.
- Done
- Done

4. EXPERIMENTAL RESULTS

Experimental results are performed on different OSN datasets such as facebook, AstroPh ,enron and Gplus for privacy preserving and community clustering. Table 1, summarization of different datasets and its properties.

Table 1: Different graph based online social networking data and its properties

Dataset	No of Nodes	Number of edges
AstroPh	18,772	198,110
Facebook	4,039	88,234
Gplus	107,614	12,238,285
Enron	36,692	183,831

a) Average Mutual information

Table 2: Average Mutual Information on AstroPh

Model	AMI
LDPGen	0.55
LF-GDPR	0.81
Proposed	0.94

Table 1,describes the comparative study of average mutual information of existing graph based PPDM models to the proposed model on AstroPh dataset.

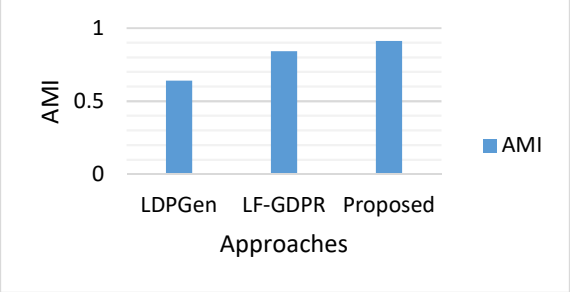

Figure 2: Average Mutual Information on Facebook

Figure 2,describes the comparative study of average mutual information of existing graph based PPDM models to the proposed model on facebook dataset.

Table 3: Average Mutual Information on Gplus

Model	AMI
LDPGen	0.62
LF-GDPR	0.86
Proposed	0.96

Figure 2,describes the comparative study of average mutual information of existing graph based PPDM models to the proposed model on Gplus dataset.

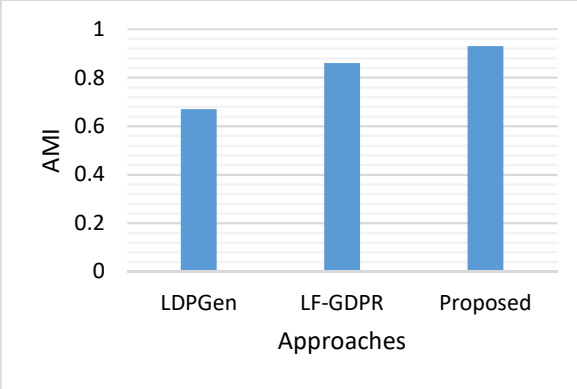

Figure 4: Average Mutual Information on Enron

Figure 4,describes the comparative study of average mutual information of existing graph based PPDM models to the proposed model on Enron dataset.

b)Runtime

Table 4: Average runtime analysis of proposed model to conventional graph based PPDM models on AstroPh dataset

Model	Runtime(ms)
LDPGen	3844
LF-GDPR	2844
Proposed	2193

Table 4,describes the comparative study of average runtime analysis of existing graph based PPDM models to the proposed model on AstroPh dataset.

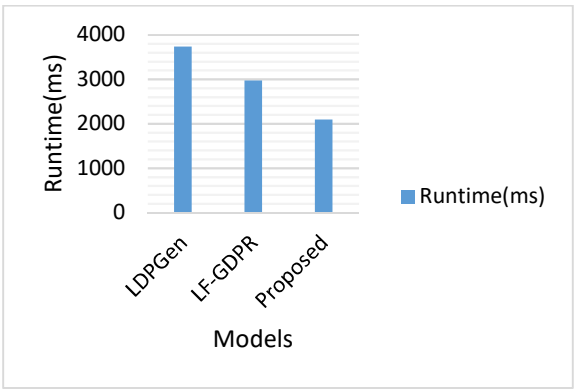

Figure 5: Average runtime analysis of proposed model to conventional graph based PPDM models on Facebook dataset

Figure 5,describes the comparative study of average runtime analysis of existing graph based PPDM models to the proposed model on Facebook dataset.

Table 5: Average runtime analysis of proposed model to conventional graph based PPDM models on Gplus dataset

Model	Runtime(ms)
LDPGen	3574
LF-GDPR	2643
Proposed	2104

Table 5,describes the comparative study of average runtime analysis of existing graph based PPDM models to the proposed model on Gplus dataset.

Table 6: Average runtime analysis of proposed model to conventional graph based PPDM models on Enron dataset

Model	Runtime(ms)
LDPGen	3974
LF-GDPR	2974
Proposed	2294

Table 6, describes the comparative study of average runtime analysis of existing graph based PPDM models to the proposed model on Enron dataset.

