DOI: https://doi.org/10.52756/ijerr.2024.v41spl.015 *Int. J. Exp. Res. Rev.*, Vol. 41: 180-194 (2024)

Original Article | Peer Reviewed | Open Access

International Journal of Experimental Research and Review (IJERR)
© Copyright by International Academic Publishing House (IAPH)
ISSN: 2455-4855 (Online)
www.iaph.in
ISSN 2455-4855

Hybrid Filtering and Probabilistic Techniques for Privacy-Preserving Community Detection in OSNs

Shamila.M[1], G. Rekha[2] and K. Vinuthna Reddy[3]

[1]Department of CSE, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur, A.P, India;
[2]Department of CSE, Koneru Lakshmaiah Education Foundation, Aziz Nagar, Hyderabad, Telangana, India;
[3]Department of CSE, Neil Gogte Institute of Technology, Uppal, Hyderabad, Telangana, India

E-mail/Orcid Id:

SM, m.shamila2024@gmail.com, https://orcid.org/0000-0001-9105-9531;
GR, gillala.rekha@klh.edu.in, https://orcid.org/0000-0003-2688-2323;
KV, vinuthna.dwh@gmail.com, https://orcid.org/0000-0002-2753-3216

Article History:

Received: 09th Jun., 2024
Accepted: 23rd July, 2024
Published: 30th July, 2024

Keywords:
Data preserving, Graph cluster detection, Social network data

How to cite this Article:
Shamila.M, G. Rekha and K. Vinuthna Reddy (2024). Hybrid Filtering and Probabilistic Techniques for Privacy-Preserving Community Detection in OSNs. *International Journal of Experimental Research and Review*, *41*(spl.), 180-194.
DOI:
https://doi.org/10.52756/ijerr.2024.v41spl.015

Abstract: Online Social Networks (OSNs) face the major challenge of protecting participant's privacy, due to the high dimensionality and volume of the data. In real-time social networks, where hundreds of personal details of people are shared every day, there remains a significant threat to privacy. Privacy preservation is challenging for a community detection problem due to the high computational complexity and memory requirements, especially in larger real-world OSN graphs. Although weighted nodes provide better results, as they allow capturing the frequencies of the values, the privacy preservation of sensitive attributes such as specific profiles becomes harder compared to these models. This problem can result in queries and subsequent learnings from social network profiles of specific individuals, which may be of personal, political or otherwise concern. Online social networks (OSNs) grapple with significant privacy challenges due to the extensive dimensions and vast quantities of data involved. This research fills the void in current privacy-preserving community detection methodologies, which face problems in computational complexity and memory usage in large-scale OSNs. The proposed framework seeks to bolster privacy preservation through a comprehensive multi-step process. Data filtering uses a blended data filter system to remove outliers and irrelevant data, thus enhancing the quality of the input data. Density-Oriented Clustering phase employs a density-oriented clustering model to identify communities, with each cluster representing a distinct community. Privacy Preservation component introduces a new privacy preservation technique for sensitive OSN attributes, surpassing existing k-anonymization methods. The developed density-based social network community detection model and its novel privacy-preserving scheme are evaluated using the datasets Yelp, Football, Zachary and Dolphin from the SNAP dataset. Experimental results on these datasets embed a comprehensive evaluation based on the order of each node and the graph-based networks, where each node is laden with weights as proximity values, indicating the semantic proximity between communities and individuals. The proposed framework employs the normalized mutual information (NMI), modularity (Q), Rand index and runtime measurements to demonstrate widespread advantages in the multi-dimensional functional space, including greater accuracy, cluster compatibility, and computational tractability over the existing prominent traditional models for OSNs.

Introduction

The term 'social network' was first used in 1954 by J.A. Barnes to signify the social structure of a population consisting of nodes (like people or things) and edges (like dependencies or relationships between the nodes). With the advent of online social networks (OSNs), the concept of a social network has taken new and novel forms. OSNs have gained widespread popularity for a

*Corresponding Author: m.shamila2024@gmail.com

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

variety of purposes, such as communication, information-sharing and entertainment. Individual services in OSNs are targeted to the purpose and form of usage. For example, Facebook is a social networking site, LinkedIn is a professional network, Instagram is a photo-sharing site, YouTube is a video-sharing site, Tumblr is a blogging and content-sharing site, and Twitter is a microblogging site. These platforms have become part of the daily activities of millions of users. OSNs provide a broad range of functions (networking, microblogging, video sharing, etc), but the core features of OSNs look remarkably similar across platforms. OSN user numbers have risen exponentially from less than 10 million in 2005 to 2 billion today, allowing people to create virtual friendships with unknown and known others. OSNs reward users for sharing information with services that link them to others, boost their friend count, enhance their fame and help them build trust with their affinity groups, often leading to herding behaviour (Zhou et al., 2015). This dynamic helps to explain the rapid rise of OSNs and their daily use.

Hybrid filtering techniques have become vital tools in improving data quality and ensuring robust privacy measures. These techniques leverage the strengths of various filtering methods to remove outliers, reduce noise, and filter out irrelevant values from the input data. By enhancing data quality, hybrid filtering boosts the accuracy of community detection algorithms and fortifies the overall privacy-preserving framework.

Increase in OSN Users Platforms like Facebook, Instagram, and Twitter have seen a dramatic rise in user's data. This upward trend is likely to continue, with more users joining these platforms and sharing personal information. Users increasingly share a vast array of personal information, including location data, videos, preferences, and photos. While this data is valuable for social networking, it poses significant privacy risks if not adequately protected. These regulations require more rigorous data protection measures, compelling OSNs to adopt advanced privacy-preserving techniques. There have been substantial advancements in privacy-preserving methods, including homomorphic encryption, secure multi-party computation and differential privacy as well. These techniques provide new ways to protect user data while allowing for meaningful data analysis. Hybrid filtering techniques (Kumar et al., 2023) are particularly effective in addressing privacy challenges in OSNs due to their efficiency in handling large-scale data. By integrating multiple filtering methods, hybrid techniques can more effectively remove outliers and irrelevant values, resulting in higher-quality data. This improved data quality is crucial for accurate community detection and the implementation of robust privacy-preserving schemes.

The openness is embedded in the underlying principle of the modern internet, which is an ecosystem where everyone is a contributor, of data, of information, of everything. Most of this is intentional sharing: users share information they think is appropriate. But a lot of it is unintentional sharing, where the users' interactions are captured, aggregated and correlated. Every time you use the internet, you're creating something bigger and richer than you were before and leaving a trail that benefits others, whether you're aware of it or not. This is how the modern internet works. It is nothing if not connected. Sharing personal information has become so widely recognized as useful that it is delivering huge utility to OSN applications. However, by doing so, it is also exposing the user to inherent privacy risks; and this cannot be ignored either. Therefore, it is necessary to balance the utility of information sharing against the user's privacy requirements in OSN applications (Prasad et al., 2019). This balancing act can be achieved only if the privacy measures are robust enough, that the sharing of information continues to deliver value to the user while at the same time building trust and continuing the use of the platform by the user.

Online Social Networks (OSNs) is a technological innovation denoting a software platform that is used for communication and exchange of information. Different types of OSNs support and facilitate different types of user behavior and interests. These behaviors and interests are diverse and vary across individuals. OSNs provide a number of functions to cater to different tastes and needs. OSNs facilitate the development and exchange of messages and ideas (Bandara et al., 2021) among different groups of people in different communities. They have profoundly altered the way people communicate, spread information and develop relationships, be it in the field of communication and information diffusion.

There is a vast amount and diversity of work still in progress on the issues of privacy in OSNs. The research focuses on some of the most challenging questions in this field – how to protect one of the most crucial factors in our world, our personal information. Westin's theory of privacy defines privacy as that guarantees individuals control over if, how and with whom personal information is shared. One way privacy protections in OSNs could achieve this is by providing mechanisms that allow people to control access to their personal information, thereby enabling them to if not completely

DOI: https://doi.org/10.52756/ijerr.2024.v41spl.015

retain at least some level of autonomy over their personal information.

In OSNs, people or organizations are linked by relationships or communications. OSNs are a virtual playground where users write, think, and share, enabling a host of activities ranging from advertising to blogging, reviews and political campaigns. These OSNs – such as Facebook, WhatsApp, Twitter, Flickr, Instagram, and others – are a part of our daily lives and make the world smaller by connecting people globally and transmitting information across the world.

