Knowledge and Information Systems (2025) 67:29–108
https://doi.org/10.1007/s10115-024-02289-y

REVIEW

A comprehensive bibliometric analysis on social network anonymization: current approaches and future directions

Navid Yazdanjue[1] · Hossein Yazdanjouei[2] · Hassan Gharoun[1] · Mohammad Sadegh Khorshidi[1] · Morteza Rakhshaninejad[3] · Babak Amiri[3] · Amir H. Gandomi[1,4]

Received: 24 October 2023 / Revised: 6 October 2024 / Accepted: 7 November 2024 /
Published online: 3 January 2025

Abstract
In recent decades, social network anonymization has become a crucial research field due to its pivotal role in preserving users' privacy. However, the high diversity of approaches introduced in relevant studies poses a challenge to gaining a profound understanding of the field. In response to this, the current study presents an exhaustive and well-structured bibliometric analysis of the social network anonymization field. To begin our research, related studies from the period of 2007–2022 were collected from the Scopus Database and then preprocessed. Following this, the VOSviewer was used to visualize the network of authors' keywords. Subsequently, extensive statistical and network analyses were performed to identify the most prominent keywords and trending topics. Additionally, the application of co-word analysis through SciMAT and the Alluvial diagram allowed us to explore the themes of social network anonymization and scrutinize their evolution over time. These analyses culminated in an innovative taxonomy of the existing approaches and anticipation of potential trends in this domain. To the best of our knowledge, this is the first bibliometric analysis in the social network anonymization field, which offers a deeper understanding of the current state and an insightful roadmap for future research in this domain.

Keywords Social network · Anonymization · Privacy preservation · Bibliometric analysis · Data publishing

1 Introduction

Over the past century, social network applications have experienced exponential growth, fundamentally transforming how we communicate, share information, and connect with others. With millions of users worldwide, these platforms have become integral to our daily lives. They are used for a wide variety of purposes, ranging from staying in touch with friends and

Hossein Yazdanjouei, Hassan Gharoun, Mohammad Sadegh Khorshidi and Morteza Rakhshaninejad have contributed equally to this work.

Extended author information available on the last page of the article

family, to business networking, accessing news and entertainment, and even political engagement and social activities. In addition to their vast user bases, social network applications have also evolved in terms of complexity and sophistication, now offering features such as live streaming, eCommerce, virtual reality, and more. Furthermore, with the advent of technologies like Artificial Intelligence (AI) and Machine Learning (ML), social networks have become more personalized and interactive, enhancing user experience while significantly expanding the amount of personal data that is collected and processed.

However, the rapid growth and increasing complexity of social network applications have also raised significant privacy concerns. Given the volume and sensitivity of data shared on these platforms, ensuring user privacy has become paramount [60, 81, 208]. Privacy threats, such as identity exposure, relationship mapping, and attribute disclosure, emerge when social network data is made publicly accessible or shared with third parties for research or commercial purposes. To address these risks, social network anonymization has emerged as a key solution. Anonymization techniques aim to protect users by masking their identities, relationships, or specific attributes, while preserving the utility of the data for analysis.

With the exponential growth of social networks and the abundance of user data, there has been a surge of scholarly interest in anonymization techniques specific to these platforms. Academics, researchers, and developers have been focusing on creating and enhancing anonymization techniques for social networks, recognizing the urgency to reconcile the dual demands of data utility and privacy protection. This high interest is evidenced by the increasing number of publications related to this topic in esteemed academic journals and conferences. Researchers are continually proposing novel anonymization techniques to keep up with the evolving landscape of social network applications. A key aspect of these research studies is to find a balance between ensuring user privacy and maintaining data utility. While the primary objective is to protect user privacy, it is equally important to ensure that anonymized data remain useful for research and business purposes. Hence, many proposed techniques focus on generating synthetic data or modifying social network graphs in a way that anonymizes user identities while preserving the overall network characteristics.

In recent years, several surveys and bibliometric studies in adjacent fields, such as privacy-preserving techniques in data mining and computer science [3, 80, 128], have provided valuable insights. However, these studies often address a broader range of privacy challenges, lacking the depth necessary for understanding the unique challenges of social network anonymization. For instance, a recent bibliometric analysis of privacy-preserving data mining techniques identified key research trends and influential works but did not account for the complexities of anonymizing social networks. Social networks, unlike traditional datasets, involve intricate user interactions and relationships, which introduce distinct privacy risks. Thus, anonymization techniques for social networks must protect not just individual data points, but also the relationships between users, making the process far more complex.

Recognizing the importance of social network anonymization in protecting user privacy, this study focuses exclusively on this domain. Our primary goal is to analyze and categorize existing research, particularly studies that propose anonymization techniques tailored to social network privacy. Additionally, to the best of our knowledge, the lack of an existing bibliometric review in the field of social network anonymization, combined with the strengths of bibliometric analysis, motivated us to undertake this study and contribute to the growing body of research in this domain. By mapping key trends and developments, our study could provide valuable insights for scholars and industry professionals, facilitating the development of more effective anonymization techniques suited to the unique challenges of social networks.

In this direction, we first extracted publications from 2007, the year that this concept emerged, to 2022 from the open-source Elsevier Developer Portal using Scopus API (Elsevier 2022). From the extracted articles, a total of 315 relevant articles were selected for inclusion in this study. Then, the authors' keywords were preprocessed and visualized using network visualization tools. Afterward, we conducted two primary analyses, namely statistical measures and network analysis, along with a co-word analysis. These were performed to detect the themes and topics and study their evolution, interrelations, and trends within the social network anonymization domain. Ultimately, the findings allowed us to categorize the popular approaches from the inception of this field of study into a novel taxonomy.

The rest of this study is organized as follows. Section 2 delivers an overview of the associated surveys and studies that have been conducted in the realm of social network anonymization. In Sect. 3, we delve into the methodology applied in this review and explain the process of gathering pertinent papers as well as the preprocessing phase. The statistical measures and analysis of the network are also presented. Moving forward, the co-word analysis and trajectory of various themes and topics' evolution are provided. At the end of this section, the prevalent approaches utilized in the field of social network anonymization are categorized based on the conducted analyses. Section 4 discusses the emerging research trends and areas of potential interest in the field of social network anonymization. Consequently, Sect. 5 wraps up the paper with concluding remarks.

2 Related works

In recent years, the field of social network anonymization has been a focal point of intensive research, leading to a substantial and ever-growing body of literature dedicated to the domain. This extensive research activity reflects the critical importance of privacy in our increasingly interconnected world. Several surveys and literature reviews have been conducted, providing a comprehensive overview of the various anonymization techniques, their evolution, and their effectiveness in different social network types. Such studies have played a significant role in advancing our understanding of the challenges and complexities involved in social network anonymization. They have also helped identify areas that require further exploration. These outstanding surveys and literature review studies are elaborated on in the following paragraphs.

In 2008, Zhou et al. conducted a comprehensive review of existing techniques for social network anonymization [210]. They performed an organized examination of the methods employed to maintain user privacy during the sharing or publication of such data. They provided a systematic overview of the field, offering insights into the strengths and weaknesses of various approaches, their applicability to different types of data, and their effectiveness in maintaining a balance between privacy and data utility. Also, the authors recognized the challenges in privacy preservation methods within social network data, especially when compared to the traditional relational data cases that have been extensively studied. The analyses of social network anonymization methods were focused on three key aspects: preserving privacy, understanding the background knowledge available to the adversary, and maintaining data utility, which is the value of the data for further use or analysis. To better structure their review, the authors categorized the existing anonymization methods into two principal types: clustering-based approaches and graph modification approaches. Clustering-based approaches function by grouping nodes (individuals or entities) and their edges (connections) into larger collective units known as super-nodes and super-edges, respectively.

Each of these larger units (super-nodes and super-edges) is then subjected to an anonymization process, effectively hiding individual identities within the clusters. The clustering-based approach is further divided into the following four subcategories.

- Vertex Clustering: This method groups nodes, or vertices, based on some similarity metric. The similarity might be based on the attributes of the nodes or on the structure of the graph like nodes having similar connectivity [58, 59]. Once the nodes are grouped into clusters, they can be replaced with a single super-node to anonymize the individual nodes in each cluster. For each super-node, two features are represented, namely the number of nodes and the number of edges within the cluster, to maintain the utility of the anonymized network.
- Edge Clustering: In this method, the edges are the main focus to preserve the sensitive relationships [208]. Edges with similar properties (such as weight or type of relationship) are grouped into clusters. Similar to vertex clustering, these grouped edges can be replaced with a super-edge, anonymizing the original individual relationships. Also, the number of edges within each super-edge will be represented to maintain the utility of the network.
- Vertex and Edge Clustering: This approach, also called generalization, combines both vertex and edge clustering [17], grouping both nodes and edges. It provides a more comprehensive anonymization, as it anonymizes both individual identities (nodes) and their relationships (edges). Once nodes and edges are grouped into clusters, they can be replaced with super-nodes and super-edges, effectively hiding individual information within the network. More precisely, the process of anonymizing node attributes utilizes a generalization technique, which is extensively researched in the context of relational data. For structural anonymization, this method utilizes edge generalization, which is similar to the method outlined by [208]. However, a significant distinction exists in this method, as it incorporates both the loss of generalization information loss and structural information loss during the clustering process.
- Vertex-Attribute Mapping Clustering: This method was originally used for anonymizing the bipartite graphs [34, 35]. When publishing a bipartite graph, the structure of the graph is preserved. The nodes are organized into clusters, and the relationship or mapping between these clusters in the original graph and those in the published graph is made publicly available. This process enables the anonymization of individual nodes while maintaining the overall structure and relationships of the graph for analysis or research purposes. Consequently, it is necessary to carefully construct the mapping between the clusters. This helps ensure that the anonymization process is effective and maintains the integrity of the structural relationships present in the original graph.

On the other hand, graph modification approaches alter the structure of the network graph to conceal individual identities while preserving the overall characteristics of the network. The authors also considered three subcategories for the graph modification approach, which are subsequently defined.

- Optimization Graph Construction: This method involves constructing a new optimized version of the original graph with a new degree sequence that is K-degree anonymous, maintaining the overall structure while ensuring that individual identities are concealed [95]. Specifically, the main idea behind K-degree is to modify the original network so that each node in the network has the same degree as at least "K-1" other nodes.
- Randomized Graph Modification: This method introduces randomness into the graph modification process to ensure anonymity. They may randomly add, delete, or modify nodes and edges within the network while attempting to preserve the overall structure and characteristics. The randomness introduced by these methods makes it more difficult to de-anonymize the data, thus providing an additional layer of privacy protection. Zhou et al. categorized the

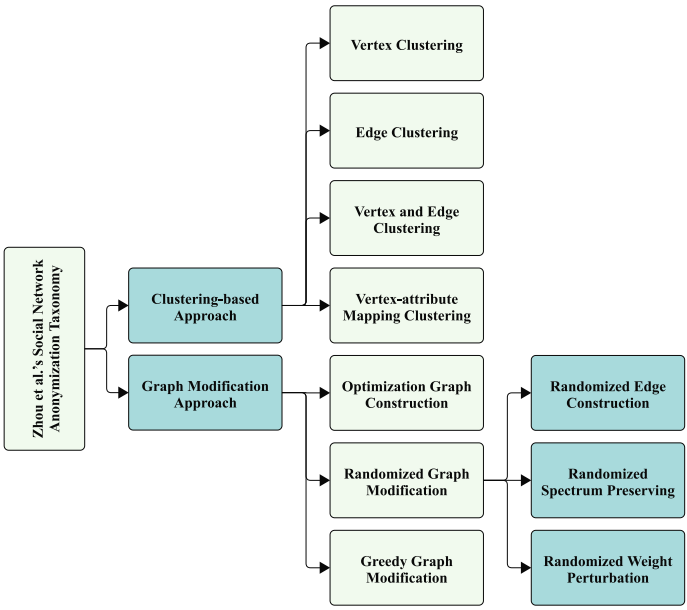


Fig. 1 Social network anonymization taxonomy introduced by Zhou et al. [210]

introduced techniques in this method into three general categories, including randomized edge construction, randomized spectrum preserving, and randomized weight perturbation.

- Greedy Graph Modification: This method iteratively modifies the graph to provide the K-anonymity-based models, which results in protecting individual identities. At each iteration, the method makes the modification that provides the greatest immediate benefit according to a specific objective, such as maximizing privacy or minimizing information loss. While these methods can be simpler and faster than other methods, they may not always provide the optimal solution, as the modifications are done based on immediate gains rather than long-term optimization.

The social network anonymization taxonomy proposed by Zhou et al. is presented in Fig. 1.

In 2010, Wu et al. investigated a review of the advancements in research concerning the privacy-preserving publishing of graph and social network data (X. [179, 180]. Their analysis is oriented toward understanding how to anonymize and publish such data while maintaining user privacy. Besides, the authors divided the anonymization strategies for simple graphs into three main categories. The first category is K-anonymity-based privacy preservation via edge modification, which involves altering edges within the graph in a way that at least "K" nodes share similar identifiable characteristics, thereby preserving privacy. In this category, they consider three subcategories, including K-degree generalization, K-neighborhood anonymity, and K-automorphism anonymity, which are subsequently described.

In the K-degree generalization or K-degree anonymity model, the structure of the graph is modified in such a way that at least "K" nodes share the same degree. Also, the K-neighborhood anonymity focuses on the neighborhood of each node. A node's neighborhood in a graph is all the nodes it is directly connected to. K-neighborhood anonymity ensures that each node shares the same neighborhood with at least "K-1" other nodes. This means that

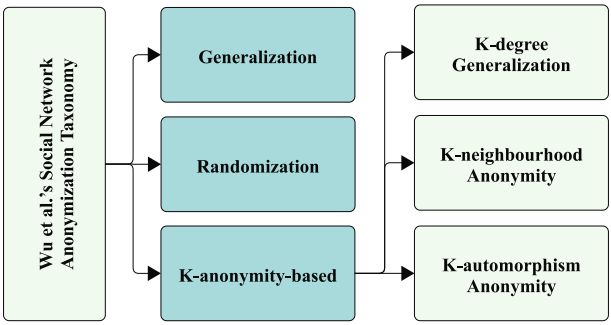


Fig. 2 Social network anonymization taxonomy proposed by [180]

for each node, there are at least "K" nodes in the graph, including itself, that are connected to the same set of nodes. Besides, an automorphism in graph theory is an isomorphism from a graph to itself. It essentially means a reordering of the nodes of the graph in such a way that the new graph is exactly the same as the original. In the context of the K-anonymity model, K-automorphism anonymity means that each node belongs to a set of at least "K" nodes that are indistinguishable from each other when considering the entire structure of the graph.

The second category is probabilistic privacy preservation via randomization. In this approach, the connections between nodes are randomized to add a level of uncertainty, effectively anonymizing the data while still preserving the overall structure and characteristics of the graph. The authors explored two edge-based randomization strategies that are frequently used in social networks: random addition/deletion and random switch. The random addition/deletion approach involves adding or deleting edges at random, while the random switch approach involves randomly switching a pair of existing edges.

The third category is privacy preservation via generalization which is identical to the clustering-based approach. Following the categorization of simple graph anonymization methods, the authors turned their attention to rich graphs that are more complex and contain additional information, like edge direction, edge weights, or additional attributes on the nodes or edges. Since anonymizing these graphs requires more sophisticated methods, the authors reviewed current approaches for handling this increased complexity. Their proposed social network anonymization taxonomy is illustrated in Fig. 2.

In 2016, Abawajy et al. provided an overview of the recent advancements in techniques for social network anonymization when releasing social network data publicly, as well as the challenges and potential research directions [1]. They also covered various privacy threats and attacks that adversaries might use to exploit anonymized data from social networks.