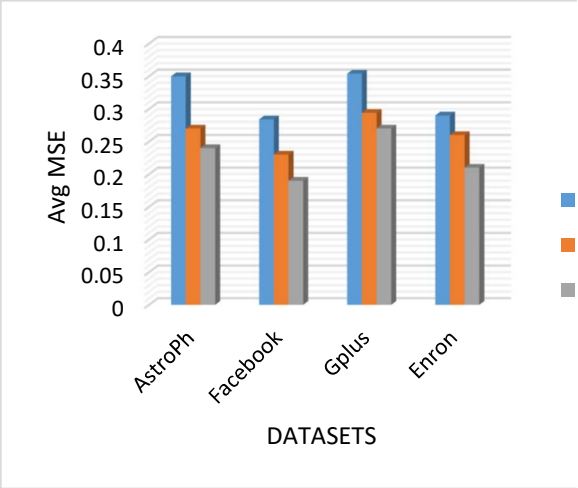

Figure 6: Average MSE analysis of proposed model to conventional graph based PPDM models on different OSN dataset

Figure 6, describes the comparative study of average mean squared error analysis of existing graph based PPDM models to the proposed model on different OSN datasets.

Analysis :

In this research work, proposed intra and intra graph-based community detection approach is tested and compared on different traditional approaches such as LDP-Gen, LF-GDPR on different computational metrics. In this work, average mutual information, average runtime and average mean squared error are used in order to improve the analysis. From the results, it is concluded that the proposed approach has better optimization than the conventional approaches on different large OSN datasets.

5. CONCLUSION

Most of the conventional graph based community clustering and privacy preserving models are independent of local and global optimization in the graph clustering process. Traditional graph based models are independent of weighted node density and weighted connection ranking. In this work, an efficient meta-heuristic optimization based community clustering and privacy preserving model is implemented on different online social networking databases. In the proposed model, local and global optimization measures are used to improve the node selection for the privacy preserving process along with the community clustering process. Experimental results show that the proposed model has better runtime, average mutual entropy and error rate than the conventional approaches. The main limitations of this work is to improve the large size graph nodes with parallel processing framework in order to minimize the overall computational time and memory.

REFERENCES:

[1] O. Abul, "Location-privacy preserving partial nearby friends querying in urban areas," Data & Knowledge Engineering, vol. 139, p. 102006, May 2022, doi: 10.1016/j.datak.2022.102006.

[2] E. Aghasian, S. Garg, and J. Montgomery, "An automated model to score the privacy of unstructured information—Social media case," Computers & Security, vol. 92, p. 101778, May 2020, doi: 10.1016/j.cose.2020.101778.

[3] L. M. Aiello and G. Ruffo, "LotusNet: Tunable privacy for distributed online social network services," Computer Communications, vol. 35, no. 1, pp. 75–88, Jan. 2012, doi: 10.1016/j.comcom.2010.12.006.

[4] L. Alkhariji, S. De, O. Rana, and C. Perera, "Semantics-based privacy by design for Internet of Things applications," Future Generation Computer Systems, vol. 138, pp. 280–295, Jan. 2023, doi: 10.1016/j.future.2022.08.013.

[5] A. Altameem, V. Kovtun, M. Al-Ma'aitah, T. Altameem, F. H, and A. E. Youssef, "Patient's data privacy protection in medical healthcare transmission services using back propagation learning," Computers and Electrical Engineering, vol. 102, p. 108087, Sep. 2022, doi: 10.1016/j.compeleceng.2022.108087.

[6] A. Andersen and M. Saus, "Privacy preserving distributed computation of community health research data," Procedia Computer Science, vol. 113, pp. 633–640, Jan. 2017, doi: 10.1016/j.procs.2017.08.319.

[7] U. I. Atmaca, C. Maple, G. Epiphaniou, and M. Dianati, "A privacy-preserving route planning scheme for the Internet of Vehicles," Ad Hoc Networks, vol. 123, p. 102680, Dec. 2021, doi: 10.1016/j.adhoc.2021.102680.

[8] Y. Benkaouz and M. Erradi, "Towards a Decentralized OSN for a Privacy-preserving e-health System," Procedia Computer Science, vol. 63, pp. 284–291, Jan. 2015, doi: 10.1016/j.procs.2015.08.345.

[9] S. Y. Bhat and M. Abulaish, "Using communities against deception in online social networks," Computer Fraud & Security, vol. 2014, no. 2, pp. 8–16, Feb. 2014, doi: 10.1016/S1361-3723(14)70462-2.