Social networks seek constant expansion – their user base and features. Linked networks are especially useful because it takes less time for information to travel across them, making them more efficient and valuable to their users. This expansion and linking of social networks characterizes how digital social networks evolve and grow ever more useful in the digital age. Link prediction simply entails the prediction of links between two unconnected nodes within a network (Singh and Singh, 2023; Jha et al., 2024). The main objective of link prediction (Pensa and Blasi , 2017) observes, is to increase the number of links and hence connectivity within the network, growing the network and strengthening the ties of the users within it. This helps the social network to be more robust and valuable.

The predictive querying engine at the heart of these features combines multiple factors, including common friends, interests and employers, to suggest new people who might make informative or beneficial connections. A substantial amount of new linkage in these networks comes from such features. This is no small matter; the ability of predictive tools to orient users toward more relevant connections could vastly accelerate network growth by pointing users towards other people they might find interesting or valuable to know. Evidence of the role played by authenticity and pseudo-anonymity in the ecosystem of OSNs: a number of OSNs, such as Facebook and LinkedIn, require real names and sometimes verification of Facebook identities for people to register and participate. Significantly, this is because the relationships in these sites are intended to mirror as closely as possible the real world, and thus are highly critical of who gets to participate in the network. The ecosystem of trust and credibility requires that people's real names be used. By contrast, on dating sites, many of which allow pseudo-anonymity, users can pursue genuine social interactions in an ecosystem of people whose identities, such as banking details and home addresses, are protected, and they can maintain their privacy.

The dual approach of encouraging authenticity at large OSNs and allowing pseudo-anonymity at dating sites encourages users of different types of social networks to have different expectations of privacy. In order to remain relevant to their users, OSNs have to safeguard their balance between boosting real interactions and protecting the user's privacy. Virtual relationships can only be sustained if users have a sense of security and comfort in sharing personal information; the balance becomes an essential element between privacy and authenticity for an OSN to stay afloat in social networking. In an effort to keep up with the times, OSNs must grow and continue to evolve to keep up with the accommodation of user needs with respect to privacy and authenticity.

Among the major risks to user privacy on OSNs is their vulnerability against linking attacks. In the former, a linking attack occurs when, compared to some others, an attacker is able to learn more about a person by linking that person's profiles across different OSNs, or from pieces of information, such as revealing user profiles, attributes and friend lists (Nicolazzo et al., 2020). The final stage in this process is privacy leaks and information misuse. It, therefore, needs robust privacy protections within the OSNs. The users should also be aware of the risks involved and how to overcome the vulnerabilities to linking attacks.

Another critical OSN security issue is related to the level of trust in spam messages sent by friends. Spammers send spam messages that appear to be from trusted contacts using e-mail lists, which probably increases the likelihood of the target trusting enough of the message to click on a link. TPAs can also provide user attributes that can help spammers make their phishing look more convincing and tailored (Madhuri et al., 2022). With spamming and phishing, the worst case can involve financial loss, identity theft and other types of cybercrime. Users need to be alert and OSNs must effectively detect and prevent spam.

Cyberbullying is still common on OSNs, and such messages should not be allowed to go viral. In the context of link prediction and recommendation systems, which are essential for suggesting new connections, it should be able to compute comparability scores between unconnected nodes and allow users to understand the effect of their recommendation on the chance of two people linking. In parallel, data-mining activities should include data encryption before it is released and the design of task-application-agnostic privacy tools that support sensitive data and protect processing time while ensuring robustness of PPDM models. Adversarial

models should also be developed to estimate risk and sanitize data using k-anonymous tables and, L-diversity and other techniques to avoid homogeneity attacks and attacks that allow sensitive attributes to be breached. This will preserve individual privacy as more and more of the world goes online.

The transformation-based approach for data sharing is critical as it allows for the secure release of sensitive attributes to third parties. In this case, the sensitive data is released to a third party which transforms the data and obscures the original value of the dataset. After transformation, the mining results are released to relevant stakeholders who can benefit from the knowledge of the data without having access to the original data, thus preserving individual privacy. In the case of the transformation-based approach, a risk assessment of data disclosure is required to ensure that the transformation effectively prevents re-identification and unauthorized access. In the case of social networks online, information security systems face a set of unique challenges. We must consider how to deal with the scalability of these systems – think of the problems involved with exponentially growing networks – and the heterogeneity and evolution of data in those networks. How do we evaluate the effectiveness of the systems? How can we harness collective intelligence and maintain some sort of privacy? These challenges require increasingly sophisticated solutions that balance privacy, functionality and user trust.

The problem of scaling is important in community detection as well. Social networks grow exponentially, and many algorithms cease to perform optimally when applied to networks of that size. We need scalable algorithms in order to deal with an exponential increase in the size and complexity of social networks. When networks become too large, the ability to analyze and interpret the output of social network analysis diminishes. As a result, we are prevented from applying this technology in areas as varied as marketing or security.

The second challenge is heterogeneity: complex graph/network representations such as hypergraphs or k-partite graphs can describe social media networks. Intuitive network representations tend to be simple and may not be sufficient to capture the heterogeneity of social media interactions. For instance, simplistic network representations of social media interactions are likely to lack critical insights. We may need more sophisticated modelling tools to capture the heterogeneity of social media interactions in a computationally feasible manner. These tools, such as multi-typed networks, are capable of handling heterogeneous data types and relationships. Handling the temporal dynamics of social media networks is a key part of understanding social dynamics over time. The fact that networks are continuously evolving means grasping the specificity of time dynamics is crucial in measuring and interpreting network evolution. For instance, tracking relationships' formation, change and disappearance is important for analyzing social dynamics. Understanding how the different parts of networks evolve over time is also important. Thus, developing ways to handle social networks' temporal dynamics will help keep social media analytics up to date and make social data more useful in making managerial decisions.

Social networks are difficult to evaluate as the community-detection algorithms are regularly given only an arbitrary ground-truth set because there is no objective means of defining the community set otherwise. This means that the performance of an algorithm can only be judged by the subjective assessments of people who are the users of these networks. There are, therefore, few established standards for measuring community-detection performance.

The third challenge lies in extracting this collective wisdom from the user-generated content. As we have seen, user-generated content, such as comments, reviews or ratings, can hold a lot of valuable information but is often difficult to extract and analyze in a timely and cost-effective manner because of its unstructured nature. The potential value of such information is often lost because of its sheer volume and diversity. Thus, another challenge is to develop sophisticated techniques for efficient extraction and analysis of collective intelligence. An issue that has persistently drawn attention is privacy. Social media companies such as Facebook and Google are often involved in privacy debates. Privacy-protecting mechanisms that keep user data secure are crucial but often difficult to design. This is especially due to the limitations of anonymization techniques and the need for users to maintain trust in the privacy system. Guaranteeing a proper level of privacy while enabling the utility of the data is a challenging balance, and it requires permanent improvements in privacy-enabling techniques and policies that can adapt to the evolution of social media.

The research into community detection aims to compare benchmark datasets against crawled datasets, which would return information about the detected communities' quality and the detection algorithms' scalability. This provides an analysis of how well they are

at identifying meaningful, cohesive subgroups from large, complex networks. The algorithms should be scalable since social networks are never growing exponentially. One would like to contrast performance across these different types of datasets to understand the strengths and weaknesses and develop a better community detection technique.

Challenges in social media data (Aghaalizadeh et al., 2021; Beg et al., 2021; Kahate and Raut, 2022; Kavinpour et al., 2019), in terms of its noisy, distributed, unstructured, and dynamic nature, make community detection a still harder problem (Rao, M.S. et al., 2022). In this respect, the objective will be to develop efficient algorithms that could map the social network analysis to graph partition problems and enable the detection of useful communities. The proposed algorithms should be tested on the sampled networks—for instance, a Twitter network of sports persons—to prove their efficiency (Keerthana et al., 2023). The authors investigate overlapping of community detection, sub-community-based, and community-based, and give new approaches to find them (Tran et al., 2021). Results will be presented in graphs to enable visualization of performance and effectiveness of the different community detection approaches to understand the positives and negatives of various methods.