The authors of this work organized the methods for anonymizing social network data into two primary groups: Non-perturbation Privacy Preservation Models and Differential Privacy Models. Four subdivisions are identified within the Non-perturbation Privacy Preservation Models: random graph editing, k-anonymization methods, clustering-based techniques, and probabilistic privacy preservation approaches. Importantly, the probabilistic privacy preservation approach was introduced as new category for anonymizing social networks. This innovative method, also known as the uncertain graph approach, attributes a specific probability value to each network connection, indicating the likelihood of its existence [11]. This means each connection in the network is not certain but has a chance of being there, which is represented by a probability value. In the context of social network anonymization, an

uncertain graph can help preserve user privacy. By introducing uncertainty into the connections between individuals, it becomes more challenging to infer specific details about an individual based on their connections within the network. The presence of uncertainty can make it harder to re-identify individuals in the network, thereby protecting their privacy.

Furthermore, differential privacy is a mathematical definition of privacy that was first proposed by Cynthia Dwork in 2006 and has since become a popular method for ensuring privacy in data analysis [42, 43]. In tabular databases, it gives a guarantee that the removal or addition of a single database entry does not significantly affect the results of any statistical queries. Within the Differential Privacy category in the context of social network anonymization, Abawajy et al. [1] delineated two subgroups based on whether differential privacy is applied at the node or edge level, which are elaborated in the following:

- Node-level Differential Privacy: This technique focuses on protecting the privacy of the individual nodes and their adjacent edges (their direct connections). When node-level differential privacy is applied, it makes it hard to infer the presence of a specific individual in the social network. The noise is introduced in such a way that whether a particular person is part of the network or not cannot be confidently determined. Moreover, node-level differential privacy provides protection to the edges adjacent to a node. This means that even if an individual is known to be part of the network, the presence or absence of a particular edge of that individual in the anonymized data cannot be determined with certainty.
- Edge-level Differential Privacy: This technique focuses on protecting the connections within a social network, represented as edges in the network graph. The principle behind it is that the presence or absence of specific edges should be concealed to maintain privacy, while the overall pattern of edges can be made public. Edge-level differential privacy aims to add enough "noise" to these edges such that it becomes hard to determine with certainty whether any particular edge exists or not.

The taxonomy proposed by Abawajy et al. [1] is demonstrated in Fig. 3.

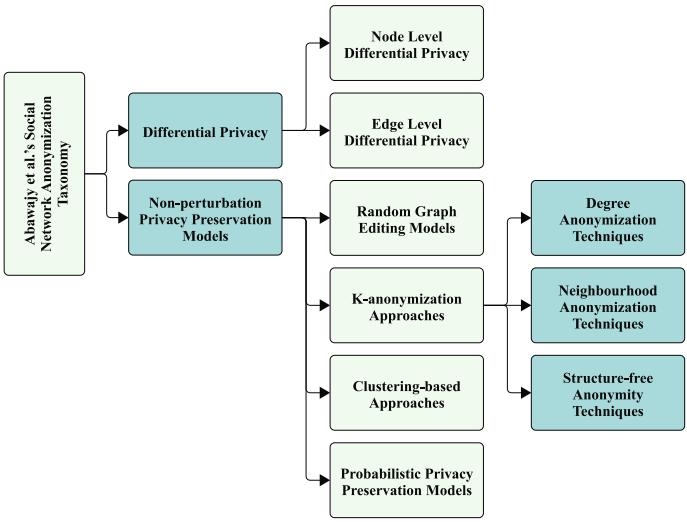


Fig. 3 Social network anonymization taxonomy developed by Abawajy et al. [1]

During the same year, Ji et al. conducted a study on the methods of both anonymizing and de-anonymizing graph data [70]. In their analysis, they reviewed various scholarly articles and classified graph anonymization strategies into six main categories: Naive ID Removal, Edge Editing-based Anonymization that is identical to the edge modification in the graph modification approach; K-anonymity; Aggregation/Class/Cluster based Anonymization; Differential Privacy; and Random Walk-based Anonymization. In their taxonomy, aside from the Naive ID Removal and Random Walk-based Anonymization methods, all other approaches align with those specified in earlier taxonomies. Therefore, Naive ID Removal and Random Walk-based Anonymization are elaborated as follows.

- Naive ID Removal is one of the simplest and earliest approaches for social network anonymization, in which explicit identifiers of the nodes (such as names, usernames, and any unique identifiers) in the network are simply removed or replaced with non-identifying labels.
- Random Walk-based Anonymization is a privacy preservation approach designed to protect the edges between nodes in a social network [115]. This approach replaces an existing edge between two nodes with a new edge that is determined through a process known as a random walk. This approach effectively randomizes the relationships in the network, making it difficult to ascertain the true relationships while maintaining the overall structure and characteristics of the network.

The taxonomy proposed by Ji et al. [70] is represented Fig. 4.

In 2017, Casas-Roma et al. conducted a review of research papers on social network anonymization, focusing on those that suggested graph modification techniques to ensure network anonymity [21]. They provided a comprehensive discussion about the advantages and drawbacks of each method. The researchers organized the graph modification strategies into three primary categories: Edge and Vertex Modification, Uncertain Graphs, and Generalization and Clustering-based approaches. Further, they subdivided the Edge and Vertex

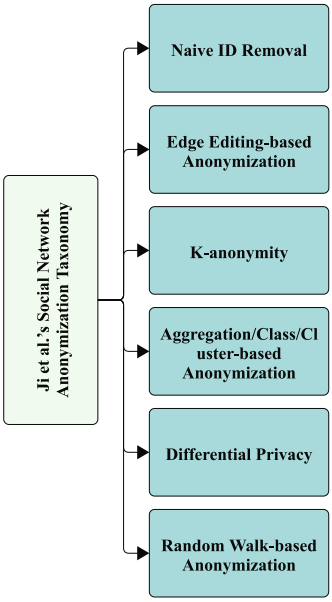


Fig. 4 Social network anonymization taxonomy by [70]

Modification techniques into three additional subcategories, namely K-anonymity, Extending K-anonymity, and Beyond K-anonymity methods. Their taxonomy is illustrated visually in Fig. 5.

In 2018, Siddula et al. conducted a review of the methods designed to preserve the privacy of users and their relationships in social networks. Based on a previous study [21], they classified the privacy concerns of social networks into three categories: Node Privacy, Attribute Privacy, and Link Privacy. Nevertheless, their article focused only on studies suggesting anonymization methods to protect node and link privacy. As previously highlighted, the authors' focus was exclusively on node and link privacy issues. For the category of node privacy, they examined studies that implemented anonymization through Naïve Anonymization and Node Perturbation methods. After analyzing the surveyed research, they concluded that there are predominantly two strategies to perturb nodes in social networks, specifically Random Perturbation and Constrained Perturbation. Additionally, in terms of link privacy, they examined research introducing social network anonymization techniques based on Edge Perturbation and Random Walk methodologies, from which they determined that there are generally five distinct approaches to perturb the edges in a social network. These methods include Intact Edges, Partial-Edge Removal, Cluster-Edge Anonymization, Cluster-Edge Anonymization with Constraints, and Removed Edges.

It is worth mentioning that Siddula et al. did not propose a taxonomy for techniques used in social network anonymization. Nevertheless, we have derived a classification scheme based on the studies they reviewed, which is graphically illustrated in Fig. 6.

In 2019, Sathiya Devi et al. conducted a literature survey concerning anonymization methods in social networks [142]. Their primary focus was on techniques used to maintain the attribute privacy of social network users. In line with this, they grouped these methods into five distinct categories: K-anonymity, L-diversity, T-closeness, Slicing, and Differential Privacy. In their suggested classification, three concepts, namely L-diversity, T-closeness, and Slicing, are elaborated upon as follows.

- L-diversity: This model is an extension of K-anonymity, introduced to overcome some of its limitations. The principle of L-diversity is that within each group of "K" indistinguishable individuals, there should be at least "L" "well-represented" values for each sensitive attribute.
- T-closeness: This model is another extension of K-anonymity and L-diversity, introduced to address their remaining weaknesses. The principle of T-closeness is that the distribution of a sensitive attribute within any group of "K" indistinguishable individuals must be close to the overall distribution of that attribute in the entire dataset.

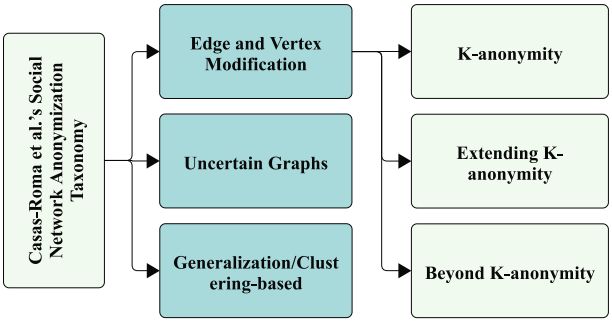


Fig. 5 Social network anonymization taxonomy by [21]

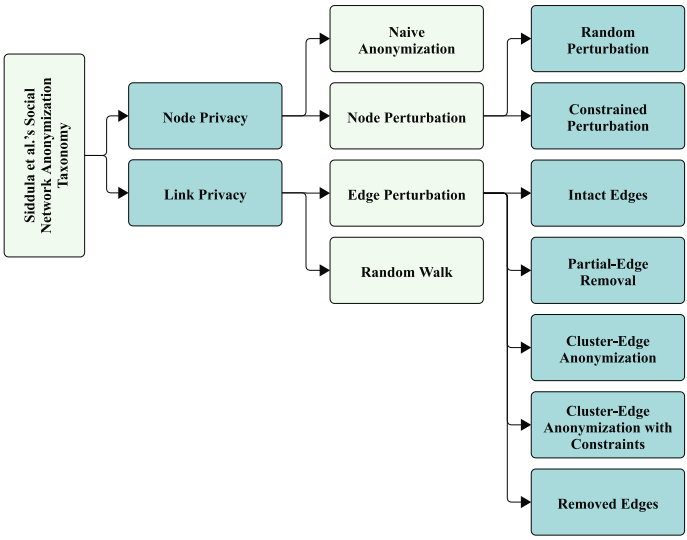


Fig. 6 Social network anonymization taxonomy based on Siddula et al. [146]

- Slicing: This technique divides data both horizontally and vertically. The horizontal division groups together similar records. The vertical division separates the dataset into different "slices," each containing a subset of the attributes. Each slice is then independently anonymized by permuting the order of the records within the slice.

A visual representation of their taxonomy is provided in Fig. 7.

In 2020, Majeed et al. reviewed the techniques employed in anonymizing data to preserve the privacy of published data [107]. As part of their study, they offered a taxonomy for the anonymization techniques used to protect the privacy of social network data. In their proposed taxonomy, privacy-aware graph computation refers to a strategy where instead

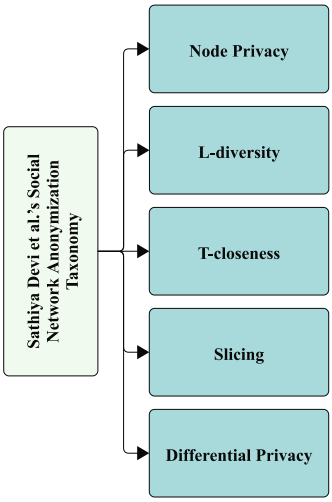


Fig. 7 Social network anonymization taxonomy proposed by Sathiya Devi et al. (2019)

of sharing the entire graph data (which might include sensitive information), only specific aggregate properties or statistics of the graph are computed and released in response to queries from data analysts. Moreover, hybrid graph anonymity methods combine different techniques for anonymizing social networks, aiming to create an appropriately anonymized version of the network. This approach seeks to address the trade-off between maintaining privacy and preserving usefulness. The schematic representation of their proposed taxonomy can be seen in Fig. 8.

In a recent study, Kiranmayi et al. scrutinized research focused on anonymizing social networks in order to preserve user privacy [79]. They proposed a classification of methods used for social network anonymization, which is visually represented in Fig. 9.

While the previously mentioned studies have indeed provided a valuable understanding of the various social network anonymization approaches and taxonomies, to the best of our

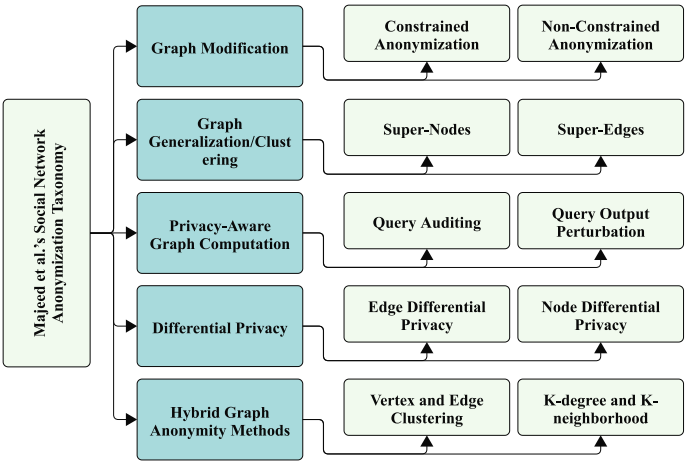


Fig. 8 Social network anonymization taxonomy developed by Majeed et al. (2020)

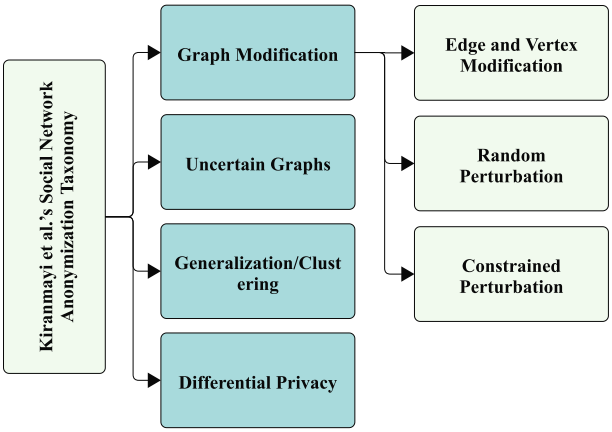


Fig. 9 Social network anonymization taxonomy proposed by Kiranmayi et al. (2021)

knowledge, there has yet to be a comprehensive bibliometric literature review conducted within this area. Such a study could offer insight into emerging trends, key themes, and potential areas of interest for future research. Additionally, bibliometric analysis is able to provide significant benefits in identifying critical research directions and key enabling technologies [153–155].

Consequently, this article presents a bibliometric analysis of social network anonymization studies to identify the current trends in this field, the main approaches of social network anonymization, as well as the recent advances. In this regard, we collected all the papers conducted from 2007 to 2022 from the open-source Elsevier Developer Portal using the Scopus API (Elsevier 2022). After preprocessing these works, we performed statistical, network, and co-word analyses to detect the high-trend topics and themes in this field. Additionally, a new taxonomy of the current social network anonymization approaches is provided. The main contributions of this work can be summarized as follows:

(1) All published studies related to the social network anonymization field were collected and preprocessed.
(2) A network visualization of the used keywords was created by the authors to understand the most frequent keywords.
(3) Statistical and network analyses were performed to identify the main keywords, themes, and topics.
(4) A co-word analysis was conducted to detect the prominent themes and topics.
(5) The evolution of the social network anonymization themes was investigated.
(6) A novel taxonomy of the social network anonymization approaches was developed based on the conducted analyses.
(7) Future research trends in the social network anonymization field are presented.

3 Research methodology

The methodology of this study draws upon the bibliometric analysis of the previous research studies conducted in the field of social network anonymization. Bibliometric analysis involves systematically evaluating academic publications to identify trends, prominent themes, and research patterns [153–155]. As shown in Fig. 10, we began by gathering all available studies related to social network anonymization. These studies were then preprocessed, which involved cleaning the data, removing duplicates, and ensuring that the relevant documents were included for analysis.