[10] X. Bi and X. Shen, "Distribution-invariant differential privacy," Journal of Econometrics, Jun. 2022, doi: 10.1016/j.jeconom.2022.05.004.
[11] P. Pham, L. T. T. Nguyen, B. Vo, and U. Yun, "Bot2Vec: A general approach of intra-community oriented representation learning for bot detection in different types of social networks," Information Systems, vol. 103, p. 101771, Jan. 2022, doi: 10.1016/j.is.2021.101771.
[12] D. Naik, D. Ramesh, A. H. Gandomi, and N. Babu Gorojanam, "Parallel and distributed paradigms for community detection in social networks: A methodological review," Expert Systems with Applications, vol. 187, p. 115956, Jan. 2022, doi: 10.1016/j.eswa.2021.115956.
[13] I. Koc, "A fast community detection algorithm based on coot bird metaheuristic optimizer in social networks," Engineering Applications of Artificial Intelligence, vol. 114, p. 105202, Sep. 2022, doi: 10.1016/j.engappai.2022.105202.
[14] 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, Jul. 2022, doi: 10.1016/j.physa.2022.127314.
[15] M. Huang, Q. Jiang, Q. Qu, L. Chen, and H. Chen, "Information fusion oriented heterogeneous social network for friend recommendation via community detection," Applied Soft Computing, vol. 114, p. 108103, Jan. 2022, doi: 10.1016/j.asoc.2021.108103.
[16] N. E. Díaz Ferreyra, T. Hecking, E. Aïmeur, M. Heisel, and H. U. Hoppe, "Community detection for access-control decisions: Analysing the role of homophily and information diffusion in Online Social Networks," Online Social Networks and Media, vol. 29, p. 100203, May 2022, doi: 10.1016/j.osnem.2022.100203.
[17] G. M. Garrido, J. Sedlmeir, Ö. Uludağ, I. S. Alaoui, A. Luckow, and F. Matthes, "Revealing the landscape of privacy-enhancing technologies in the context of data markets for the IoT: A systematic literature review," Journal of Network and Computer Applications, vol. 207, p. 103465, Nov. 2022, doi: 10.1016/j.jnca.2022.103465.
[18] A. Guarino, D. Malandrino, and R. Zaccagnino, "An automatic mechanism to provide privacy awareness and control over unwittingly dissemination of online private information," Computer Networks, vol. 202, p. 108614, Jan. 2022, doi: 10.1016/j.comnet.2021.108614.
[19] C. Guo, J. Jia, K.-K. R. Choo, and Y. Jie, "Privacy-preserving image search (PPIS): Secure classification and searching using convolutional neural network over large-scale encrypted medical images," Computers & Security, vol. 99, p. 102021, Dec. 2020, doi: 10.1016/j.cose.2020.102021.
[20] Y. Himeur, S. S. Sohail, F. Bensaali, A. Amira, and M. Alazab, "Latest trends of security and privacy in recommender systems: A comprehensive review and future perspectives," Computers & Security, vol. 118, p. 102746, Jul. 2022, doi: 10.1016/j.cose.2022.102746.
[21] D. H. Ho, Y. Lee, S. Nagireddy, C. Thota, B. Never, and Y. Wang, "OpenComm: Open community platform for data integration and privacy preserving for 311 calls," Sustainable Cities and Society, vol. 83, p. 103858, Aug. 2022, doi: 10.1016/j.scs.2022.103858.
[22] M. Kamal et al., "Privacy-aware genetic algorithm based data security framework for distributed cloud storage," Microprocessors and Microsystems, vol. 94, p. 104673, Oct. 2022, doi: 10.1016/j.micpro.2022.104673.
[23] S. Kavianpour, A. Tamimi, and B. Shanmugam, "A privacy-preserving model to control social interaction behaviors in social network sites," Journal of Information Security and Applications, vol. 49, p. 102402, Dec. 2019, doi: 10.1016/j.jisa.2019.102402.
[24] I. Kayes and A. Iamnitchi, "Privacy and security in online social networks: A survey," Online Social Networks and Media, vol. 3–4, pp. 1–21, Oct. 2017, doi: 10.1016/j.osnem.2017.09.001.
[25] J. W. Kim, K. Edemacu, and B. Jang, "Privacy-preserving mechanisms for location privacy in mobile crowdsensing: A survey," Journal of Network and Computer Applications, vol. 200, p. 103315, Apr. 2022, doi: 10.1016/j.jnca.2021.103315.
[26] S. Kumar and P. Kumar, "Upper approximation based privacy preserving in online social networks," Expert Systems with Applications, vol. 88, pp. 276–289, Dec. 2017, doi: 10.1016/j.eswa.2017.07.010.
[27] L. La Cava, S. Greco, and A. Tagarelli, "Information consumption and boundary spanning in Decentralized Online Social Networks: The case of Mastodon users," Online Social Networks and Media, vol. 30, p. 100220, Jul. 2022, doi: 10.1016/j.osnem.2022.100220.