The surge in Online Social Networks (OSNs) has led to a tremendous expansion in the amount of personal data shared, making the protection of privacy a paramount concern. Current statistics reveal that the number of OSN users has escalated from below 10 million in 2005 to more than 2 billion today, emphasizing the enormous quantity of personal information at risk. This rapid escalation underscores the importance of implementing strong privacy-preserving strategies to safeguard user data. One of the key issues in OSNs is the extensive dimensionality and volume of data. With users sharing numerous personal details daily, the threat of privacy violations grows. Traditional privacy preservation methods often struggle with the computational demands and memory requirements necessary to process such large data sets in real time. Therefore, there is a pressing need for more effective and scalable solutions.

Related works

PPDM is a topic that has sparked enormous interest lately due to an increasing public need to protect sensitive information from probable exposure in the course of data analysis (Barsocchi et al., 2021). They mentioned that PPDM essentially takes the form of a collaboration between two or multiple parties that have sensitive input databases and participate in a common task of data mining while keeping their data confidential.

Privacy-Preserving Techniques Described

The privacy-preserving techniques of data mining can be considered as the collective efforts of multiple parties having sensitive databases. These various parties jointly do data mining tasks, preserving the privacy of their datasets. One of the prime objectives of this approach is to gain useful results from data mining without giving away critical information about the databases (Gupta et al., 2018).

Objective: Obtain Results Without Exposing Critical Information

The aim of PPDM is to ensure that the results are achieved but with no revelation of the sensitive information. This objective is the central concept in PPDM, hence the protection of the privacy of data during mining. This objective applies to three broad aspects of PPDM: input privacy, output privacy, and minimizing discrepancies. These aspects make certain conditions whereby the input data and the result of data mining remain confidential, along with the preservation of data integrity through any transformation.

Input Privacy

Input privacy deals with ensuring the privacy of submission information. In the process of privacy-preserving data mining, the security and privacy of data provided for mining by different parties should be guaranteed.

Output Privacy

Output privacy ensures that the results obtained from activities of Datamining are private and integrity guaranteed. This simply means ensuring that the results mined will only be available to the intended recipient and no other person can have access, hence avoiding leaking out data. Ensuring output privacy conserves the confidentiality of information derived from the mining process by privacy-preserving data mining.

Minimizing Discrepancy

The next is the reduction of discrepancy—minimizing the difference between the original data and the one to be used for mining. In doing this, it is important to ensure that the integrity and accuracy of the data are observed. By minimizing the discrepancies, privacy-preserving data mining ensures that the transformation does not alter the real data greatly; therefore, the utility of the data is maintained.

Two Common Methodologies for Mining Distributed Databases

The centralized and distributed approaches are the most popular methodologies of mining distributed databases in privacy-preserving data mining. The centralized approach involves centralizing the data in one

DOI: https://doi.org/10.52756/ijerr.2024.v41spl.015

place, whereas replicas of data at each site and then transferred it to a common central location made by the distributed approach. Both methodologies have advantages and challenges with respect to the speed, accuracy, and privacy of knowledge discovery.

Centralized Approach

In centralization privacy, data for data mining resides at a single centralized point (Bourahla et al., 2020; Sai and Li, 2020), which is faster and more efficient in computation. However, this comes with high privacy risks since aggregation of sensitive data from different sources in one point makes the data more at risk of data breaches and unauthorized access. One of them is the centralized privacy-preserving location-sharing system, a system named CenLocShare (Xiao et al., 2018), which combines a location-based server and social network server into one other server – Location-storing Social Network Server employs dummy locations and uses dedicated mapping protocols between the LSSNS and the Cellular Tower while using dummy locations to share the privacy-preserving locations.

Distributed Approach

Data at each site are duplicated and sent to a central location (Bahri et al., 2018; Sun et al., 2019) in the distributed approach. This strategy favors accuracy and privacy since the data will remain distributed across several locations, which, though slower in speed compared to the centralized approach, has the added advantage of increased protection from the risk reduction involved in centralizing sensitive data.

Drawback: Introduction of False Transactions

The major disadvantages of dummy transaction addition are the introduction of false transactions, as it is bound to increase the size of the dataset, extend the transformation time, and reduce the overall utility of data. As much as this methodology has these drawbacks, it stands a great chance of effectively protecting sensitive rules by reducing the likelihood of their exposure during mining.

Alternative Technique: Addition of Noise in Decision Tree Classifiers

Another alternative technique is to introduce noise into the decision tree classifiers themselves (Zhang et al., 2019). It counts class label comparability between the original dataset and transformed through shuffling or other alteration of attribute values. While this offers privacy protection to some degree, it is at lowering utility costs in the classification model. The similarity between how much of the transformed dataset's characteristics reflect in the original dataset and how useful the latter can be onboarding utility are delicate balancing factored into this method alone.

Another Approach: Safeguard Extracted Knowledge

Safeguarding against extracted knowledge can be done using randomization and k-anonymization. Randomization attempts to make the patterns of data distorted, which can be treated as working on attribute transitional probability matrices followed by applying k-anonymity on the randomized data to maintain privacy. However, k-anonymity has problems dealing with issues such as the homogeneity of the dataset, which might degrade the overall efficiency of this technique.

Link Prediction Techniques

In privacy-preserving data mining, these link prediction techniques (Daud et al., 2020; Zareie & Sakellariou, 2020) are reductive, using keyword-based matching and text similarity. All of these methods calculate the similarity of nodes pair-wise and have proved to be successful in different scenarios; however, effective privacy controls and concerns regarding data leakage remain a significant problem, especially in the case of online social networks. This points to the very relevant cyberbullying issue of the finite privacy-control capability of OSNs and the high effectiveness of detection solutions, which demands robust privacy-preserving techniques (Zhao et al., 2019; Zheng et al., 2019) at the same time.

PPDM Techniques: Transform Data While Maintaining Privacy

Several PPDM techniques focus on transforming data in such a way that privacy is guaranteed and striking a good balance between data utility and privacy preservation. These include item-control designs, SIF-IDF, and tree structures to reduce the number of noises added or datasets removed. These proposed techniques aim to improve performance and reduce the number of scans taken from the database to ensure effective and efficient PPDM. The other techniques intend to add random noise (Wei et al., 2019) to the data to ensure individuals' privacy. This technique guards input privacy because it ensures the confidentiality of input data. One of the most influential approaches is the k-anonymity model, which is for guarding output privacy by anonymizing the data in such a way that every record is indistinguishable from at least k-1 other records. However, k-anonymity has its flaws, such as the risk of a homogeneity attack, which impelled researchers to work on more advanced models like l-diversity and t-closeness. It contributes to the methodologies for mining distributed databases with a concept of secure multi-party computation as a tool, enabling data mining across

DOI: https://doi.org/10.52756/ijerr.2024.v41spl.015

multiple parties without revealing the individual datasets. This work (Kantarcioglu & Clifton, 2004) pointed out the trade-off between the centralized and distributed approaches. On the other hand, while centralized approaches could provide results faster, distributed methods enhance privacy since data remains distributed.

Other recent techniques in PPDM include randomization and k-anonymization, with the former referring to the use of attribute transitional probability matrices that are used to confuse patterns in data (Li et al., 2020; Liu et al., 2020; Kayes and Iamnitchi, 2017; Yang et al., 2021). While these techniques are quite effective in enhancing privacy, they are hindered by several challenges, such as the homogeneity of the dataset and the utility of the transformed data. These techniques would apply to include personal information (Jain et al., 2021) shared online for its preservation in an SNS context.

Material and Methods

Data privacy preservation using a filtered-based technique in online social networking

The presented work constitutes a community detection framework with a privacy-preserving data filter mechanism for the datasets of online social networking. In the pre-processing step, input data is filtered. Since social networking data is often in the form of numbers, a hybrid data filter mechanism is applied to remove outliers. In the next step, the filtered data is given to the privacy-preserving model, a clustering approach based on the density community. The clustering model of the probabilistic model is applied to the filtered data to detect the human communities. In this particular framework, each cluster denotes a single community and all the data points lying within the clusters contribute to intra-cluster variance, while the objects between the clusters add inter-cluster variance due to the difference in the data attribute. At the last step, it adopts a privacy-preserving approach to sensitive attributes such as user profile information. The presented framework is illustrated in Figure 1 below.