Next, we examined the keywords used by the authors in these studies. By analyzing these keywords, we created a visual network map that illustrates how often specific keywords appeared and how they are related to one another. Afterward, we conducted both statistical and network analyses. These analyses provided us with insights into the significant keywords and key focus areas.

To go deeper, we performed a co-word analysis, which identified groups of related terms that appeared together frequently. This allowed us to pinpoint the most prominent themes and topics in the field. Additionally, we tracked the evolution of themes over time to observe how research in social network anonymization has progressed. Finally, based on the results of the mentioned analyses, we developed a new taxonomy of social network anonymization approaches, highlighting the most important techniques and trends in the field.

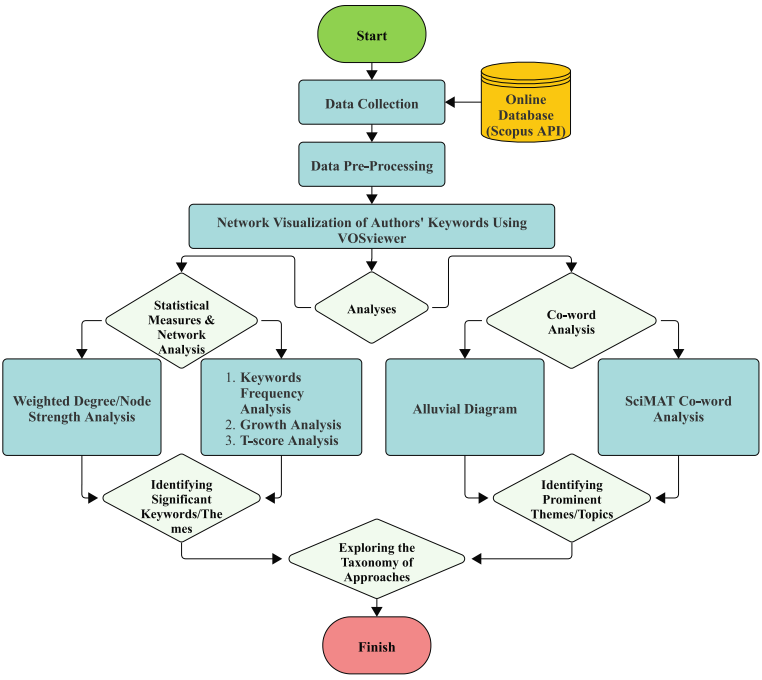


Fig. 10 The research methodology

3.1 Data collection and preprocessing

As mentioned earlier, this study used the author keywords of the papers published in the social network anonymization field as the dataset. The articles were gathered from the open-source Elsevier Developer Portal via the Scopus API (Elsevier 2022), offering a detailed assortment of features associated with the retrieved papers, such as article title, journal title, year of publication, authors, authors' affiliations, authors' keywords, the number of citations, and references. Scopus database is chosen for this study primarily because of its extensive coverage of peer-reviewed journal articles and conference proceedings, particularly in the fields relevant to social network anonymization, such as computer science and engineering. While databases like Web of Science (WoS) also offer valuable resources, Scopus covers a larger number of journals and has a wider inclusion of interdisciplinary research [120], making it more suitable for capturing the broad landscape of social network anonymization research. Additionally, Scopus provides comprehensive citation data, which is essential for conducting detailed bibliometric analyses. Although it may not capture every type of publication (such as preprints or some conference papers), its robust and reliable dataset has been widely used in previous bibliometric studies across various fields [3, 5, 154, 155]. Therefore, using Scopus allows this study to offer a well-rounded analysis of the core research trends and developments in the social network anonymization domain.

The search query "social network anonymization" and related searches, such as "online social network anonymization," "graph anonymization," "social network privacy preservation," "graph privacy preservation," and "social graph anonymization," were employed to extract all relevant articles in this area of research.

From the collected articles, we found that 315 papers were published in the field of social network anonymization over the past 16 years (2007–2022). The information displayed in Fig. 11 suggests a consistent rise in the quantity of research papers published in this domain from 2007 to 2019, reflecting a growing interest from researchers. However, from 2020 to 2022, the number of publications declined in the field of social network anonymization.

These papers have been published in 193 different journals and conferences. Some of these publication venues, specifically those with a higher volume of published papers, are ranked and visually represented in Fig. 12.

Furthermore, Fig. 13 provides a graphical representation of the number of articles published by each country, limiting the scope to those countries that have produced at least five articles. The top contributors are China, the USA, India, Spain, and Iran. China stands as the leading publisher, contributing 33% of the total articles sourced for the current research study. Following China, the USA has produced 22%, India has published 10%, Spain has contributed 7%, and Iran accounts for 5% of the total published articles. Several factors contribute to the higher productivity of countries like China, the USA, and India in the field

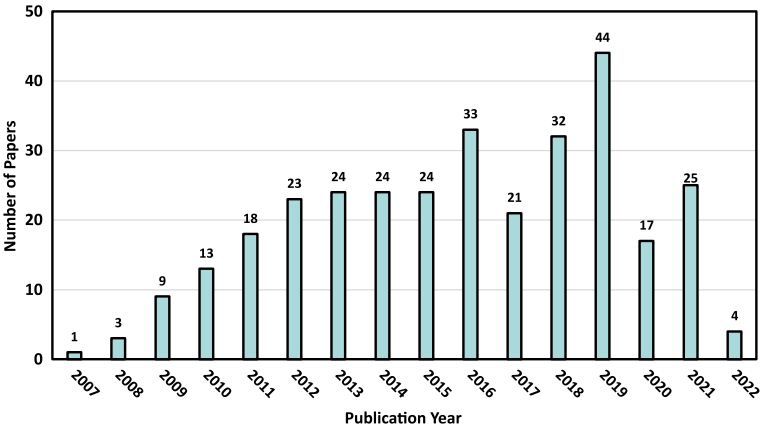


Fig. 11 Number of papers per year

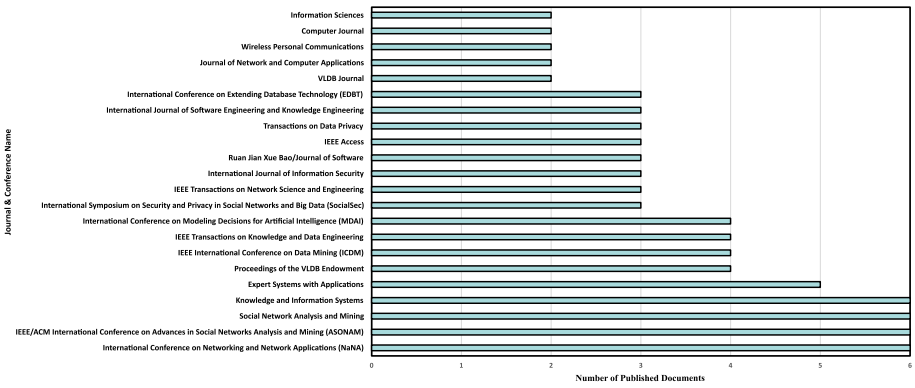


Fig. 12 Top-ranked publication venues for the social network anonymization field

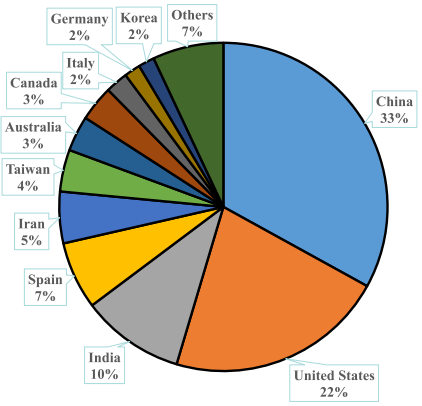


Fig. 13 The Countries with most publication in social network anonymization field

of social network anonymization. These countries invest significantly in research and development, particularly in data privacy, artificial intelligence, and information security. China and the USA, for instance, benefit from strong academic institutions, robust funding, and active tech industries, which lead to higher publication output. Similarly, India's focus on technological research and its vast academic network have contributed to its presence in this field.

In the current study, the research period was segmented into four consecutive subperiods: 2007–2010, 2011–2014, 2015–2018, and 2019–2022. Also, the number of publications in each subperiod is demonstrated in Fig. 14. An R-squared value of 0.57 indicates that the regression model explains 57% of the variation in the number of publications. This means that there are other factors that affect the number of publications besides the time period alone. The remaining 43% of the variation may be due to other factors, such as changes in the research landscape, funding availability, or other external factors.

With respect to the data preprocessing stage, it is worth mentioning that the following processes were performed to improve the quality and clarity of the keywords in the dataset, making it more suitable for analysis:

- Removing duplicates: Duplicate keywords were removed to reduce the noise in the dataset.

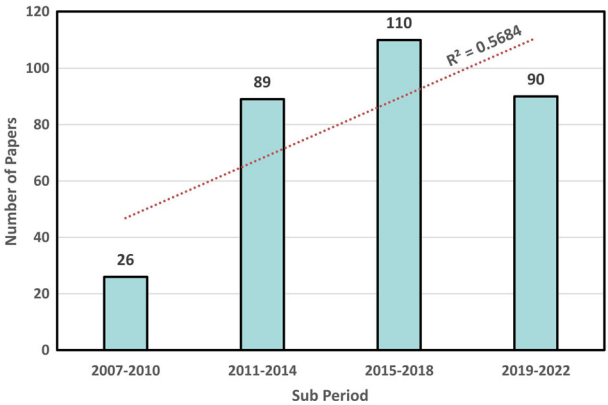


Fig. 14 Number of papers in each subperiod and the fitted regression line

- Stemming: Words were reduced to their root form to group different variations of the same word.
- Lemmatization: Words were converted to their base form to capture their core meaning.
- Removing special characters: Special characters were removed to improve the readability of the dataset.
- Missing values: For papers lacking specific authors' keywords, indexed keywords supplied by the Elsevier Scopus API (Elsevier 2022) were used as substitutes.
- Misspelled words correction: Some papers contained misspelled keywords, like "anonimisation." To address this problem, we developed Python scripts to identify keywords with the most similar spelling and replace the incorrect ones. Thus, the incorrectly spelled keyword "anonymisation" was replaced with the correct term, "anonymization."

Following the application of the above-mentioned preprocessing steps, the dataset's overall attributes are as follows: The total count of keywords in the dataset amounts was determined to be 1547, out of which only 361 keywords are unique. Therefore, each paper contains an average of 4.91 keywords.

Additionally, the most frequently used keyword in the dataset appeared 244 times, while the least frequent ones appeared only once. These statistics provide a general idea about the distribution of keywords in the dataset and help to understand the characteristics of the research in the field of social network anonymization. The information will be used to identify the dominant themes and topics in the field and gain insights into the areas that have received more attention from researchers.

3.2 Authors' keywords network visualization

The frequency information mentioned earlier was utilized in constructing a keyword-keyword network using the VOSviewer software. This network represents keywords as nodes and establishes edges based on the frequency of co-occurrence between pairs of keywords. By examining this network, it becomes possible to identify relationships between keywords and gain a comprehensive understanding of their connections. Moreover, the network incorporates clusters of keywords that depend on the strength of their relationship [4, 171]. Additionally, the frequency information aids in ranking the keywords according to their popularity, offering valuable insights into the most relevant topics within the field of social network anonymization.

Following the construction of the keyword-keyword network, an analysis was conducted to unveil concealed patterns and relationships. This analysis involved the presentation of network diagrams, density visualizations in the form of heatmaps, and overlay visualization. By utilizing the VOSviewer tool, it was possible to identify the most important and influential keywords, examine the relationships between keywords, determine the time period during which each keyword was utilized, and comprehend the overall structure of the keyword-keyword network.

In the following figures, the co-occurrence mapping of the keywords for the four defined subperiods is depicted.

In the network in Fig. 15 for the 2007–2010 subperiod, "Social Network," "Privacy Preservation," and "Anonymization" are influential keywords that are positioned near the center with larger nodes. This implies that these keywords have persistently been significant keywords in the field of social network anonymization. The "Structural Properties," "K-anonymity," "Generalization Approach," and "Social Network Analysis (SNA)" keywords are closely related

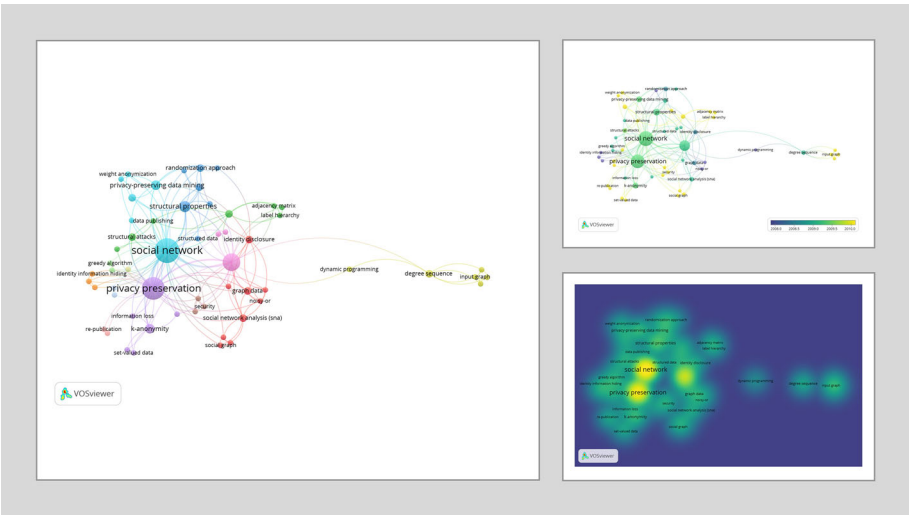


Fig. 15 Co-occurrence mapping of 2007–2010 authors' keywords

to the mentioned influential keywords, demonstrating the co-occurrence of these keywords in the social network anonymization field.

Furthermore, in the 2011–2014 network illustrated in Fig. 16, "Social Network," "Privacy Preservation," "Anonymization," and "K-anonymity" emerge as prominent keywords. Again, these keywords are situated close to the center with larger nodes, indicating their significant influence. It is worth mentioning that "Data Publishing," "Clustering Algorithms," "Structural Properties," "Utility," and "L-diversity" keywords have a strong association with the

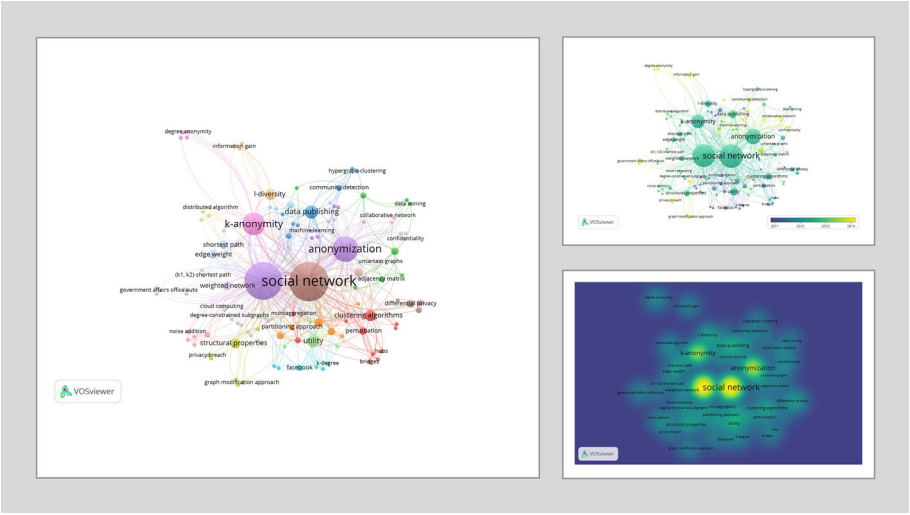


Fig. 16 Co-occurrence mapping of 2011–2014 authors' keywords

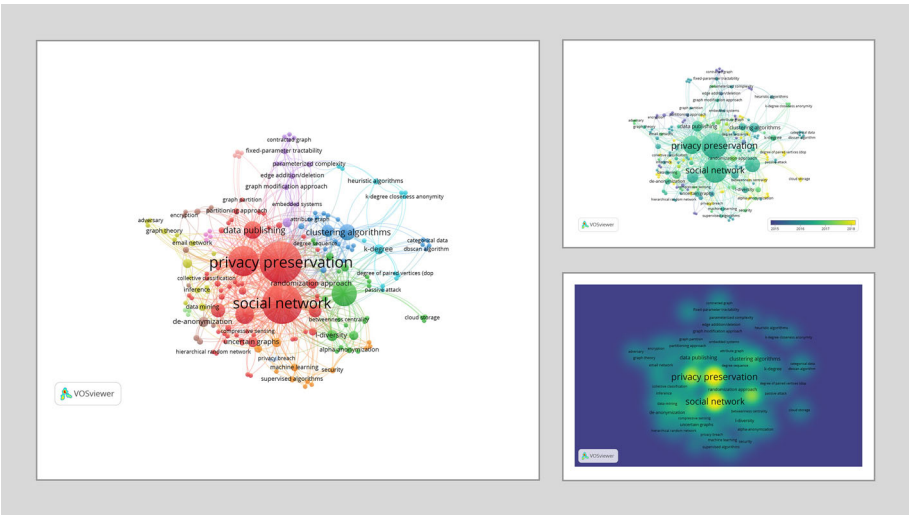


Fig. 17 Co-occurrence mapping of 2015–2018 authors' keywords

aforementioned influential keywords. This illustrates their co-occurrence and interrelation in the field of social network anonymization.