Figure 1. Overview of the proposed framework. Privacy of the profile-sensitive attributes is maintained not through traditional k-anonymization algorithms but rather through a new privacy-preserving scheme.

PPDM OSN Graph Data Pre-processing

The pre-processing of online social network (OSN) graph data involves handling multiple sources of data (Ψ) and systematically processing each training dataset ($\Delta[\iota]$). The attributes must be normalized and encoded appropriately depending on their type for each record ($\Pi[\rho]$) in a given training dataset.

Continuous Attribute Normalization

For each continuous attribute ($\alpha\tau[\Pi]$) that is not empty

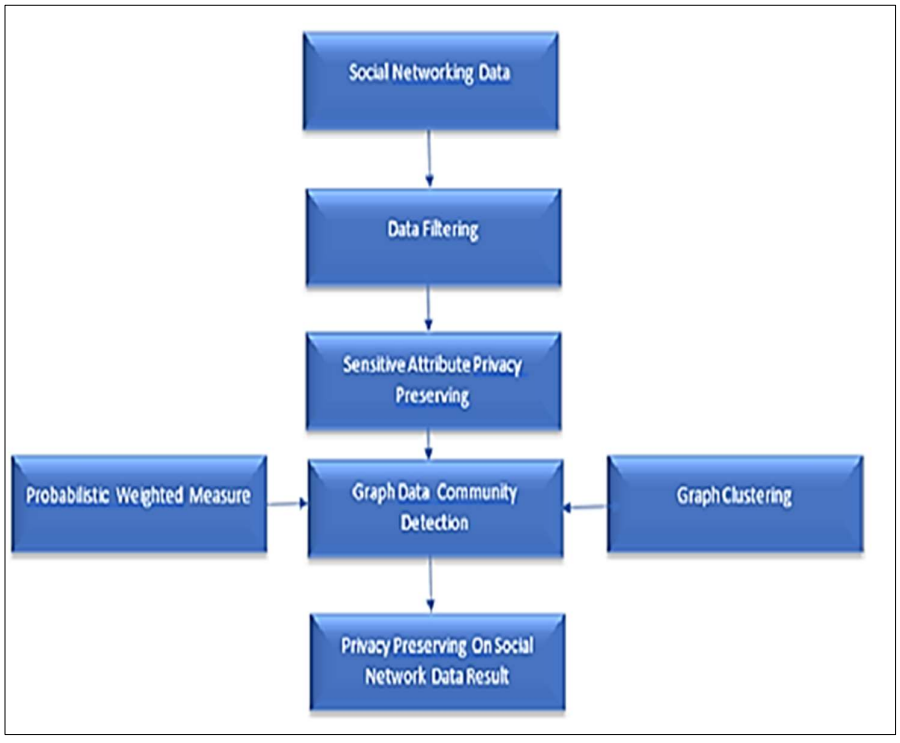

Figure 1. Proposed Framework.

DOI: https://doi.org/10.52756/ijerr.2024.v41spl.015

(φ), the attribute value is normalized using the following equation:

ατ[Π] = (ατ[Π] - (μχ(ατ) + μν(ατ)) / 2) / (μαχ_c(ατ) - μιν_c(ατ))

where:

- ατ[Π] is the attribute value of the current record.
- μχ(ατ) is the mean value of the attribute.
- μν(ατ) is another measure of central tendency (e.g., median).
- μαχ_c(ατ) and μιν_c(ατ) represent the maximum and minimum values of the attribute.

This equation ensures that all the continuous attribute values are on the same scale, which is essential in many machine-learning algorithms.

Categorical Attribute Encoding

For each categorical attribute (ατ[Π]) that is not empty (φ), the attribute value is encoded using the following equation:

ατ[Π] = Σ [Φ(ατ[ι] / κμ) - Φ(κμ)] / (Μχ * Prob(ατ[ι] / κμ))

where:

- ατ[Π] is the attribute value of the current record.
- Φ denotes a function (e.g., cumulative distribution function).
- κμ is a normalization constant.
- Prob(ατ[ι] / κμ) represents the probability of the attribute value given the normalization constant.

This process is simply encoding categorical data into a numerical format, yet it is still able to hold onto the statistical relationships of the data.

It is the pre-processing stage of OSN graph data, which helps for successive analysis and mining tasks by making sure to handle continuous and categorical attributes appropriately.

Algorithm 2: Weighted Probabilistic Community Detection

1. **Initialize an empty list for weighted nodes:**
   - Create an empty list named nodes_with_weight.
2. **Compute the centralized mean weighted measure for each node:**
   - For each node in the social networking graph, calculate the centralized mean weighted measure λ_α using the formula: $\lambda_\alpha = \frac{\mu_{\alpha 1} - \mu_{\alpha 2}}{2\sqrt{\min(\sigma_{\alpha 1}, \sigma_{\alpha 2})}}$ $\lambda_\alpha = \frac{\mu_{\alpha 1} - \mu_{\alpha 2}}{2\sqrt{\min(\sigma_{\alpha 1}, \sigma_{\alpha 2})}}$
   - To calculate $\mu_{\alpha 1}$ $\mu_{\alpha 1}$ and $\mu_{\alpha 2}$ $\mu_{\alpha 2}$: $\mu_{\alpha 1} = \text{CalculateAverageAttributeA1(node)} \mu_{\alpha 2} = \text{CalculateAverageAttributeA2(node)}$ $\mu_{\alpha 1} = \text{CalculateAverageAttributeA1(node)} \mu_{\alpha 2} = \text{CalculateAverageAttributeA2(node)}$
   - After computing λ_α, add the tuple (node, λ_α) to the nodes_with_weight list.
3. **Initialize an empty dictionary for community membership:**
   - Create an empty dictionary named community membership.
4. **Determine the community nodes and compute probabilistic measures:**
   - Make a new list as community nodes and initialize it for every community in the network..
   - In the social networking graph for each node, calculate the following probabilities:
     - The probability of the node's attribute A1 given the community, $\text{Prob(A1 | Cm)} \text{Prob(A1 | Cm)}$.
     - The probability of the node's attribute A2 given the community, $\text{Prob(A2 | Cm)} \text{Prob(A2 | Cm)}$.
   - Compute the maximized weighted probabilistic measure λ_β using the formula: $\lambda_\beta = \frac{\max(\text{Prob(A1|Cm)}, \text{Prob(A2|Cm)})}{2 \cdot |N| \cdot |Cm|}$ $\lambda_\beta = \frac{\max(\text{Prob(A1|Cm)}, \text{Prob(A2|Cm)})}{2 \cdot |N| \cdot |Cm|}$
   - Add the tuple (node, λ_β) to the community nodes list.
5. **Sort and identify community leaders:**
   - After processing all nodes, sort the list of community nodes from most to least of λ_β.
   - Find the highest node to serve as the head of the group.

Incorporate the community leaders role into the community membership.

Given the input dataset Δ and the output privacy-enabled dataset Δ^', the process involves the following steps:

1. Read and load the input dataset Δ
2. Apply the initial data filtering method on each numerical attribute in Δ.
3. Check for the presence of any sensitive attributes in the list Σα:
   - For each sensitive attribute identified:
   - Apply privacy-preserving transformations:
   - Utilize advanced data transformation techniques with adaptable permutation matrices.
   - Compute key parameters based on the provided secret key and permutation matrices.
   - Use bitwise XOR operations for enhanced data security.
4. Produce the privacy-enabled dataset Δ^' by performing anonymization techniques on Δ.

DOI: https://doi.org/10.52756/ijerr.2024.v41spl.015

5. Process each instance within the nearest neighbor groups:\
   - For each instance pair:
   - Determine distances using the Chebyshev metric.
   - From the sorted list of distances, identify the closest objects.
   - Use local density estimation methods to refine the object selection.
6. For every reducer in a MapReduce framework:
   - Identify the objects which are nearest using the probabilistic nearest neighbor approach.
   - Calculate relevant statistical values, including mean and distance metrics.
   - Estimate probabilities as needed.
   - Determine the local density for each case using unique equations.
   - Filter nearest neighbor objects based on local

Yelp dataset and connection to user profiles

	A	B	C	D	E	F	Formula Bar	H	I	J
1	business_id	date	review_id	stars	text	type	user_id	cool	useful	funny
2	9yKzy9PApeiPPOUJEtnvkg	26-01-2011	fWKvX83p0-ka4JS3dc6E5A	5	My wife	review	rLtl8ZkDX5vH5nAx9C3q5Q	2	5	0
3	ZRJwVLyzEJq1VAihDhYiow	27-07-2011	IjZ33sJrzXqU-0X6U8NwyA	5	I have no	review	0a2KyELOd3Yb1V6aivbluQ	0	0	0
4	6oRAC4uyJCsJl1X0WZpVSA	14-06-2012	lESLBzqUCLdSzSqm0eCSxQ	4	love the g	review	0hT2KtfLiobPvh6cDC8JQg	0	1	0
5	_1QQZuf4zZOyFCvXc0o6Vg	27-05-2010	G-WvGaISbqqaMHlNnByodA	5	Rosie,	review	uZetl9T0NcROGOyFfughhg	1	2	0
6	6ozycU1RpktNG2-1BroVtw	05-01-2012	1uJFq2r5QfJG_6ExMRCaGw	5	General	review	vYmM4KTsC8ZfQBg-j5MWkw	0	0	0
7	#NAME?	13-12-2007	m2CKSsepBCoRYWxiRUsxAg	4	Quiessen	review	sqYN3lNgvPbPCTRsMFu27g	4	3	1
8	zp713qNhx8d9KCJJnrw1xA	12-02-2010	riFQ3vxNpP4rWLk_CSri2A	5	Drop	review	wFweIWhv2fREZV_dYkz_1g	7	7	4
9	hW0Ne_HTHEAgGF1rAdmR	12-07-2012	JL7GXJ9u4YMx7RzsO5NfiQ	4	Luckily, I	review	1ieuYcKS7zeAv_U1SAB13A	0	1	0
10	wNUea3IXZWD63bbOQaO	17-08-2012	XtnfnYmnJYi71yIuGsXIUA	4	Definitely	review	Vh_DlizgGh5qQh4qfZ2h6A	0	0	0
11	nMHhuYan8e3cONo3Porn	11-08-2010	jIAIXA46pU1swYyRCdfXtQ	5	Nobuo shc	review	sUNkXg8-KFtCMQDV6zRzQg	0	1	0
12	AsSCv0q_BWqIe3mX2JqsO	16-06-2010	E11jzpKz9Kw5K7fuARWfRw	5	The	review	#NAME?	1	3	1
13	e9nN4XxjdHj4qtKCOPq_vg	21-10-2011	3rPt0LxF7rgmEUrznoH22w	5	Wonderful	review	C1rHp3dmepNea7XouwB6Q	1	1	0
14	h53YuCilDfEFSJCQpk8v1g	11-01-2010	cGnKNX3l9rthE0-TH24-qA	5	They	review	UPtysDF6cUDUxq2KY-6Dcg	1	2	0
15	WGNIYMeXPyoWav1APUq	23-12-2011	FvEEw1_OsrYdvwLV5Hrliw	4	Good tatt	review	Xm8HXE1JHqscXe5BKf0GFQ	1	2	0
16	yc5AH9H71xJidA_J2mChLA	20-05-2010	pfUwBKYYmUXeiwrhDluQcw	4	I'm 2	review	JOG-4G4e8ae3lx_szHtR8g	1	1	0
17	Vb9FPCEL6Ly24PNxLBaAFw	20-03-2011	HvqmdqWcerWO3Gs6zbrOw	2	Was it	review	ylWOj2y7TV2e3yYeWhu2QA	0	2	0
18	supigcPNO9lKo6olaTNV-g	12-10-2008	HXP_0Ul-FCmA4f-k9CqvaQ	3	We went	review	SBbftLzfYYKltOMFwOTIJg	3	4	2

1	JkeCKyEaQlbLd9uZYl4DjA::LiiLii C.::http://www.yelp.com/user_details?userid=JkeCKyEaQlbLd9uZYl4DjA
2	cs91PAsv6esdWAaSkzm2lg::Jan Ellen T.::http://www.yelp.com/user_details?userid=cs91PAsv6esdWAaSkzm2lg
3	cMgGj2FXHEbzdNZdLN_EwA::Saki U.::http://www.yelp.com/user_details?userid=cMgGj2FXHEbzdNZdLN_EwA
4	KXJbnHT4PDS1JZNCFKdmMg::stephanie h.::http://www.yelp.com/user_details?userid=KXJbnHT4PDS1JZNCFKdmMg
5	Tpmvufw1eea1DrjLAY2jLg::Theodore J.::http://www.yelp.com/user_details?userid=Tpmvufw1eea1DrjLAY2jLg
6	L22W35q3Ci3TytpA2LW34g::Doug H.::http://www.yelp.com/user_details?userid=L22W35q3Ci3TytpA2LW34g
7	xiJ6e1qN3Sqv4dS4D8CpNg::Amelia M.::http://www.yelp.com/user_details?userid=xiJ6e1qN3Sqv4dS4D8CpNg
8	pu96s510jutWeFOuofgY2g::Ty G.::http://www.yelp.com/user_details?userid=pu96s510jutWeFOuofgY2g
9	zcOlcoYhVgEgWxRpTVjUJA::Steve K.::http://www.yelp.com/user_details?userid=zcOlcoYhVgEgWxRpTVjUJA
10	_NH7Cpq3qZkByP5xR4gXog::Chris M.::http://www.yelp.com/user_details?userid=_NH7Cpq3qZkByP5xR4gXog
11	9YLxlEqpEjiugcA1NCbQnw::Christina W.::http://www.yelp.com/user_details?userid=9YLxlEqpEjiugcA1NCbQnw
12	PiSHysV8QdhgzU7QlRn1Kg::Cheri W.::http://www.yelp.com/user_details?userid=PiSHysV8QdhgzU7QlRn1Kg
13	PB3OGUgRSajFx18EBUwwEQ::Lord H.::http://www.yelp.com/user_details?userid=PB3OGUgRSajFx18EBUwwEQ
14	e9ATa_PlOWiyTS4QhdsJKA::Derek F.::http://www.yelp.com/user_details?userid=e9ATa_PlOWiyTS4QhdsJKA
15	urQ0TUF3uhc-oOHzb3F5tQ::Yuan D.::http://www.yelp.com/user_details?userid=urQ0TUF3uhc-oOHzb3F5tQ
16	DUJLlGEHUerY8feKhjO85A::Ryan H.::http://www.yelp.com/user_details?userid=DUJLlGEHUerY8feKhjO85A
17	ti5uWqAxf7pOkKs7csMlDA::Tiffany R.::http://www.yelp.com/user_details?userid=ti5uWqAxf7pOkKs7csMlDA
18	JssCk4of-CQ7j81ZLrZMZg::Jennifer C.::http://www.yelp.com/user_details?userid=JssCk4of-CQ7j81ZLrZMZg
19	dMEMCWkbmi2hzr24_9J-ZA::Michael P.::http://www.yelp.com/user_details?userid=dMEMCWkbmi2hzr24_9J-ZA
20	Y_v3O9q_-vSOk25cOEjlkQ::Dan S.::http://www.yelp.com/user_details?userid=Y_v3O9q_-vSOk25cOEjlkQ

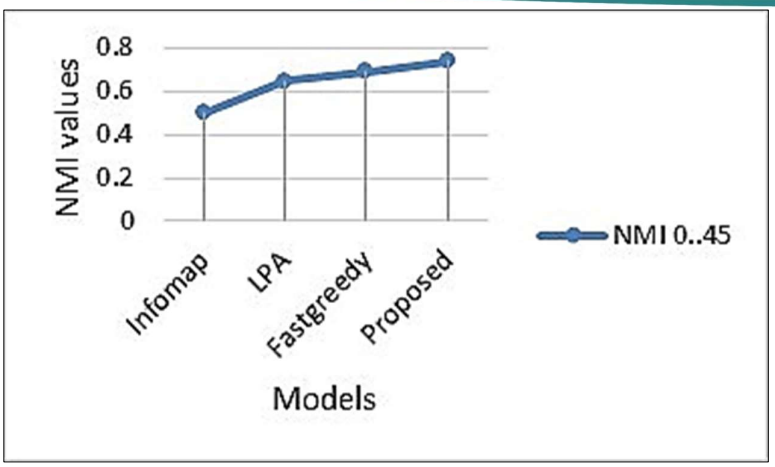

Figure 2. Performance Comparison on Yelp data.