In the 2015–2018 network shown in Fig. 17, once again, "Social Network," "Privacy Preservation," "Anonymization," and "K-anonymity" stand out as influential keywords. Moreover, a number of keywords, such as "Utility," "Data Publishing," "Clustering Algorithms," "Generalization Approach," "Differential Privacy," "Graph Modification Approach," "Background Knowledge," "Machine Learning," "Uncertain Graphs," and "L-diversity" exhibit a strong connection with the influential nodes. This illustrates that the co-occurrence of these keywords was of considerable interest to researchers during the 2014–2018 subperiod.

In the network of the last subperiod 2019–2022 in Fig. 18, keywords "Social Network," "Privacy Preservation," "Anonymization," "K-anonymity," "Clustering Algorithms," "Generalization Approach," "Differential Privacy," "Utility," "Data Publishing," and "Graph Theory" are influential nodes. Additionally, keywords "Graph Modification Approach," "Structural Properties," "Information Loss," "Genetic Algorithm (GA)," "Randomization Approach," "Perturbation Techniques," "Background Knowledge," "K-degree," "Graph Matching Technique," "Uncertain Graphs," "Optimization," "Fuzzy Set Theory," and "Artificial Neural Network (ANN)" demonstrate a strong link with the influential nodes mentioned earlier, indicating their interrelation in the field of social network anonymization.

The co-occurrence mapping of all the authors' keywords used in the 2007–2022 period is provided in Fig. 19.

As shown in this figure, keywords "Social Network," "Privacy Preservation," "Anonymization," "K-anonymity," "Data Publishing," "Clustering Algorithms," "Generalization Approach," "Differential Privacy," "Graph Modification Approach," "Uncertain Graphs," "Information Loss," "Structural Properties," "Randomization Approach," "L-diversity," "Graph Theory," "Background Knowledge," "Social Network Analysis (SNA)," "K-degree," "Machine Learning," and "Perturbation Techniques" are among the most influential keywords

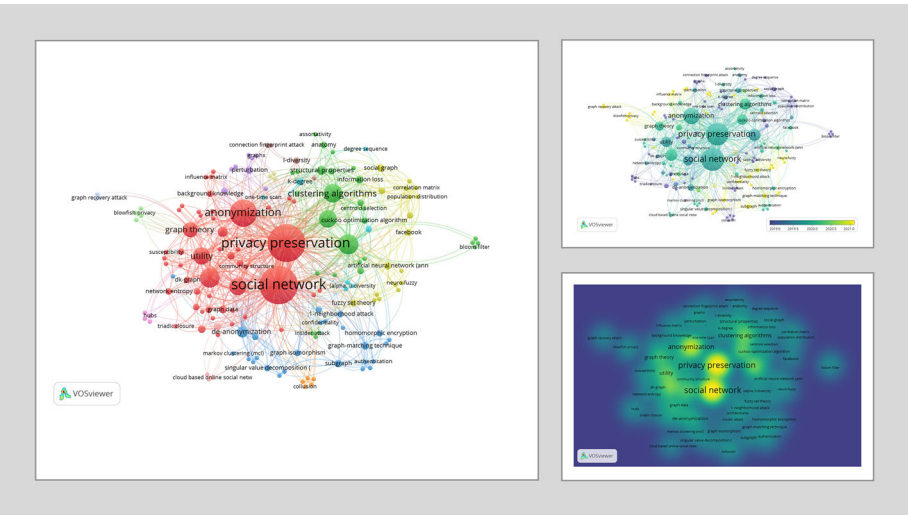


Fig. 18 Co-occurrence mapping of 2019–2022 authors' keywords

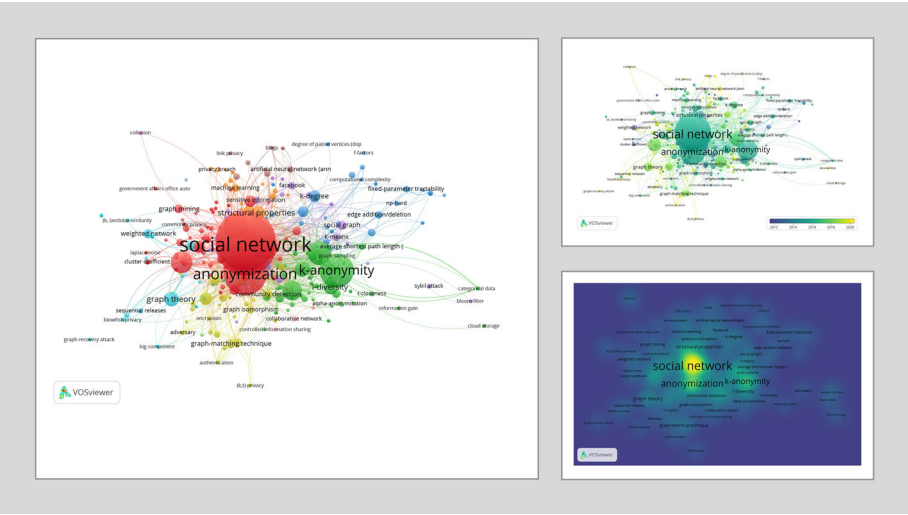


Fig. 19 Co-occurrence mapping of 2007–2022 authors' keywords

used by researchers in the social network anonymization field. Moreover, the overlay visualization provides key insights into the period of usage for particular keywords by researchers in the specified field. For instance, the most recent keywords utilized by researchers include "Graph matching Technique," "Artificial Neural Network (ANN)," "Optimization," "Particle Swarm Optimization Algorithm (PSO)," "Cuckoo Optimization Algorithm (COA)," "Firefly Algorithm (FA)," "Hierarchical Clustering," "Blowfish Privacy," and "Assortativity."

3.3 Statistical measures and network analysis

In the following section, statistical analyses and network analysis, including the frequency analysis of the keywords, their relative growth, the T-score values, and weighted degree/node strength, are provided on the research dataset during the successive subperiods.

3.3.1 Most significant keywords

This section outlines the statistical and network analysis techniques employed to analyze the authors' keywords. Specifically, the study leverages growth analysis, T-score, and weighted node strength metrics to evaluate the significance of frequently occurring keywords across four distinct subperiods. Growth analysis identifies the expansion of research topics by measuring changes in keyword frequency, while T-score analysis evaluates the statistical significance of keywords based on their frequency, mean, and standard deviation. Additionally, weighted node strength assesses the importance of a keyword by summing the weights of its network connections. These methods provide insights into the evolution and relevance of key topics in the research area.

We report these metrics not only to assess the impact and importance of the keywords thus far but also to predict which keywords are more likely to become future trends. To this end, a combined importance score was calculated, which integrates the normalized values of growth, T-score, and weighted node strength. In this study, equal weights were assigned to these metrics to ensure balanced importance across all measures. The detailed formulas for each metric are provided in Appendix A for further reference.

Table 1 presents the distribution of each frequent keyword in the 2007–2022 period, and Fig. 20 illustrates the frequency distribution of the commonly used keywords over the four determined subperiods. This figure provides a more insightful representation of the growth of the unique keywords presented in Table 1. Specifically, a large portion of the bar plots for each keyword pertains to the most recent subperiods, indicating that they are currently experiencing increased usage and positive growth.

Furthermore, Table 2 presents the growth, T-score, weighted node strength, and importance factor associated with the frequent keywords. Notably, the growth, T-score, and weighted node strength values for each keyword hold more significance compared to their frequency alone. This is because our goal is to assess the impact and relevance of keywords and topics up to the present time as well as to predict which ones are more likely to become future trends. It is noteworthy that all the frequent keywords listed in Table 1 exhibit statistically significant and positive growth, while none indicate a statistically significant decline according to the data presented in Table 2.

The keywords with the highest relative significant growth rate, T-score, and weighted node strength in Table 2 are "Social Network," "Privacy Preservation," "Anonymization," "K-anonymity," "Data Publishing," "Utility," "Clustering Algorithms," "Differential Privacy," and "Generalization Approach." Therefore, these keywords are considered pertinent to the social network anonymization field. Furthermore, given their high relative importance factor, it is more likely that we will encounter them in future works within this domain.

Additionally, some keywords did not exhibit noticeable growth or high weighted node strength compared to the ones previously mentioned. However, their corresponding T-score significance values suggest that they still represent important topics. These keywords, as shown in Table 2, include "Structural Properties," "L-diversity," "Information Loss," "Graph Modification Approach," "Randomization Approach," "Social Network Analysis (SNA),"

Table 1 The number of frequent keywords in each year

Frequent Keywords	Social network	Privacy preservation	Anonymization	K-anonymity	Data publishing	Utility	Clustering algorithms	Differential privacy	Generalization approach	Structural properties
2007	1	1	0	0	0	0	0	0	0	0
2008	1	1	2	0	0	1	0	0	0	1
2009	8	6	4	2	0	0	1	1	0	1
2010	9	8	4	1	1	0	0	0	1	2
2011	18	16	7	7	2	2	1	1	1	0
2012	21	18	8	7	1	1	2	1	1	3
2013	21	18	10	5	3	3	3	1	1	0
2014	17	18	8	7	3	2	0	0	0	2
2015	17	21	10	4	3	6	5	1	3	0
2016	23	26	12	16	3	3	3	2	1	1
2017	16	16	7	3	6	4	2	4	1	0
2018	23	23	15	8	3	4	3	3	2	0
2019	35	32	16	12	8	7	7	6	6	2
2020	13	14	9	1	4	4	3	6	3	1
2021	19	18	8	6	1	3	6	5	5	1
2022	2	3	2	0	1	0	1	1	1	0

Frequent Keywords	L-diversity	Information loss	Graph modification approach	Graph theory	Randomization approach	Social network analysis (sna)	De-anonymization	Edge weight anonymization	K-degree	Perturbation techniques
2007	0	0	0	0	0	0	0	0	0	0
2008	0	0	0	0	1	1	0	0	0	0
2009	0	1	0	0	1	0	0	0	0	0

Table 1 (continued)

Frequent Keywords	L-diversity	Information loss	Graph modification approach	Graph theory	Randomization approach	Social network analysis (sna)	De-anonymization	Edge weight anonymization	K-degree	Perturbation techniques
2010	0	0	1	2	0	1	0	1	0	0
2011	2	0	0	0	0	0	0	1	1	1
2012	1	3	0	0	1	2	0	1	0	2
2013	1	1	0	0	0	1	0	1	0	0
2014	2	0	1	0	3	0	0	1	0	0
2015	1	3	0	0	0	1	2	0	0	0
2016	3	2	1	1	0	2	0	0	2	1
2017	1	1	1	1	4	0	1	0	2	1
2018	1	2	0	0	1	1	2	0	1	1
2019	1	1	1	5	2	1	3	0	2	2
2020	0	2	2	2	2	1	2	0	1	0
2021	1	0	2	3	2	0	1	2	0	1
2022	0	0	2	1	0	0	0	0	0	0

Frequent Keywords	Graph matching technique	Neighborhood attack	Community detection algorithms	Background knowledge	Partitioning approach	Uncertain graphs	Weighted network	Machine learning	Genetic algorithm (ga)	Directed graph
2007	0	0	0	0	0	0	0	0	0	0
2008	0	0	1	1	0	0	0	0	0	0
2009	0	0	0	0	1	0	0	0	0	0
2010	1	1	0	0	0	0	1	0	0	0

Table 1 (continued)

Frequent Keywords	Graph matching technique	Neighborhood attack	Community detection algorithms	Background knowledge	Partitioning approach	Uncertain graphs	Weighted network	Machine learning	Genetic algorithm (ga)	Directed graph
2011	0	1	0	0	0	0	1	0	0	0
2012	1	0	0	0	2	1	1	0	1	0
2013	0	0	1	0	0	0	2	1	0	0
2014	0	1	1	1	0	0	0	1	0	1
2015	1	0	1	2	1	2	0	0	0	1
2016	0	0	1	0	2	0	1	2	1	0
2017	0	1	0	2	0	2	0	1	0	0
2018	1	1	1	0	0	1	0	0	0	0
2019	0	0	0	1	1	3	0	0	0	2
2020	1	0	1	0	0	0	0	0	1	1
2021	1	1	0	0	1	0	0	1	1	1
2022	1	1	0	1	0	0	0	0	0	0

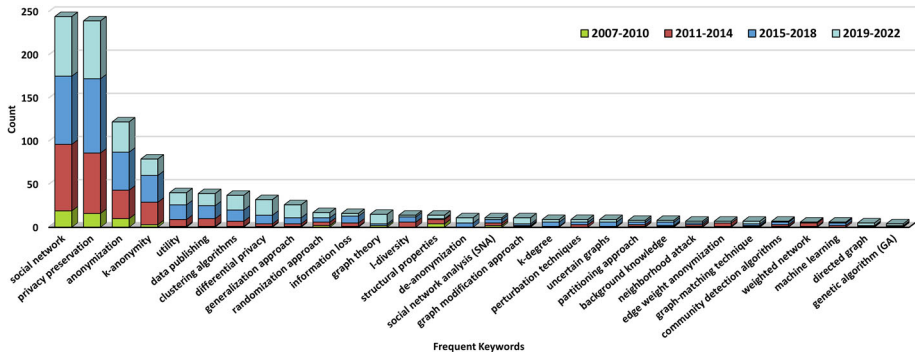


Fig. 20 Analysis of frequency distribution of frequent keywords across four subperiods

"Perturbation Techniques," "Graph Matching Techniques," "Neighborhood Attack," "Community Detection Algorithms," "Uncertain Graphs," "Machine Learning," and "Genetic Algorithm (GA)."