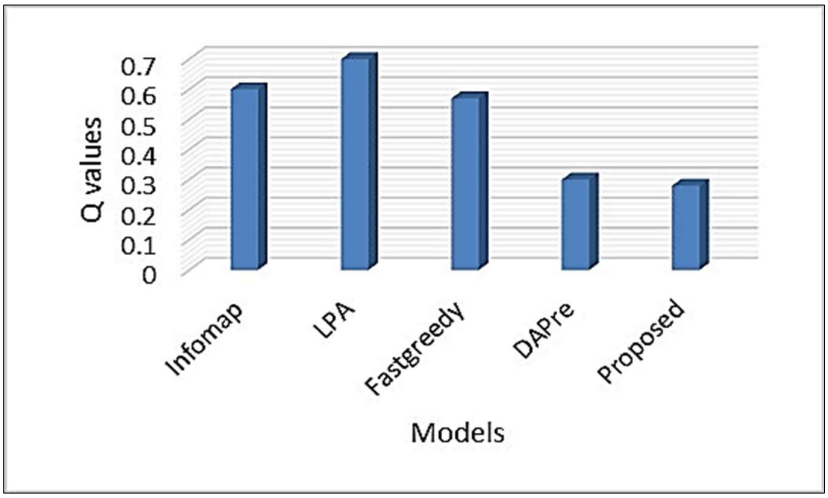

Figure 3. compares the Q value of the suggested model to dolphin data.

kernel density estimation.

7. Finalize the process.

Results and Discussion

The presented experimental results were simulated on the Java programming environment, using our third-party graph and similarity libraries to achieve the goal. Four different datasets were used to analyze the performance of the proposed model. These are Yelp, Football, Zachary and Dolphin dataset. The results were analyzed with Q, NMI, VI, and RI. These estimation metrics were used for analyzing the results obtained from the training dataset.

Here is how the Q metric is calculated:

$$Q = \sum (e_{ii} - a_i^2)$$

The NMI between P and Q is calculated as:

$$NMI(P|Q) = \frac{e(P) + e(Q) - e(P,Q)}{(e(P) + e(Q))/2}$$

Here, the X and Y are the original and the predicted communities. $e(P)e(P)$ and $e(Q)e(Q)$ are the entropy values of the corresponding communities.

Figure 2 plots the result of applying our probabilistic model to the Yelp dataset and comparing it to different traditional models. We can see that this proposed model works much more efficiently than traditional models with respect to Yelp. Here, the value of NML represents a measure of the quality of the Yelp dataset both inter-community and intra-community detection.

Figure 3 compares to Figure 2, but specifically for the Dolphin dataset. It also represent that our suggested model outperforms standard models in terms of Q value on Dolphin dataset. Figure 4, shown below, compares our model's Q value with other existing models in terms of the Football dataset.

From Figure 4 we can see that the proposed Q value is more efficient than traditional models, having better performance in discovering intracommunity and intercommunity communities over the Football dataset. Table 1 shows how well the proposed model counts local patterns across Yelp data samples compared to standard methods. Table 2 shows how the proposed model counts local patterns compared to standard models throughout Football dataset.

DOI: https://doi.org/10.52756/ijerr.2024.v41spl.015

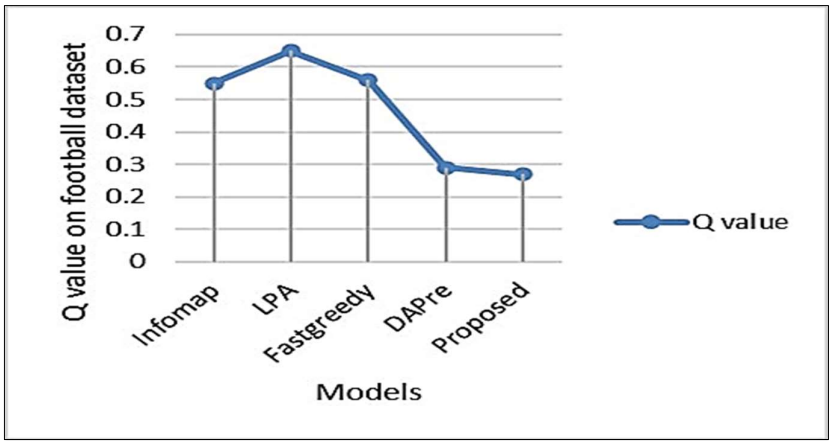

Figure 4. compares the Q value of suggested model to Football dataset

Table 1. Proposed Model Performance on Yelp Dataset.

Test	LPA	Infomap	DAPre	Fast greedy	Proposed
1	21075	11737	20336	13190	29259
2	16547	15637	22145	10693	26581
3	17890	11500	17979	14861	29772
4	19837	12479	11269	13801	29617
5	13389	19570	10708	14716	27752
6	14327	15895	13322	17124	28896
7	17339	18648	18318	10278	29383
8	18219	21102	22273	18177	25866
9	13253	15002	23572	13552	25064
10	13356	12695	11646	11550	26852
11	12328	17531	14017	20366	26774
12	17422	15066	15169	11627	25522
13	19719	18763	12624	18379	29506
14	14149	10375	18199	23866	27310
15	20603	12114	16179	10821	25657

Table 2. Proposed Model Performance on Football Dataset.

Test	LPA	Infomap	DAPre	Fast greedy	Proposed
1	5739.011	5552.53	5255.61	5799.24	3854.64
2	5728.31	5922.76	5189.04	5334.34	3951.82
3	5330.94	4743.78	5465.60	5790.27	4108.85
4	5122.21	5317.62	5395.72	5124.76	3848.57
5	5097.51	5501.61	4582.05	5709.85	4144.79
6	5553.19	5111.42	4683.31	5277.69	3848.45
7	4871.26	5291.48	4750.42	4972.49	3845.19
8	5099.17	5518.15	5678.79	5731.41	3854.74
9	5616.19	4617.71	5476.26	5353.48	4014.18
10	5253.88	5743.15	5240.42	5428.42	4201.14
11	4715.14	5043.28	5556.72	5166.29	3898.91
12	4928.71	5378.38	5120.36	5712.82	3800.67
13	4913.88	4564.85	4570.60	5330.68	3812.51
14	4599.89	5631.24	4683.97	5583.48	4188.79
15	4981.05	5076.56	5342.90	4824.76	4246.71

DOI: https://doi.org/10.52756/ijerr.2024.v41spl.015

Figure 5 shows how the proposed model and standard models (DAPre, Infomap, LPA, Fastgreedy) perform in terms of local pattern counts across different test cases using the Twitch data samples.

Figure 6 compares the local pattern counts of the proposed model with standard models throughout different test cases using the Twitch dataset. Both Figures 5 and 6 plot a visual representation of the performance of the proposed model compared to other standard models on the Twitch OSN dataset. The Twitch Social Network Dataset contains the social network of users from the Twitch platform, a popular live-streaming service focused on video game live streaming. The dataset includes nodes representing users and edges representing friendships between them. The main dataset features are as given below

- **Nodes**: Users on the Twitch platform.
- **Edges**: Friendships between the users.
- **Attributes**:
  - user_id: Unique identifier for each individual user.
  - friend_id: Unique identifier for each friend of the user.

creation_date: Date on which the friendship was established.

Conclusion

Traditional approaches to protecting privacy in social networking datasets are based more on data-perturbation techniques than on data-transformation techniques because the volume of social networking data to be processed in real-time requires huge computational memory and time to perform intensive, fine-grained operations. Likewise, most of the conventional privacy-preserving methods rely on preset metrics and static notions of communities based on structural concepts extracted from social networks. We propose a probabilistic framework for community detection that can be cast into the filtering framework. The proposed algorithm extends standard community detection approaches in two ways. First, our approach in the Bayesian network includes dynamic metrics, leading to a fast computation. Second, our framework based on filtering yields NMI and Q rates higher than the corresponding ones from related methods, as illustrated through our experiments. Our future work includes optimizing meta-heuristics for both global and local search methods expanding the search space within the privacy-preserving procedure.

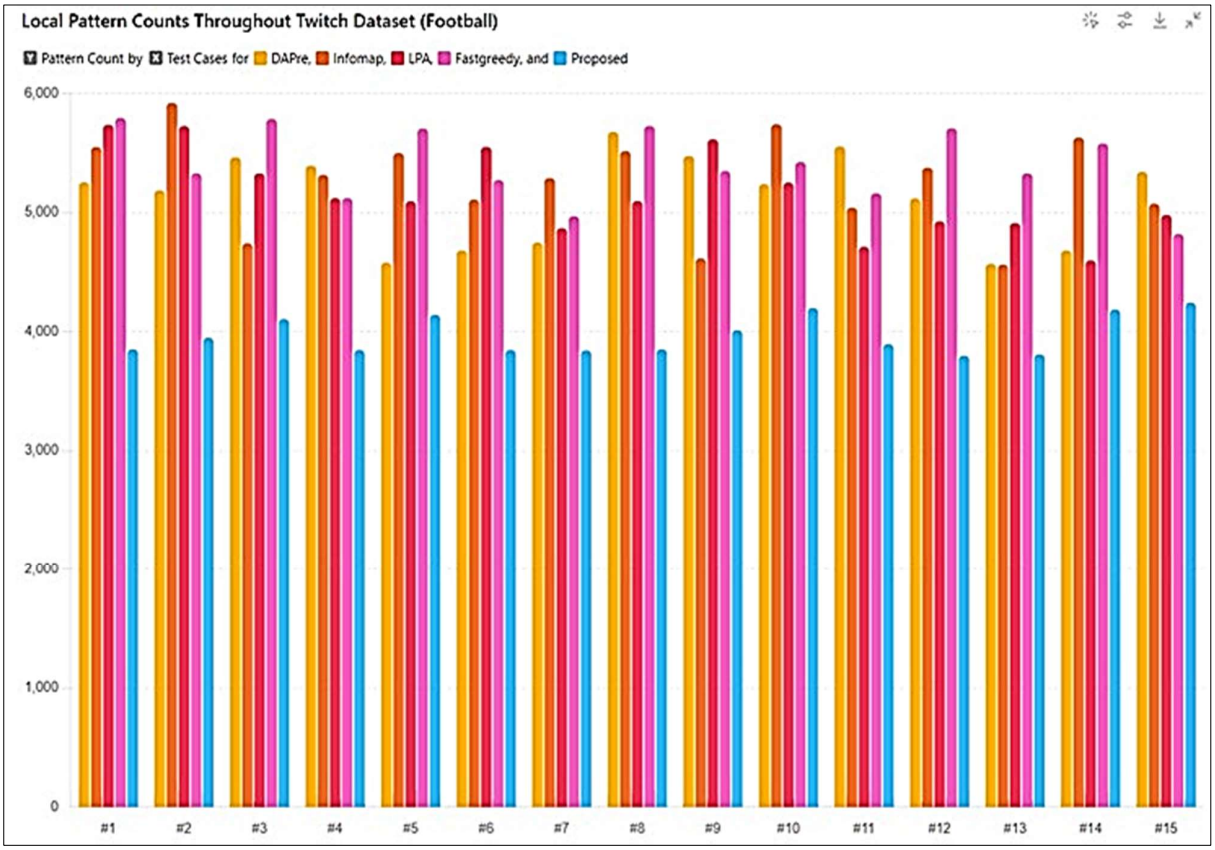

Figure 5. Local Pattern Counts across Twitch Data Samples.

DOI: https://doi.org/10.52756/ijerr.2024.v41spl.015

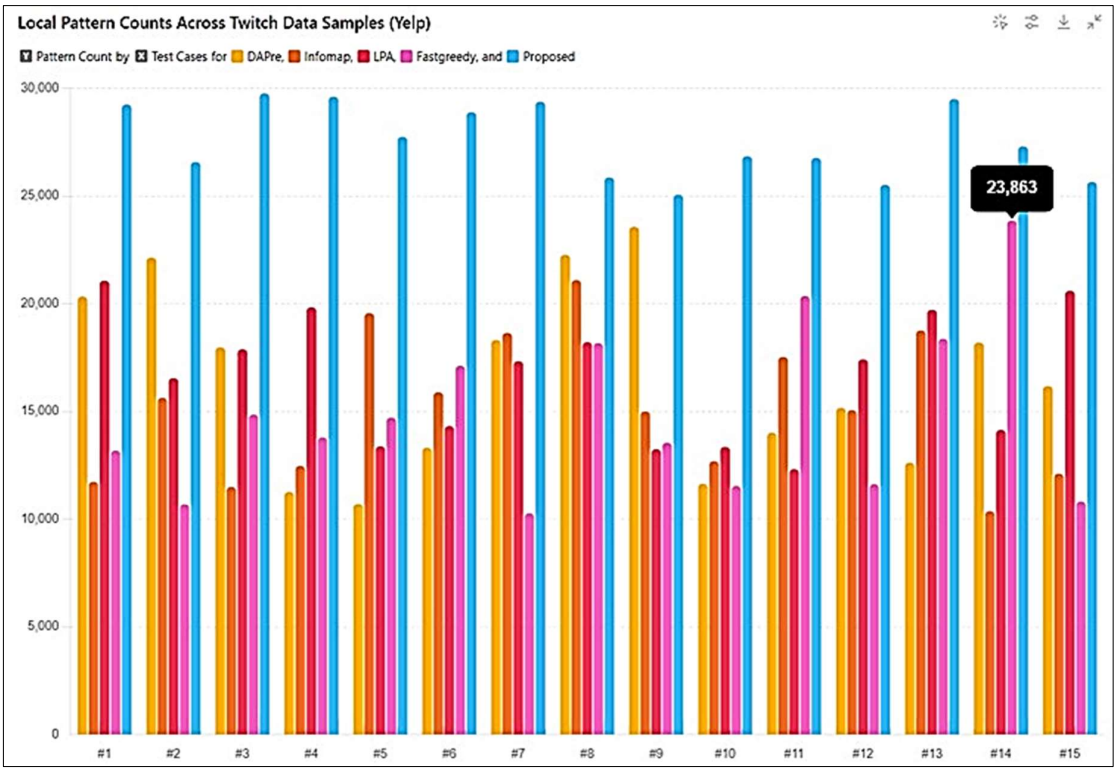

Figure 6. Local Pattern Counts throughout Twitch Dataset.

Conflict of Interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Aghaalizadeh, S., Afshord, S. T., Bouyer, A., & Anari, B. (2021). A three-stage algorithm for local community detection based on the high node importance ranking in socialnet works. *Physica. A: Statistical Mechanics and its Applications, 563*, 125420. https://doi.org/10.1016/j.physa.2020.125420

Bahri, L., Carminati, B., & Ferrari, E. (2018). Decentralized privacy preserving services for Online Social Networks. *Online Social Networks and Media, 6*, 18–25. https://doi.org/10.1016/j.osnem.2018.02.001

Bandara, E., Liang, X., Foytik, P., Shetty, S., Hall, C., Bowden, D., Ranasinghe, N., & De Zoysa, K. (2021). A blockchain empowered and privacy preserving digital contact tracing platform. *Information Processing & Management, 58*(4), 102572. https://doi.org/10.1016/j.ipm.2021.102572

Barsocchi, P., Calabrò, A., Crivello, A., Daoudagh, S., Furfari, F., Girolami, M., & Marchetti, E. (2021). COVID-19 & privacy: Enhancing of indoor localization architectures towards effective social distancing. *Array, 9*, 100051. https://doi.org/10.1016/j.array.2020.100051.