3.4 Co-word analysis

In this section, we describe the co-word analysis [15], which builds upon the earlier co-citation analysis method [151, 152]. Unlike citation-based methodologies, the co-word analysis allows us to explore the relationships and connections between keywords, as well as identify patterns among them within the research context. In other words, the co-word analysis explores the frequency with which pairs of keywords appear simultaneously in the literature, enabling the identification of relationships between clusters of keywords or themes [37, 87] and tracking their developmental trends [36, 88]. The analysis of keyword co-occurrence patterns can reveal the intellectual structure of a particular research field and the connections among its various themes. This method allows for a deeper understanding of the relationships between different topics within the field being studied (Ronda-Pupo & Guerras-Martin, 2012). Over the years, the co-word analysis method has been enhanced and refined by incorporating innovative techniques, such as co-word clustering [14], social network analysis [40], and strategic diagrams [157]. Of particular interest is the strategic diagram, which utilizes measures of density and centrality to map the dynamics of themes and topics within a research field. The strategic diagram serves as a valuable tool in visualizing and analyzing the interplay and evolution of various themes, providing valuable insights into the intellectual landscape of the research field.

The SciMAT software, developed by [33], was used to perform co-word analysis using an algorithm to identify themes across the four different subperiods. The software creates networks of keywords, and edges represent the co-occurrence of keyword pairs in the analyzed documents.

The edge's weight indicates the importance of the relationship in the entire set of documents related to the research field being studied. The results of the analysis are then used to create strategic diagrams that show how thematic areas evolved throughout each subperiod. The SciMAT software follows three main stages:

- Extract clusters of keywords for each subperiod.

Table 2 The Frequent Keywords' Related Growth, T-score, Weighted Node Strength, and Importance Factor

Frequent Keywords	Social network	Privacy preservation	Anonymization	K-anonymity	Data publishing	Utility	Clustering algorithms	Differential privacy	Generalization approach	Structural properties
2007–2010	19	16	10	3	1	1	1	1	1	4
2011–2014	77	70	33	26	9	8	6	3	3	5
2015–2018	79	86	44	31	15	17	13	10	7	1
2019–2022	69	67	35	19	14	14	17	18	15	4
Total	244	239	122	79	39	40	37	32	26	14
Growth	1	0.88295	0.56618	0.47703	0.43732	0.40764	0.34049	0.2794	0.24946	0.29382
T-score ($\|\alpha - p\_value\|$)	0.99992	0.99992	0.99994	0.99398	0.99474	0.99672	0.99285	0.98098	0.97596	0.97657
Weighted Node Strength	1	0.97007	0.51186	0.33436	0.17750	0.18988	0.17750	0.12899	0.14035	0.59855
Importance Factor	0.99997	0.95098	0.69266	0.60179	0.53652	0.53142	0.50361	0.46312	0.45526	0.44342

Frequent Keywords	L-diversity	Information loss	Graph modification approach	Graph theory	Randomization approach	Social network analysis (sna)	De-anonymization	Edge weight anonymization	K-degree
2007–2010	0	1	1	2	2	2	0	1	0
2011–2014	6	4	1	0	4	3	0	4	1
2015–2018	6	8	2	2	5	4	5	0	5
2019–2022	2	3	7	11	6	2	6	2	3
Total	14	16	11	15	17	11	11	7	9
Growth	0.24295	0.1993	0.20499	0.34165	0.15716	0.10629	0.23688	0.22777	0.20955
T-score ($\|\alpha - p\_value\|$)	0.98728	0.97637	0.9655	0.80343	0.96331	0.98594	0.83892	0.86029	0.85537
Weighted Node Strength	0.0650	0.07533	0.05056	0.06914	0.04850	0.05572	0.05159	0.02373	0.04024
Importance Factor	0.43175	0.417	0.40702	0.40474	0.38966	0.38265	0.3758	0.3706	0.36839

Table 2 (continued)

Frequent Keywords	Perturbation techniques	Graph matching technique	Neighborhood attack	Community detection algorithms	Background knowledge	Partitioning approach	Uncertain graphs	Weighted network	Machine learning	Genetic algorithm (ga)	Directed graph
2007–2010	0	1	1	1	1	1	0	1	0	0	0
2011–2014	3	1	2	2	1	2	1	4	2	1	1
2015–2018	3	2	2	3	4	3	5	1	3	1	1
2019–2022	3	3	2	1	2	2	3	0	1	2	3
Total	9	7	7	7	8	8	9	6	6	4	5
Growth	0.13666	0.11388	0.09111	0.08352	0.15944	0.0987	0.20955	0.1025	0.08352	0.09111	0.13666
T-score ($\|\alpha - p\_value\|$)	0.92578	0.96167	0.96167	0.96167	0.84769	0.84769	0.66096	0.71472	0.71472	0.59031	0.44513
Weighted Node Strength	0.04231	0.02683	0.02579	0.02270	0.03921	0.01857	0.02889	0.02373	0.01960	0.02063	0.01547
Importance Factor	0.36825	0.36746	0.35953	0.35596	0.34878	0.32166	0.2998	0.28032	0.27261	0.23402	0.19909

- Investigate the evolution of the extracted clusters over time, with the aim of identifying the primary themes of the research field, their origins, and the connections among them.
- Analyze the performance of the identified themes within each subperiod using quantitative measures such as the number of documents, average citations, and h-index.

Each stage is further explained in the following subsections.

3.4.1 Process of detecting themes

In this section, we discuss the SciMAT software that was applied to analyze the keywords from papers related to social network anonymization and detect the themes. To evaluate the effectiveness and quality of the identified themes and thematic areas, a quantitative and impact analysis was conducted for each subperiod. This analysis involved examining the quantity of documents linked to each theme, encompassing both core documents and secondary documents. Core documents are defined as those that contain at least two keywords appearing within the network of a particular theme, whereas secondary documents are those that have only one keyword associated with the theme's network. It is important to note that both core and secondary documents can potentially belong to multiple thematic networks [32]. Additionally, the quantity of documents, citations, average citation counts, and h-index value of each identified theme were tracked to evaluate their quality and impact.

In the following subsections, the themes are depicted visually. Then, the evolution of the themes is discussed in relation to their performance measures.

3.4.2 Visualization of social network anonymization themes

To explore themes related to the social network anonymization field in different time periods, two types of strategic diagrams are illustrated using the SciMAT software. These diagrams show the spheres' sizes, with the first diagram representing the number of citations received and the second one representing the number of core documents published for each theme (Appendix B displays the core documents relating to each theme within each subperiod). It is worth mentioning that, based on their placement within the strategic diagram, there are four distinct types of themes [13, 14, 36, 37, 61]:

- The themes related to motors located in the upper-right quadrant of the strategic diagram are highly central and dense, indicating their significance for the organization and development of the research field. In other words, these themes are well established and fundamental to the field.
- The upper-left quadrant of the strategic diagram contains the specialized and peripheral themes, which have high density, indicating their advanced development. However, their low centrality implies that these themes have little significance or importance for the research field.
- Themes that are either emerging or disappearing are located in the lower-left quadrant of the strategic diagram. These themes exhibit low density and centrality, suggesting that they are underdeveloped and not very significant for the research field.
- The lower-right quadrant of the strategic diagram contains basic, transversal, and general themes that are significant for the research field but are not yet well developed.

The strategic diagrams and their related tables with quantitative measures for each subperiod are subsequently presented.

During the first subperiod, from 2007 to 2010, a total of 26 documents was extracted and analyzed. The strategic diagrams are presented in Fig. 21.

Table 3 provides the quantitative measures related to the extracted themes.

During the 2007–2010 subperiod, the strategic map indicates that the "Relational Data" and "Structural Attacks" themes were emerging as key areas of focus, while the "Generalization Approach" appeared as an underdeveloped theme that could either grow or fade in importance. Note that the "Structural Attacks" and "Relational Data" are highly central because they address the fundamental challenges in protecting user data within social networks, making them widely applicable and connected to various aspects of privacy research. The relationships between these motor and specialized themes reveal shifting research priorities. For instance, "Structural Attacks" laid the groundwork for more sophisticated anonymization methods in later periods, as researchers increasingly focused on addressing vulnerabilities in social networks. The interplay between themes, such as "Relational Data"

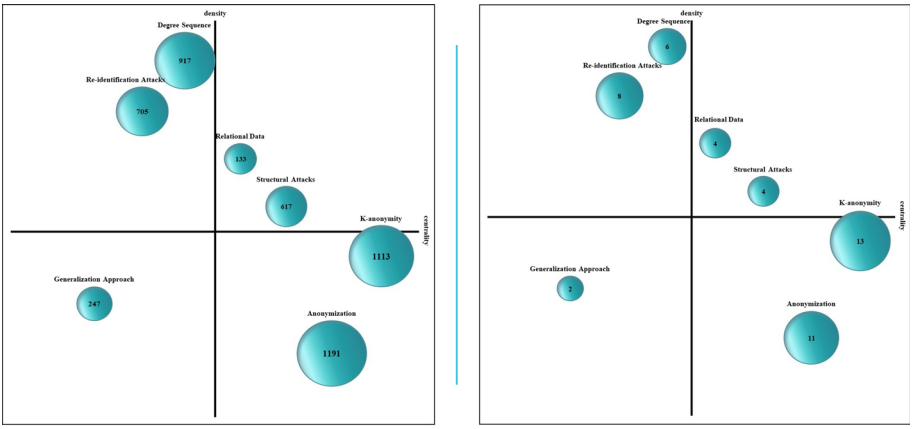


Fig. 21 The strategic diagrams designed for the initial subperiod of 2007–2010 utilizing two parameters for analysis: **a** Number of citations received by these documents and **b** Total count of core documents

Table 3 The metrics used to evaluate the performance of the themes during the subperiod of 2007–2010

Theme name	Core documents count	Core documents h-index	Core documents citations	Core documents average citations
Relational data	4	3	133	33.25
Re-identification attacks	8	6	705	88.12
K-anonymity	13	8	1,113	85.62
Degree sequence	6	5	917	152.83
Structural attacks	4	3	617	154.25
Anonymization	11	9	1,191	108.27
Generalization approach	2	2	247	123.5

and "Re-identification Attacks," also highlights how early research centered on key privacy risks, setting the stage for the development of more complex privacy-preserving techniques as the field matured.

In the following paragraphs, we provide a detailed explanation of each theme during the 2007–2010 subperiod based on the obtained strategic map and the quantitative measures:

The "Structural Attacks" and "Relational Data" themes are classified as motor themes due to their central role in addressing key privacy challenges in social network anonymization. "Structural Attacks" focus on exploiting the inherent connections and patterns within social networks to compromise privacy, such as identifying users through structural vulnerabilities. With 4 core documents, an h-index of 3, and a high total citation count of 617, this theme had a substantial impact, reflecting the urgency of addressing these risks. Similarly, "Relational Data" involves protecting the relationships between users, which are often targeted in privacy breaches. With 4 core documents, an h-index of 3, and 133 total citations, this theme also played an important role, though slightly less prominent. Together, these themes were foundational in shaping research on safeguarding user connections and preventing privacy breaches in social networks.

The "Degree Sequence" and "Re-identification Attacks" themes are categorized as specialized and peripheral due to their higher density but lower centrality, indicating that they are highly developed in their specific subfields but remain somewhat isolated from the broader research landscape. Both themes are well developed, with substantial citation counts and h-index values, reflecting a deep focus within specific areas. However, their low centrality suggests that they were not yet widely integrated into the overall research conversation. Nevertheless, their high average citation counts, 152.83 for "Degree Sequence" and 88.12 for "Re-identification Attacks," demonstrate that these themes had a noticeable impact within specialized areas of the field.

The "Generalization Approach" emerges as an underdeveloped theme with low centrality and density. With only 2 core documents and an h-index of 2, this theme exhibits limited significance during the 2007–2010 subperiod. However, its relatively high average citation count (123.5) indicates that although this area did not gain widespread attention during this time, the documents it produced had some influence in the field. This suggests that the generalization approach might be an emerging theme, showing potential for growth in subsequent periods.

The themes of "Anonymization" and "K-anonymity" are considered basic, transversal, and general, appearing in the lower-right quadrant of the strategic diagram. These themes are marked by high core document counts, strong h-index values, and substantial total citations, which reflect their broad applicability and importance to the research field. Their positioning, however, indicates that while these themes were crucial, they were still in the process of maturing. Their central placement suggests that both "Anonymization" and "K-anonymity" are poised to continue influencing the direction of the field and have the potential to further develop into key research areas in the long term.

Figure 22 depicts the strategic diagram of the second subperiod of 2011–2014.

Also, the quantitative measures related to the extracted themes are provided in Table 4.

During the 2011–2014 subperiod, the strategic map indicates that the "Social Network" and "Facebook" themes were emerging as key areas of focus, while the "Weighted Maximum Common Subgraph (WMCS)" and "Combinatorial Graph" appeared as underdeveloped themes that could either grow or fade in importance. The "Social Network" and "Facebook" themes maintained high centrality because they represent the core domain of anonymization research, making them essential references for nearly all studies in this field. Their relevance to the majority of anonymization studies explains why they dominate the research

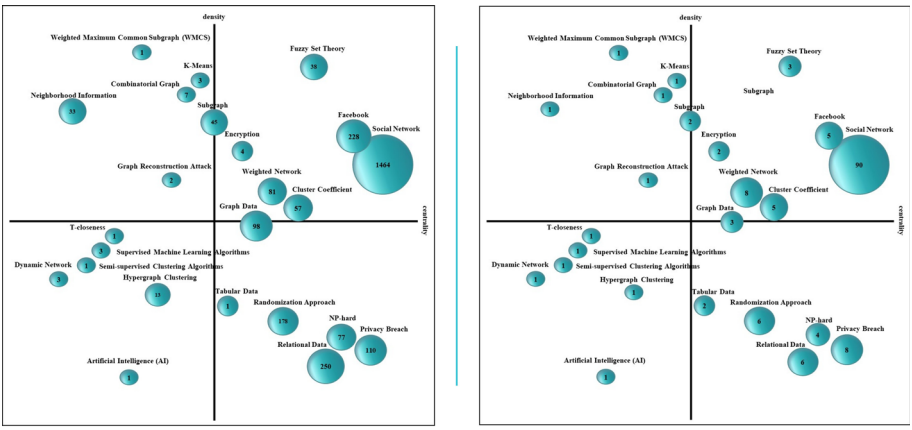


Fig. 22 The strategic diagrams designed for the initial sub period of 2011–2014 utilizing two parameters for analysis: **a** Number of citations received by these documents and **b** Total count of core documents

landscape. The relationships between these themes reflect the evolving priorities within the field. For example, the focus on the "Social Network" theme demonstrates how researchers were addressing the increasingly complex privacy risks associated with large-scale social platforms. The prominence of "Facebook" underscores the importance of specific platforms in driving research during this period, particularly as it was often used as a dataset for testing anonymization techniques. Additionally, the interplay between "Facebook" and themes like the "Randomization Approach" reveals a shift toward using real-world datasets to explore more advanced, technical methodologies for privacy preservation, thereby laying the groundwork for future advancements in anonymization strategies.

The next paragraphs present a thorough analysis of each theme from the 2011–2014 subperiod, guided by the strategic map and quantitative data:

"Social Network," "Facebook," "Fuzzy Set Theory," "Cluster Coefficient," "Encryption," and "Weighted Network" are considered motor themes because they are highly central and dense. The "Social Network" theme is the most prominent theme with the highest core document count (90) and h-index (20), which indicates its central role in shaping the social network anonymization research landscape. "Facebook" is another motor theme, highlighting the importance of using this social network as the experimental dataset in social network anonymization research. "Fuzzy Set Theory," "Cluster Coefficient," "Encryption," and "Weighted Network" have lower core document counts and citations compared to "Social Network" and "Facebook," but their placement in the upper-right quadrant indicates their importance in the research field during this period.

"Weighted Maximum Common Subgraph (WMCS)," "K-Means," "Combinatorial Graph," "Graph Reconstruction Attack," and "Neighborhood Information" are considered as specialized and peripheral themes. These themes have higher density but lower centrality, suggesting they are specialized but somewhat isolated from the main research focus. "Weighted Maximum Common Subgraph (WMCS)" and "Combinatorial Graph" are well developed, with strong h-index values and average citations per document, indicating a deeper focus in specific subfields. However, their low centrality implies that, during this subperiod, they were not widely integrated into the broader research landscape. Their low citation counts

Table 4 The metrics used to evaluate the performance of the themes during the subperiod of 2011–2014

Theme name	Core documents count	Core documents h-index	Core documents citations	Core documents average citations
Facebook	5	5	228	45.6
Fuzzy set theory	3	3	38	12.67
Social network	90	20	1464	16.27
Cluster coefficient	5	3	57	11.4
Randomization approach	6	5	178	29.67
Relational data	6	4	250	41.67
Weighted network	8	3	81	10.12
Privacy breach	8	5	110	13.75
Np-hard	4	3	77	19.25
Encryption	2	1	4	2
Graph data	3	3	98	32.67
Subgraph	2	2	45	22.5
Tabular data	2	1	1	0.5
Weighted maximum common subgraph (WMCS)	1	1	1	1
K-means	1	1	3	3
Neighborhood information	1	1	33	33
Combinatorial graph	1	1	7	7
Graph reconstruction attack	1	1	2	2
Hypergraph clustering	1	1	13	13
Dynamic network	1	1	3	3
Semi-supervised clustering algorithms	1	1	1	1
Supervised machine learning algorithms	1	1	3	3
T-closeness	1	1	1	1
Artificial intelligence (AI)	1	1	4	4

(1 and 7, respectively) suggest that, while specialized, these themes had a relatively limited impact in the field.

"T-closeness," "Supervised Machine Learning Algorithms," "Semi-supervised Clustering Algorithms," "Hypergraph Clustering," "Dynamic Network," and "Artificial Intelligence (AI)" are emerging or disappearing themes. These themes, located in the lower-left quadrant, are characterized by low centrality and density, indicating that they were either emerging or disappearing during this subperiod. Most of these themes have only one core document with low citation counts, signifying limited influence and visibility at the time. However, some of these themes, such as "Dynamic Network" and "Artificial Intelligence (AI)," may represent promising areas that could gain more traction in future periods.

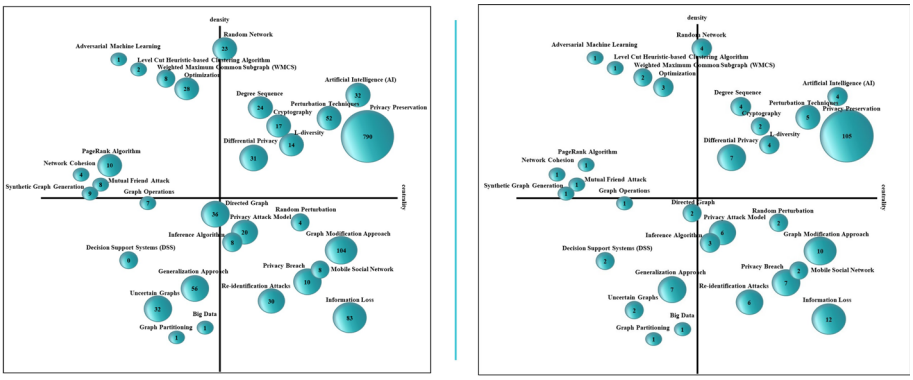
"Tabular Data," "Randomization Approach," "NP-hard," "Privacy Breach," and "Relational Data" are classified as basic and transversal themes, as they have high centrality but lower density. These themes are essential for advancing anonymization methods in social networks. "Relational Data" and "Privacy Breach" focus on safeguarding structured and relational datasets, with an average of 41.67 and 13.75 citations per document, respectively. Also, placing the "NP-hard" and "Randomization Approach" themes in the lower-right quadrant suggests their potential to grow in importance and influence the in the future development of social network anonymization techniques.

Moreover, it is worth nothing that themes on the border lines, i.e., "Graph Data" and "Subgraph," have mixed characteristics, suggesting they might be transitioning between different levels of significance and development within the field. Hence, with respect to the quantitative values of both "Graph Data" and "Subgraph" themes from Table 4, they can be considered motor themes because they are fundamental and crucial themes for the development of the social network anonymization field.

Figure 23 illustrates the strategic map of the extracted themes for the 2015–2018 subperiod.

In addition, the quantitative measures related to the themes of this subperiod are shown in Table 5.

During the 2015–2018 subperiod, the strategic map indicates that the "Privacy Preservation" and "Artificial Intelligence (AI)" themes were emerging as key areas of focus, while the "Optimization" and "Adversarial Machine Learning" themes appeared as underdeveloped



(a) (b)

Fig. 23 The strategic diagrams designed for the initial sub period of 2015–2018 utilizing two parameters for analysis: **a** Number of citations received by these documents and **b** Total count of core documents

Table 5 The metrics used to evaluate the performance of the themes during the subperiod of 2015–2018

Theme name	Core documents count	Core documents h-index	Core documents citations	Core documents average citations
Random network	4	2	23	5.75
Perturbation techniques	5	2	52	10.4
Privacy preservation	105	13	790	7.52
Degree sequence	4	3	24	6
Privacy breach	7	2	10	1.43
Generalization approach	7	5	56	8
Artificial intelligence (AI)	4	4	32	8
Privacy attack model	6	3	20	3.33
Graph modification approach	10	4	104	10.4
Information loss	12	4	83	6.92
L-diversity	4	2	14	3.5
Re-identification attacks	6	3	30	5
Differential privacy	7	4	31	4.43
Optimization	3	3	28	9.33
Weighted maximum common subgraph (WMCS)	2	2	8	4
Cryptography	2	2	17	8.5
Random perturbation	2	1	4	2
Inference algorithm	3	2	8	2.67
Directed graph	2	2	36	18
Level cut heuristic-based clustering algorithm	1	1	2	2
Adversarial machine learning	1	1	1	1
Mobile social network	2	1	8	4

Table 5 (continued)

Theme name	Core documents count	Core documents h-index	Core documents citations	Core documents average citations
Decision support systems (DSS)	2	0	0	0
Graph operations	1	1	7	7
Synthetic graph generation	1	1	9	9
Mutual friend attack	1	1	8	8
Network cohesion	1	1	4	4
PageRank algorithm	1	1	10	10
Uncertain graphs	2	2	32	16
Graph partitioning	1	1	1	1
Big data	1	1	1	1

themes that could either grow or fade in importance. "Privacy Preservation" and "Artificial Intelligence (AI)" dominate because they address widespread privacy concerns related to large-scale platforms and the integration of AI, which are applicable across a broad range of privacy research. The relationships between these themes reflect the evolving priorities within the field. For example, the focus on the "Privacy Preservation" theme demonstrates how researchers were increasingly addressing the privacy challenges posed by large-scale data-sharing platforms. The prominence of "Artificial Intelligence (AI)" underscores the growing integration of machine learning techniques into privacy research during this period, particularly as AI methods were used for both identifying privacy risks and developing anonymization strategies. Additionally, the interplay between "Artificial Intelligence (AI)," "Perturbation Techniques," and "Differential Privacy" provided new ways to anonymize data without significantly compromising its utility. By integrating these cutting-edge methodologies with established privacy-preserving techniques, researchers were laying a strong foundation for more robust and scalable anonymization solutions. These solutions would become essential as the size and complexity of social networks and data-sharing platforms continued to grow in subsequent years, requiring even more advanced methods to ensure user privacy.

The next paragraphs offer a detailed analysis of each theme from the 2015–2018 subperiod, informed by the strategic map and relevant quantitative data:

"Privacy Preservation," "Artificial Intelligence (AI)," "Perturbation Techniques," "L-diversity," "Differential Privacy," "Cryptography," "Degree Sequence," and "Random Network" are placed in the upper-right quadrant and are considered motor themes due to their high centrality and density, indicating their crucial role in shaping the social network anonymization research landscape. "Privacy Preservation" is the dominant theme, with the highest core document count (105) and an h-index of 13, reflecting its central role in the field during this period. "Artificial Intelligence (AI)" also emerged as a key motor theme, highlighting the increasing reliance on AI techniques in privacy research. Themes such as "Perturbation Techniques," "L-diversity," "Differential Privacy," "Cryptography," "Degree Sequence,"

and "Random Network" have lower core document counts and citations compared to "Privacy Preservation" and "Artificial Intelligence (AI)," but their placement in the upper-right quadrant indicates their importance to the research community during this period.

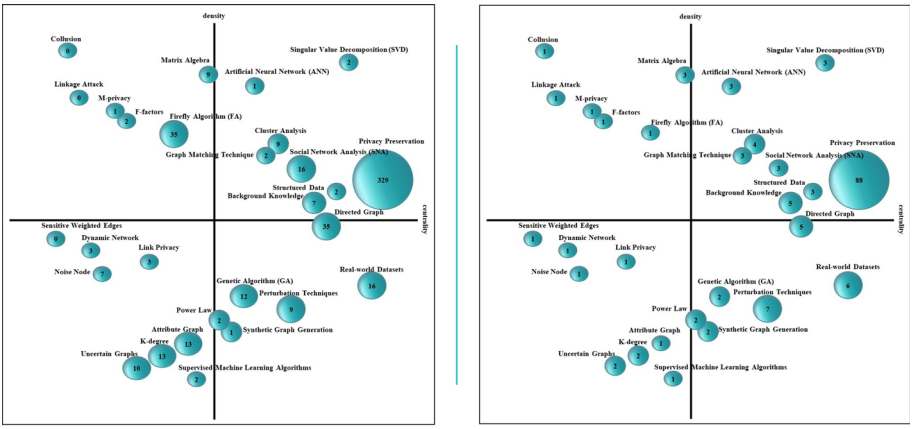
"Optimization," "Adversarial Machine Learning," "Level Cut Heuristic-based Clustering Algorithm," "Weighted Maximum Common Subgraph (WMCS)," "PageRank Algorithm," "Network Cohesion," "Mutual Friend Attack," and "Synthetic Graph Generation" are situated in the upper-left quadrant and are considered specialized and peripheral themes that are advanced but less significant in this subperiod. This suggests that while they are well developed in specific subfields, they remain somewhat isolated from the main research focus. For example, "Optimization" has 3 core documents with an average of 9.33 citations per document, indicating advanced development. However, "Adversarial Machine Learning," "Level Cut Heuristic-based Clustering Algorithm," "Weighted Maximum Common Subgraph (WMCS)," "PageRank Algorithm," "Network Cohesion," "Mutual Friend Attack," and "Synthetic Graph Generation" exhibit advanced development but had limited overall impact during this subperiod, as evidenced by their lower core document counts and citation numbers.

"Graph Operations," "Decision Support Systems (DSS)," "Generalization Approach," "Directed Graph," "Uncertain Graphs," "Big Data," and "Graph Partitioning" themes are located in the lower-left quadrant and are characterized by low centrality and density, indicating that they were either emerging or disappearing during this subperiod. Most of these themes have only one or two core documents with low citation counts, signifying limited influence and visibility at the time. However, some of these themes, such as "Uncertain Graphs" and "Directed Graph," might be underdeveloped and represent promising areas that could gain more traction in future periods.

The lower-right quadrant contains the themes "Random Perturbations," "Graph Modification Approach," "Information Loss," "Re-identification Attacks," "Privacy Breach," "Mobile Social Network," "Privacy Attack Model," and "Inference Algorithm" which are classified as basic and transversal, as they have high centrality but lower density. These themes play a crucial role in advancing anonymization methods in social networks but are still in the process of development. For example, "Random Perturbations" has 2 core documents, an h-index of 2, and an average of 2 citations per document, indicating early-stage research that has yet to gain significant traction. In contrast, "Graph Modification Approach," with 10 core documents, an h-index of 4, and an average of 10.4 citations per document, represents ongoing efforts to tackle privacy challenges through modifications of graph structures, showing more consistent impact. Similarly, "Information Loss," with 12 core documents and an average of 6.92 citations per document, along with "Re-identification Attacks," with 6 core documents and an average of 5 citations per document, provide foundational insights into the trade-offs and consequences of different anonymization techniques. While "Privacy Breach," with 7 core documents, an h-index of 2, and an average of 1.43 citations per document, has had limited impact during this subperiod, it remains an essential concern for future research. Besides, "Mobile Social Network," "Privacy Attack Model," and "Inference Algorithm" exhibit varying core document counts, h-index values, and citation counts. Despite their current moderate influence, their high centrality and low density suggest that they have the potential to grow in importance as research continues to evolve. The placement of these themes in the lower-right quadrant indicates that they will likely play a foundational role in shaping the future development of social network anonymization techniques.

Moreover, the strategic diagram of the last subperiod, 2019–2022, is depicted in Fig. 24.

In addition, the quantitative measures of the themes in the last subperiod are provided in Table 6. Since this is the final subperiod, the average number of citations is significantly



(a) (b)

Fig. 24 The strategic diagrams designed for the initial sub period of 2019–2022 utilizing two parameters for analysis: **a** Number of citations received by these documents and **b** Total count of core documents

lower than the previous subperiods. This is primarily due to the relationship between this factor and the number of papers. In the last subperiods, the citation count decreases due to the limited number of documents in the field.

During the 2019–2022 subperiod, the strategic map highlights "Privacy Preservation," "Artificial Neural Network (ANN)," and "Graph Matching Technique" as emerging key areas of focus, reflecting the growing sophistication in social network anonymization. The continued prominence of "Privacy Preservation" underscores the ongoing challenge of safeguarding user data in complex social networks, where risks are evolving alongside the expansion of data-sharing platforms. The emergence of "Artificial Neural Network (ANN)" points to the increasing integration of machine learning techniques in privacy research. AI-driven approaches such as ANN are being applied to identify vulnerabilities and enhance anonymization strategies, enabling more proactive, scalable methods for handling large datasets. Similarly, the "Graph Matching Technique" demonstrates the focus on comparing graph structures to prevent privacy breaches, a critical task in social networks where relationships between users must be carefully managed to protect sensitive data. Meanwhile, specialized themes such as "Collusion," "Matrix Algebra," and "Linkage Attack" remain peripheral, reflecting niche areas of research with specific applications. These themes, while important in certain contexts, have yet to gain broader traction in the field. For instance, "Collusion" addresses multi-attacker scenarios, which, though relevant, are less central to mainstream anonymization strategies. The relationships between motor and specialized themes reveal a clear trend; for example, foundational themes such as "Privacy Preservation" continue to anchor the field, while emerging technical methods, including AI and graph-based approaches, are becoming increasingly vital. The inclusion of real-world datasets and optimization techniques, including "Directed Graph" and "Genetic Algorithm (GA)," points to a growing emphasis on practical, scalable solutions for anonymization in complex networks.

In the next paragraphs, we present a detailed analysis of each theme from the 2019–2022 subperiod, using the strategic map and associated quantitative data:

"Privacy Preservation," "Background knowledge," "Artificial Neural Network (ANN)," "Graph Matching Technique," "Cluster Analysis," "Singular Value Decomposition (SVD),"

Table 6 The metrics used to evaluate the performance of the themes during the subperiod of 2015–2018

Theme name	Core documents count	Core documents h-index	Core documents citations	Core documents average citations
Singular value decomposition (SVD)	3	1	2	0.67
Privacy preservation	88	9	329	3.74
Artificial neural network (ANN)	3	1	1	0.33
Graph matching technique	3	1	2	0.67
Matrix algebra	3	2	9	3
Perturbation techniques	7	2	9	1.29
Cluster analysis	4	2	9	2.25
Social network analysis (SNA)	3	2	16	5.33
Structured data	3	1	2	0.67
Background knowledge	5	2	7	1.4
Real-world datasets	6	3	16	2.6
Directed graph	5	3	35	7
Genetic algorithm (GA)	2	2	12	6
Collusion	1	0	0	0
Power law	2	1	2	1
Synthetic graph generation	2	1	1	0.5
Firefly algorithm (FA)	1	1	35	35
F-factors	1	1	2	2
M-privacy	1	1	1	1
Linkage attack	1	0	0	0
Uncertain graphs	2	2	10	5
K-degree	2	1	13	6.5
Noise node	1	1	7	7
Link privacy	1	1	3	3
Dynamic network	1	1	3	3
Sensitive weighted edges	1	0	0	0

Table 6 (continued)

Theme name	Core documents count	Core documents h-index	Core documents citations	Core documents average citations
Attribute graph	1	1	13	13
Supervised machine learning algorithms	1	1	2	2

"Structured Data," and "Social Network Analysis (SNA)" are classified as motor themes due to their high centrality and density, reflecting their importance in driving research within the social network anonymization field. "Privacy Preservation" is the most significant theme, with 88 core documents and 329 citations, underscoring its crucial role in shaping privacy-preserving techniques. This theme focuses on protecting the personal information and sensitive relationships of social network users, remaining central to the research landscape. While "Artificial Neural Network (ANN)," "Graph Matching Technique," and "Singular Value Decomposition (SVD)" had lower core document counts, their placement in the upper-right quadrant indicates that they are becoming essential in refining data protection methodologies in social networks. These methods, along with "Background knowledge," play a key role in creating privacy-preserving techniques that address the complex relationships in social network datasets.