Beg, S., Anjum, A., Ahmad, M., Hussain, S., Ahmad, G., Khan, S., & Choo, K. K. R. (2021). A privacy-preserving protocol for continuous and dynamic data collection in IoT enabled mobile app recommendation system (MARS). *Journal of Network and Computer Applications,174*, 102874. https://doi.org/10.1016/j.jnca.2020.102874

Bourahla, S., Laurent, M., & Challal, Y. (2020). Privacy preservation for social networks sequential publishing. *Computer Networks, 170*, 107106. https://doi.org/10.1016/j.comnet.2020.107106

Daud, N. N., Hamid, S. H. A., Saadoon, M., Sahran, F., & Anuar, N. B. (2020). Applications of link prediction in social networks: A review. *Journal of Network and Computer Applications, 166*, 102716. https://doi.org/10.1016/j.jnca.2020.102716

Gupta, B., Sangaiah, A., Nedjah, N., Yamaguchi, S., Zhang, Z., & Sheng, M. (2018). Recent research in computational intelligence paradigms into security and privacy for online social networks (OSNs). *Future Generation Computer Systems, 86*, 851–54. https://doi.org/10.1016/j.future.2018.05.017

Jain, A. K., Sahoo, S. R., & Kaubiyal, J. (2021). Online social networks security and privacy: comprehensive

review and analysis. *Complex & Intelligent Systems, 7*(5), 2157–2177. https://doi.org/10.1007/s40747-021-00409-7

Jha, K., Jain, A., & Srivastava, S. (2024). A Secure Biometric-Based User Authentication Scheme for Cyber-Physical Systems in Healthcare. *International Journal of Experimental Research and Review*, *39*(Spl Volume), 154-169. https://doi.org/10.52756/ijerr.2024.v39spl.012

Kahate, S. A., & Raut, A. D. (2022). Comprehensive Analysis of Privacy Attacks in Online Social Network: Security Issues and Challenges. *International Journal of Safety and Security Engineering, 12*(4), 507–518. https://doi.org/10.18280/ijsse.120412

Kantarcioglu, M., & Clifton, C. (2004). Privacy-preserving distributed mining of association rules on horizontally partitioned data. *IEEE Transactions on Knowledge and Data Engineering, 16*(9), 1026–1037. https://doi.org/10.1109/tkde.2004.45

Kavianpour, S., Tamimi, A., & Shanmugam, B. (2019). A privacy-preserving model to control social interaction behaviors in social network sites. *Journal of Information Security and Applications, 49*, 102402. https://doi.org/10.1016/j.jisa.2019.102402

Kayes, I., & Iamnitchi, A. (2017). Privacy and security in online social networks: A survey. *Online Social Networks and Media*, 3–4, 1–21. https://doi.org/10.1016/j.osnem.2017.09.001

Keerthana, B., Vana, T. R., Rao, M. S., Sambana, B., & Mishra, P. (2023). Using CNN technique and webcam to identify face mask violation. In *Springer proceedings in mathematics & statistics,* pp. 245–254. https://doi.org/10.1007/978-3-031-15175-0_20

Kumar, C., Bharti, T. S., & Prakash, S. (2023). A hybrid Data-Driven framework for Spam detection in Online Social Network. *Procedia Computer Science, 218*, 124-132. https://doi.org/10.1016/j.procs.2022.12.408

Li, X., Gu, Y., Dvornek, N., Staib, L. H., Ventola, P., & Duncan, J. S. (2020). Multi-site fMRI

analysis using privacy-preserving federated learning and domain adaptation: ABIDE results. *Medical Image Analysis, 65,* 101765. https://doi.org/10.1016/j.media.2020.101765

Liu, P., Xu, Y., Jiang, Q., Tang, Y., Guo, Y., Wang, L. E., & Li, X. (2020). Local differential privacy for social network publishing. *Neurocomputing, 391*, 273–279. https://doi.org/10.1016/j.neucom.2018.11.104.

Madhuri, T. N. P., Rao, M. S., Santosh, P. S., Tejaswi, P., & Devendra, S. (2022). Data Communication Protocol using Elliptic Curve Cryptography for Wireless Body Area Network. *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*. https://doi.org/10.1109/iccmc53470.2022.9753898

Nicolazzo, S., Nocera, A., Ursino, D., & Virgili, L. (2020). A privacy-preserving approach to prevent feature disclosure in an IoT scenario. *Future Generation Computer Systems, 105*, 502–519. https://doi.org/10.1016/j.future.2019.12.017

Pensa, R. G., & Di Blasi, G. (2017). A privacy self-assessment framework for online Social networks. *Expert Systems With Applications, 86*, 18–31. https://doi.org/10.1016/j.eswa.2017.05.054

Prasad, K. L., Anusha, P., Rao, M., & Rao, K. (2019). A Machine Learning based Preventing the Occurrence of Cyber Bullying Messages on OSN. *International Journal of Recent Technology and Engineering*, *8*(3), 1861–1865. https://doi.org/10.35940/ijrte.a9164.078219

Rao, M.S., Uma Maheswaran, S.K., Sattaru, N.C., Abdullah, K.H., Pandey, U.K., & Biban, L. (2022). A Critical Understanding of Integrated Artificial Intelligence Techniques for the Healthcare Prediction System. *2nd International Conference on Advance Computing and Innovative Technologies in Engineering* (ICACITE), Greater Noida, India, *2022*, pp. 728-731. https://doi: 10.1109/ICACITE53722.2022.9823678.

Sai, A. M. V. V., & Li, Y. (2020). A Survey on Privacy Issues in Mobile Social Networks. *IEEE Access, 8*, 130906–130921. https://doi.org/10.1109/access.2020.3009691

Singh, D., & Singh, S. (2023). Precision fault prediction in motor bearings with feature selection and deep learning. *Int. J. Exp. Res. Rev.*, 32, 398-407. https://doi.org/10.52756/ijerr.2023.v32.035

Sun, G., Song, L., Liao, D., Yu, H., & Chang, V. (2019). Towards privacy preservation for "check-in" services in location-based social networks. *Information Sciences, 481*, 616– 634. https://doi.org/10.1016/j.ins.2019.01.008

Tran, A. T., Luong, T. D., Karnjana, J., & Huynh, V. N. (2021). An efficient approach for privacy preserving decentralized deep learning models based on secure multi-party computation.*Neurocomputing,422*, 245–262.

Wei, J., Lin, Y., Yao, X., & Sandor, V. K. A. (2019). Differential privacy-based trajectory community recommendation in social network. *Journal of Parallel and Distributed Computing, 133*, 136–148.

https://doi.org/10.1016/j.jpdc.2019.07.002

Xiao, X., Chen, C., Sangaiah, A. K., Hu, G., Ye, R., & Jiang, Y. (2018). CenLocShare: A centralized privacy-preserving location-sharing system for mobile online social networks. *Future Generation Computer Systems, 86*, 863–872. https://doi.org/10.1016/j.future.2017.01.035

Yang, J., Fu, C., & Lu, H. (2021). Optimized and federated soft-impute for privacy-preserving tensor completion in cyber-physical-social systems. *Information Sciences, 564*, 103–123. https://doi.org/10.1016/j.ins.2021.02.028

Zareie, A., & Sakellariou, R. (2020). Similarity-based link prediction in social networks using

latent relationships between the users. *Scientific Reports, 10*(1). https://doi.org/10.1038/s41598-020-76799-4

Zhang, J., Zhao, B., Song, G., Ni, L., & Yu, J. (2019). Maximum delay anonymous clustering feature tree based privacy-preserving data publishing in social networks. *Procedia Computer Science, 147*, 643–46. https://doi.org/10.1016/j.procs.2019.01.190

Zhao, Y., Tarus, S. K., Yang, L. T., Sun, J., Ge, Y., &Wang, J. (2020). Privacy-preserving clustering for big data in cyber-physical-social systems: Survey and perspectives. *Information Sciences, 515*, 132–155.https://doi.org/10.1016/j.ins.2019.10.019.

Zheng, X., Cai, Z., Luo, G., Tian, L., & Bai, X. (2019) Privacy-preserved community discovery in online social networks. *Future Generation Computer Systems, 93*, 1002–1009. https://doi.org/10.1016/j.future.2018.04.020

Zhou, J., Cao, Z., Dong, X., Xiong, N., & Vasilakos, A. V. (2015). 4S: A secure and privacy-preserving key management scheme for cloud-assisted wireless body area network in m-healthcare social networks. *Information Sciences, 314,* 255–276. https://doi.org/10.1016/j.ins.2014.09.003

How to cite this Article:

Shamila.M, G. Rekha and K. Vinuthna Reddy (2024). Hybrid Filtering and Probabilistic Techniques for Privacy-Preserving Community Detection in OSNs. *International Journal of Experimental Research and Review*, *41*(spl.), 180-194.

DOI : https://doi.org/10.52756/ijerr.2024.v41spl.015

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.