"Collusion," "Matrix Algebra," "Linkage Attack," "M-privacy," "F-factors," and "Firefly Algorithm (FA)" themes are positioned in the upper-left quadrant, indicating higher density but lower centrality, which suggests that while they are specialized and well developed, they remain peripheral to the broader field. For example, "Firefly Algorithm (FA)" is an optimization technique applied in specific cases of social network anonymization, but it is not yet widely integrated into the core research landscape. Similarly, "M-privacy" and "Linkage Attack" represent privacy models and attack scenarios that are important in niche areas but do not yet hold widespread relevance across all privacy preservation contexts. These themes are crucial in their respective subfields but have not significantly influenced the overall direction of social network anonymization research.

The "Sensitive Weighted Edges," "Dynamic Network," "Link Privacy," "Noise Node," "Attribute Graph," "K-degree," "Uncertain Graphs," and "Supervised Machine Learning Algorithms" themes, located in the lower-left quadrant, are characterized by low centrality and density, indicating that they may be either emerging research directions or fading interests within the field. For instance, themes like "Link Privacy," "Noise Node," "Uncertain Graphs," and "Dynamic Network" may represent emerging topics that could gain more traction as researchers continue to explore innovative methods in social network anonymization. These themes hold the potential for introducing novel privacy-preserving approaches. Conversely, themes like "Attribute Graph," "K-degree," and "Sensitive Weighted Edges" could be considered disappearing if their relevance has decreased due to advances in more prominent themes or emerging methodologies.

The "Genetic Algorithm (GA)," "Directed Graph," "Real-world Datasets," "Perturbation Techniques," "Power Law," and "Synthetic Graph Generation" themes are situated in the lower-right quadrant, indicating high centrality but lower density, which suggests that while they are broadly impactful across the research field, they are not yet fully developed. These

themes are essential building blocks that can be applied across a range of privacy preservation problems in social networks. For example, "Perturbation Techniques," with 7 core documents and an average of 1.29 citations per document, reflects general methods used to anonymize graph data by introducing noise. These techniques can be applied to a wide array of social network privacy preservation scenarios. Similarly, "Directed Graph" and "Genetic Algorithm (GA)" are fundamental concepts that can be applied in various contexts, playing a key role in shaping future research developments.

Based on the conducted analyses of this section, it is worth noting that the basic and motor themes in the social network anonymization domain achieve the highest citation scores and impacts. By recognizing these basic themes, researchers can better comprehend the foundational concepts, principles, and building blocks that support this research area. This understanding aids in establishing a robust foundation, which is essential for grasping more advanced and specialized topics. Moreover, basic themes frequently act as a launchpad for new research, allowing researchers to spot potential research gaps and opportunities for further investigation, innovation, and cooperation. Additionally, being familiar with the motor themes offers valuable insights into the field's primary challenges and prospects, assisting researchers in developing new research questions and hypotheses.

3.4.3 Evolution of social network anonymization themes

After examining how the number of common keywords changes throughout various subperiods. Subsequently, we investigated the development of thematic areas to track the progression of themes.

As depicted in Fig. 25, there is a variation in the constancy of keywords across each subperiod. Although several keywords consistently appear, some newly appear or fade away within each subperiod. In addition, specific keywords like "T-closeness" exhibit uniqueness by being exclusive to particular subperiods. Despite this, a number of keywords are consistently present across all the analyzed subperiods, such as "Social Network," "Privacy Preservation," "Anonymization," "K-anonymity," "Generalization Approach," "Differential Privacy," "Graph Modification Approach," "Randomization Approach," and "Clustering Algorithms."

It is worth noting that, in overlapping map, the inclusion similarity index refers to the degree of overlap or similarity in the set of keywords between two subperiods, with a higher value indicating a greater degree of similarity. Taking this into consideration and based on the information presented in Fig. 25, the number of shared keywords between the (2007–2010) and (2011–2014) subperiods is 49. This indicates that these two subperiods share some common research topics or themes in the social anonymization field. Also, the inclusion similarity index of 0.7 between the (2007–2010) and (2011–2014) subperiods suggests a relatively high degree of overlap/similarity in the set of shared keywords. This implies a

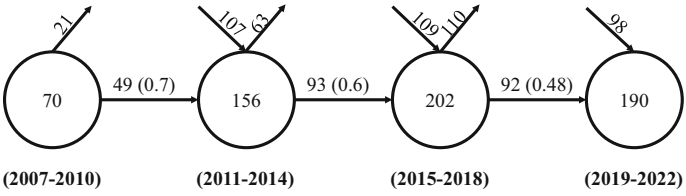


Fig. 25 Overlap mapping: the proportion of keywords that are shared between consecutive subperiods, both incoming and outgoing

certain level of continuity or similarity in research themes or topics between these two subperiods.

Furthermore, the number of shared keywords increases between the (2011–2014) and (2015–2018) subperiods, which suggests that there is more overlap in research topics or themes between these two subperiods. However, the inclusion similarity index between the (2011–2014) and (2015–2018) subperiods decreases to 0.6, indicating a lower degree of overlap or similarity in the set of shared keywords between these two subperiods. This signifies a decreasing level of continuity or similarity in research themes or topics between these two subperiods.

Interestingly, the number of shared keywords between the (2015–2018) and (2019–2022) subperiods is slightly lower at 92, which may indicate a slight decrease in the level of overlap or commonality in research topics between these two subperiods. In addition, the inclusion similarity index between the (2015–2018) and (2019–2022) subperiods decreases even further to 0.48, which suggests a reduction in the degree of overlap or similarity in the set of shared keywords between these two subperiods. This shows a further decrease in the level of continuity or similarity in research themes or topics between these two subperiods.

Overall, the shared keywords and inclusion similarity index values between the subperiods suggest a decreasing level of continuity or similarity in research themes or topics in the social anonymization field. This reduction between the subperiods may indicate a shift in research focus or divergence in scientific disciplines within the social anonymization field, which could have implications for interdisciplinary collaboration and knowledge transfer within this field.

In the ensuing subsections, we present a comprehensive analysis of the development of themes within the realm of social network anonymization by utilizing the evolution map and alluvial diagrams.

Evolution map SciMAT's evolution map analysis was used to examine the evolution of themes in the social network anonymization field over time. This map illustrates the primary topics and themes within the field and how they have transformed or progressed across different periods. The creation of this evolution map relies on the analysis of co-occurring networks of keywords, enabling us to identify the most pertinent research topics and their relationships. Therefore, through employing this map, we gained a better understanding of the field's development, identified emerging topics or research gaps, and explored potential future research directions.

Figure 26 illustrates the evolution of themes in the social network anonymization field. In this evolution map:

- A solid line indicates that either both themes have the same name or one of the themes shares a keyword with the other.
- A dashed line signifies that the themes share keywords that are not identical to the names of the themes.
- The width of the link between themes is directly proportional to the inclusion similarity index.
- The size of the spheres corresponds to the number of published core documents for each theme, meaning that a larger sphere indicates a higher number of published core documents.

This evolution map illustrates that during each subperiod, some newly emerged keywords represent new research topics or unexplored areas of study in the social network anonymization field. These keywords may indicate the introduction of new methods or techniques, novel research areas, or the participation of new authors in the field. For example, in the subperiod

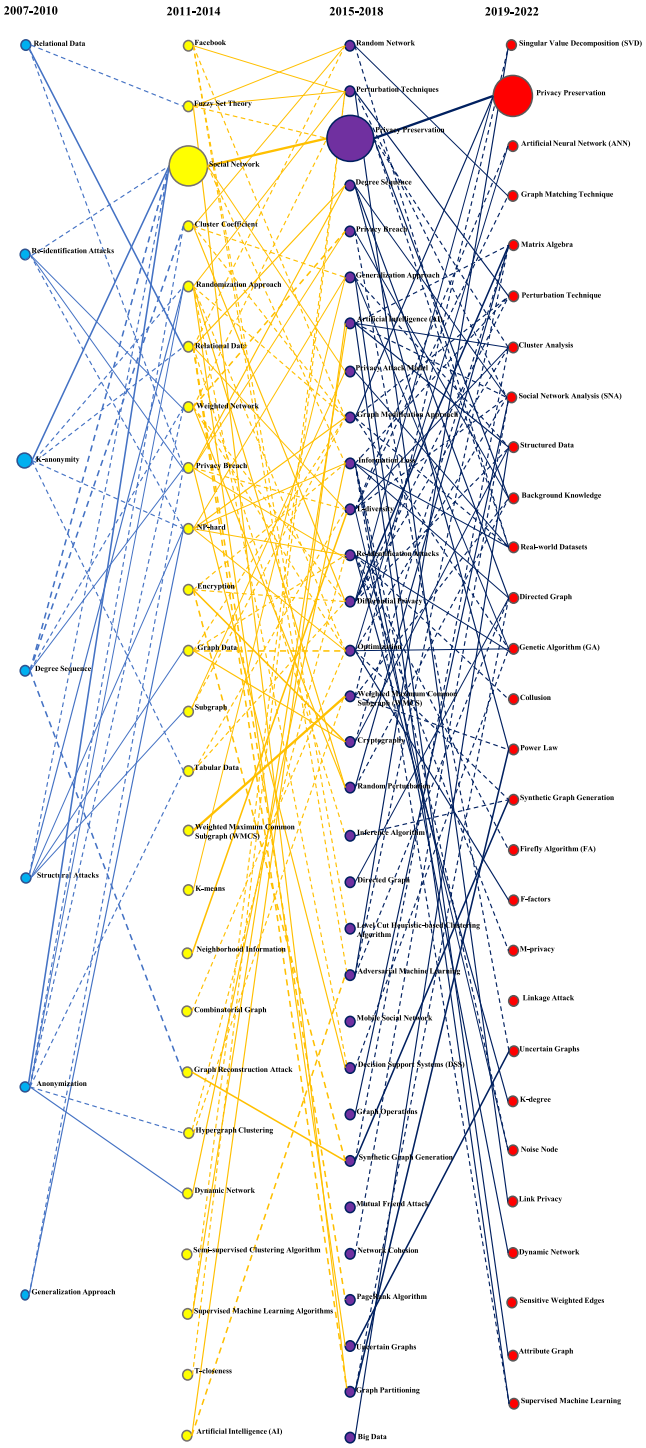


Fig. 26 The evolution map of the social network anonymization field

spanning 2011–2014, ten newly emerged keywords include "Artificial Intelligence (AI)," "T-closeness," "Supervised Machine Learning Algorithms," "Semi-supervised Clustering Algorithm," "Combinatorial Graph," "Neighborhood Information," "K-means," "Weighted Maximum Common Subgraph (WMCS)," "Encryption," and "Facebook."

The appearance of "Artificial Intelligence (AI)," "Supervised Machine Learning Algorithms," and "Semi-supervised Clustering Algorithm" as newly emerged keywords suggest that researchers are starting to explore the application of AI and machine learning techniques to address challenges of social network privacy. These algorithms could be used to identify patterns in social network data that reveal sensitive information, such as location data or personal relationships, and then obfuscate that information to protect individual privacy.

One newly emerged keyword, "T-closeness," is an anonymization model for relational data that aims to protect the privacy of social network users by ensuring that the distribution of data values in the anonymized dataset is similar to the distribution in the original dataset. This model has superiority over the two well-known K-anonymity and L-diversity anonymization models because it can deal with various types of attacks, such as homogeneity and background knowledge.

Another newly emerged keyword, "Combinatorial Graph" refers to a graph structure that can be used to model complex relationships between nodes in a social network. This keyword highlights the importance of using mathematical and computational tools to understand and solve the social network anonymization problem.

Moreover, the "Weighted Maximum Common Subgraph (WMCS)" is a graph matching algorithm that can be used to identify the largest subgraph that is common to two or more graphs. This algorithm is used in social network anonymization to identify and preserve significant structures (those with high total weights) when applying anonymization techniques.

"Neighborhood Information" is a newly emerged keyword that refers to the study of the relationships between the nodes and their neighbors in a social network. This information can be used to identify patterns in social network data and understand the structure of social networks. In the context of social network anonymization, the researchers can use this information to develop new anonymization methods that consider the relationships between nodes in a social network.

Also, "K-means" is a clustering algorithm that groups data points together based on their similarity. In the context of social network anonymization, K-means can be used to group nodes together based on their similarity (e.g., attributes or connections) to protect individual privacy.

Besides, "Encryption" is employed as a privacy protection mechanism for social network data to limit unauthorized access and keep users' information secure. However, since using encryption alone is insufficient, it must be paired with anonymization techniques to guarantee complete anonymity.

"Facebook" is a newly emerged keyword that likely refers to the growing interest in using this platform to study ways of social network anonymization to protect users' privacy.

Furthermore, in the 2015–2018 subperiod, the newly emerged keywords contain "Big Data," "Mutual Friend Attack," "Graph Operations," "Mobile Social Network," and "Directed Graph."

"Big Data" refers to the huge and complex sets of information that are generated by users in social networks and various electronic devices like smartphones, wearable tech, or home automation systems. These large datasets create significant challenges when it comes to keeping personal information private and anonymous.

In this subperiod, "Mutual Friend Attack" emerged, which is a type of attack that aims to re-identify social network users by exploiting information about their mutual friends or connections.

"Graph Operations" refers to the various computational and mathematical techniques used to analyze and manipulate social network graphs, such as algorithms for graph clustering, graph contraction, graph aggregation, graph pruning, and graph obfuscation.

The emergence of the "Directed Graph" in this subperiod shows a trend in studying the social network anonymization methods on this type of graph where relationships or interactions have a directional nature.

In the final subperiod, from 2019 to 2022, "Linkage Attack" emerged as a new theme. This novel keyword suggests that researchers are increasingly concentrating on addressing this type of attack in social network anonymization studies. During a linkage attack, an attacker could exploit the structural properties or patterns present in the anonymized network data, linking them to known properties or patterns in external data sources to re-identify a target.

It is worth noting that in different subperiods of the evolution map, there are some connections between clusters or individual nodes that are interesting to discuss and help understand the relationships and interactions among various research topics or themes over time.

The "Graph Modification Approach" was first introduced during the 2007–2010 period as a node of the "Re-identification Attacks" theme. The edge weight between these two keywords (representing the strength of the relationship between those keywords, calculated based on their co-occurrence in the articles) in the theme suggests that there is a significant overlap in the research conducted on these topics. Therefore, these two keywords have a significant tendency to appear together, indicating that graph modification approaches are frequently studied and used as a way to prevent re-identification attacks in the context of social network anonymization. This can help researchers understand the importance of developing robust graph modification techniques to enhance privacy protection in social networks. The strong relationship between these two keywords also highlights the ongoing challenge of balancing the need for privacy with the utility of the data as re-identification attacks continue to evolve and become more sophisticated. As a result, researchers working in this area should pay close attention to advances in both graph modification techniques and re-identification attack strategies to ensure that they are aware of the latest developments and can adapt their methods accordingly.

Furthermore, in the 2011–2014 subperiod, the "Graph Modification Approach" keyword appeared in the "NP-hard" theme. The edge weight between these keywords indicates a relatively low strength of association between them within the social network anonymization field. This means that although these terms are related, they may not frequently co-occur in the published papers, or their relationship is not as strong as other keyword pairs in the domain. In 2015–2018, the "Graph Modification Approach" keyword gained prominence within the social anonymization field, leading to the formation of a new theme centered around it with the same label. The edge weight between the "Graph Modification Approach" and "Edge Addition/Deletion" keywords suggests that the combination of graph modification approaches and edge addition/deletion techniques is a central theme in the field of social network anonymization. Researchers in this area focused on developing and improving methods that involve altering the structure of social network graphs, particularly by adding or removing edges between nodes, to protect user privacy while preserving the utility of the data for further analysis. Finally, in 2019–2022, the "Graph Modification Approach" keyword is placed in the "Privacy Preservation" theme, which reveals a growing interest in using this approach to protect privacy in social networks. The strong co-occurrence of this keyword with the "Structural Properties" and "Anonymization" keywords is also interesting. This strong

co-occurrence suggests that the "Graph Modification Approach" is a promising technique for privacy preservation and social network anonymization, particularly in the context of preserving structural properties while maintaining data usefulness.

The evolution map of the social network anonymization field shows that in 2007–2010, "Differential Privacy" was placed in the degree sequence theme, which is a mathematical property of networks. This suggests that the concept of differential privacy is linked to mathematical approaches to social network anonymization by limiting the risk of re-identification.

Furthermore, the keyword "Differential Privacy" has strong connections to other keywords, such as "Synthetic Graph Generation," "Privacy Preserving Data Mining (PPDM)," "Input Graph," and "Power Law." The strong connection between "Differential Privacy" and these other keywords in the evolution map suggests that researchers in the social network anonymization field during the 2007–2010 period were actively exploring the use of differential privacy techniques to create synthetic graphs, develop privacy-preserving data mining algorithms, and analyze the structural properties of input graphs (like degree sequences and power law distributions) without compromising privacy.

During the 2011–2014 subperiod, the evolution map of the social network anonymization field shows the "Differential Privacy" keyword as a part of the "Cluster Coefficient" cluster. This placement, along with its strong association to keywords like "Divide and Conquer Algorithm," "Laplace Noise," "Random Network," "Cluster Coefficient," and "Graph Mining," emphasizes the concentration of research and the interconnections among these keywords during that period. The prominent link between "Differential Privacy" and other keywords mentioned in the evolution map indicates that researchers in the social network anonymization domain were concentrating on incorporating differential privacy into graph mining activities while investigating the connection between privacy and structural characteristics like cluster coefficients. For instance, the scientists developed divide-and-conquer algorithms for effective graph analysis on random networks and employed Laplace noise to ensure privacy while analyzing the mentioned networks.

During the 2015–2018 period of the evolution map, the "Differential Privacy" keyword emerged as a theme. This signifies the increasing prominence of differential privacy as a central research topic in the mentioned subperiod. The keyword also exhibits strong connections to other keywords, such as "Topological Information," "Persistent Homology," "Laplace Noise," "Matrix Algebra," "Random Matrix," "Correlation Matrix," "Personalization," "Social Recommendations," and "Shortest Path," which highlights the interdisciplinary nature of research during this period. The strong connections between "Differential Privacy" and the above-mentioned keywords suggest that researchers were working on various aspects of differential privacy in social networks. Such methods include applying differential privacy to protect topological information, exploring advanced mathematical techniques like persistent homology, using matrix algebra and random matrices to analyze network properties, and incorporating privacy-preserving techniques into personalized social recommendations and shortest path algorithms.

In the 2019–2022 subperiod, the "Differential Privacy" keyword is placed in the "Privacy Preservation" theme, indicating the maturation of differential privacy as a well-established technique in the context of privacy preservation in social networks. Its strong connection to various other keywords, such as "Privacy Preservation," "Structural Privacy," "Social Network," and "Anonymization," underscores the continued relevance and focus on these topics during this period. The strong connections between "Differential Privacy" and these keywords in this subperiod suggest that researchers in the social network anonymization field were focusing on integrating differential privacy with other privacy preservation techniques to achieve better structural privacy in social networks. This includes exploring new methods

and techniques based on differential privacy for anonymizing social networks, as well as studying the trade-offs between privacy and utility in various anonymization approaches.

Based on the provided evolution map and during the 2007–2010 subperiod, the "Generalization Approach" keyword is placed in its own theme. This indicates that during this time, generalization approaches were a distinct and significant research focus within the field. However, based on the strategy map diagram of the mentioned subperiod, the "Generalization Approach" theme had limited importance and connectivity with other research themes, suggesting that it might have been underdeveloped or overlooked in comparison to other themes during that time. The strong connection between "Generalization Approach" and other keywords, such as "Eigenvector Centrality," "Split Algorithms," and "Decision Support Systems (DSS)," highlights the relationships and interdisciplinary nature of research during this period. The strong connection between "Generalization Approach" and the mentioned keywords in the evolution map suggests that researchers in the social network anonymization field were exploring the use of generalization approaches to anonymize social network data while considering network properties like eigenvector centrality. They also developed and applied various split algorithms, such as union-split algorithms, to improve the efficiency of generalization processes, specifically in the context of graph partitioning and distributed privacy-preserving techniques. Besides, the strong relationship between the "Generalization Approach" and "Decision Support Systems (DSS)" demonstrates that researchers conducted some studies to investigate how they can use the generalization approaches to ensure privacy preservation in decision support systems.

In the 2011–2014 subperiod, the "Generalization Approach" keyword is placed in the "NP-hard" theme. This placement indicates that during this period, the research focus related to generalization approaches in social network anonymization was strongly connected to NP-hard problems and optimization challenges. The connections to the "NP-hard," "Optimization," and "Information Loss" keywords highlight the relationships between these topics during this period. The connections between "Generalization Approach" and the keywords mentioned in this subperiod suggest that researchers in the social network anonymization field were focusing on the challenges associated with solving NP-hard problems related to generalization approaches. They were also exploring optimization techniques to minimize information loss while preserving privacy and dealing with the computational complexities of these problems.

In the 2015–2018 subperiod, the "Generalization Approach" keyword is once again placed in the "Generalization Approach" theme, suggesting that this research theme remained relevant and active during this period. The "Generalization Approach" keyword is also connected to various other keywords, including "K-means," "Node Degree," and "Suppression." The connection to "K-means" indicates that researchers explored the use of clustering algorithms, like K-means, for generalization approaches in social network anonymization. The connection to "Node Degree" suggests that researchers may have been focusing on preserving the degree distribution of nodes in the network while generalizing it for privacy protection. The connection to "Suppression" implies that researchers might have been exploring the use of suppression techniques to protect the users' attribute privacy by suppressing their sensitive data in the network.

During the 2019–2022 subperiod in the evolution map of the social network anonymization field, the "Generalization Approach" keyword is placed in the "Genetic Algorithm (GA)" theme. This indicates that during this period, researchers explored the use of genetic algorithms to improve the efficiency and effectiveness of generalization approaches in social network anonymization. The "Generalization Approach" keyword is also connected to other

keywords, including "Particle Swarm Optimization Algorithm (PSO)" and "Hybrid Algorithms." The connection to the "Particle Swarm Optimization Algorithm (PSO)" suggests that researchers used swarm intelligence-based optimization techniques to improve generalization approaches in social network anonymization. The connection to "Hybrid Algorithms" indicates that researchers combined genetic algorithms and other optimization techniques (e.g., PSO) to create more powerful and flexible algorithms for social network anonymization. As a result, the use of genetic algorithms, swarm intelligence, and hybrid algorithms highlight the interdisciplinary nature of research in this field and the importance of leveraging techniques from diverse fields to tackle complex problems.

While the "Uncertain Graphs" keyword was not present in the evolution map of the 2007–2010 subperiod, it does appear the 2011–2014 subperiod for the first time and is placed in the "Randomization Approach" theme. This indicates that during this period, the researchers tried to explore new ways of addressing the challenges associated with applying an uncertain graph approach in social network anonymization.

In the 2015–2018 subperiod, the "Uncertain Graphs" keyword is presented as a theme. This indicates that during this period, researchers investigated the challenges and opportunities associated with using uncertain graphs as an approach to anonymize the social network and focused on this topic as a distinct and important research theme. Additionally, the "Uncertain Graphs" keyword has a strong connection to the "Maximizing Variance" keyword. Variance maximization involves maximizing the variance of the uncertain graph by selecting or perturbing its edges or weights. This can help to improve the accuracy of analyses performed on the graph while preserving privacy.

During the 2019–2022 subperiod, the "Uncertain Graphs" keyword is once again placed in the "Uncertain Graphs" theme. This suggests that during this period, researchers continued to focus on uncertain graphs as an important and distinct research theme in social network anonymization. The "Uncertain Graphs" keyword is also connected to the "Node Characteristics" and "Triadic Closure" keywords in the theme. The focus on node characteristics and triadic closures highlights the importance of understanding the structural properties and dynamics of social networks to propose novel uncertain graph approaches for anonymizing social networks effectively.

Furthermore, based on the proposed evolution map of the social network anonymization field, we aim to identify connections between clusters or individual nodes across different subperiods. This can help researchers to understand the relationships and interactions among different research topics and themes over time.

As shown in the evolution map, the "K-anonymity" theme emerged for the first time in the 2007–2010 subperiod. This means that the researchers tried to apply the previously defined anonymization models used for relational and structured datasets to the graph data, which is unstructured data. They proposed various social network anonymization models to deal with the structural attacks based on the K-anonymity model along with different graph structural properties, such as degree sequence, graph isomorphism, graph automorphism, and centrality criterion. Subsequently, between 2011 and 2014, researchers focused on defeating the shortcomings of K-anonymity-based models in combating re-identification and neighborhood attacks. They introduced the "L-diversity" anonymization model to social network anonymization, which led to the development of various models. Furthermore, during this time, researchers also started to propose social network anonymization models based on the "T-closeness" anonymization model to tackle the limitations of L-diversity-based models.

The proposed evolution map highlights a significant connection pattern involving "Generalization Approach," "NP-hard," "Optimization," "Heuristics," and "Meta-heuristics." Figure 26 demonstrates a link between the "Generalization Approach" theme in the

2007–2010 subperiod and the "NP-hard" theme in the 2011–2014 subperiod. The "NP-hard" theme encompasses keywords like "Optimization," "Generalization Approach," and "Information Loss," and there is a robust relationship among these keywords within this cluster. That is, in this subperiod, the researchers' focus shifted to tackling NP-hard problems related to generalization approaches during that time. This connection suggests that as researchers explored generalization approaches in the earlier subperiod, they discovered that some problems in the field were NP-hard, which means they are computationally difficult to solve in an efficient manner. An example of this is identifying the optimal node clusters in the generalization approach that result in minimal information loss. Consequently, the research emphasis transitioned to discovering optimization methods, heuristics, and meta-heuristics to tackle these intricate issues more effectively.

In the following subperiod of the evolution map, 2015–2018, the "NP-hard" keyword is associated with the "Re-identification Attacks" theme. This association emphasizes the intricate nature of some re-identification problems and underscores the necessity of creating robust and efficient anonymization methods to counter these attacks. The evolution map illustrates the importance of comprehending the computational obstacles inherent in re-identification attacks and the unceasing efforts within the anonymization discipline to address these challenges. Besides, in this subperiod, the keyword "Optimization" arises as a distinct theme in the context of social network anonymization. This theme demonstrates the growing importance of optimization techniques in addressing privacy and anonymization challenges in social networks. Within the optimization theme, several related keywords are connected, such as "Genetic Algorithm (GA)," "Heuristics," and "Combinatorial Optimization." These keywords represent different optimization methodologies that have been employed in the field to improve anonymization techniques. Additionally, these optimization-related keywords have external connections with other relevant keywords from other themes. For instance, "Clustering Algorithms," "Graph Modification Approach," and "K-anonymity" are linked to the optimization theme. These external links signify the interdisciplinary nature of social network anonymization research and reveal that optimization techniques are being integrated with other approaches to enhance privacy protection. The emergence of the "Optimization" theme, along with its connections to other keywords and themes, highlights the increasing role of optimization techniques in the field of social network anonymization. This trend highlights the ongoing efforts to develop more effective and efficient methods for protecting users' privacy in social networks by leveraging various optimization strategies.

In the last subperiod, the "Optimization" keyword appears within the "Artificial Neural Network (ANN)" theme, signifying the increasing use of optimization techniques in the context of ANNs for social network anonymization. The keyword has connections with keywords like "Cuckoo Optimization Algorithm (COA)," "Graph Neural Network (GNN)," "Backpropagation Algorithm," and "High Degree Nodes," illustrating the diverse optimization approaches being integrated with neural networks to address privacy challenges in social networks.

Besides, in this subperiod, the emergence of "Genetic Algorithm (GA)" as a theme, along with its connections to related keywords like "Particle Swarm Optimization Algorithm (PSO)," "Hybrid Algorithms," "Generalization Approach," and "Edge Addition/Deletion," highlights the growing importance of optimization techniques in social network anonymization field. The diverse connections between GA and these keywords demonstrate the ongoing efforts to develop more effective privacy-preserving methods by integrating GA with various meta-heuristics, generalization techniques, and graph modification approaches to address privacy challenges in social networks.

Also, the placement of "Combinatorial Optimization" within the "Synthetic Graph Generation" theme, along with its external connections to "Differential Privacy" and "Machine Learning" keywords, illustrates the diverse applications of combinatorial optimization in the field of social network anonymization. Emphasizing its relevance in generating synthetic graphs that maintain user privacy, combinatorial optimization techniques are applied in conjunction with differential privacy mechanisms and machine learning-based approaches to develop more effective privacy-preserving solutions. These interdisciplinary connections highlight the ongoing efforts to enhance privacy protection in social networks by integrating combinatorial optimization with various techniques and methodologies.

The emergence of the "Firefly Algorithm (FA)" as a theme in social network anonymization is notable in this subperiod. As a meta-heuristic algorithm, it has the potential for complex optimization problems. The "Firefly Algorithm (FA)" theme is connected to other important concepts, such as "Fuzzy Clustering" and "Identity Disclosure." The connection between "Fuzzy Clustering" and the "Firefly Algorithm (FA)" theme implies that the algorithm was applied in optimizing fuzzy clustering techniques to enhance the effectiveness of social network anonymization. Also, the connection between "Identity Disclosure" and the "Firefly Algorithm (FA)" theme suggests that the algorithm was used to mitigate the risk of identity disclosure by optimizing anonymization techniques that balance data utility and privacy preservation. Additionally, it has external links with relevant keywords, including "Optimization," "K-anonymity," "Information Loss," and "Clustering Algorithms." These connections highlight its relevance in privacy preservation and optimizing social network anonymization techniques.

In the current subperiod, the "Neighborhood Attraction Firefly Algorithm (NAFA)" keyword is associated with the "F-factors" theme in social network anonymization research. This algorithm, which belongs to the meta-heuristic optimization category, is utilized to optimize privacy-preserving techniques within the field. Moreover, external connections to other keywords, such as "Graph Modification Approach," "Structural Properties," "Utility," and "Anonymization," suggest potential applications of the NAFA algorithm in enhancing privacy through graph modification, maintaining the structural properties of social networks during anonymization, optimizing the balance between privacy protection and data utility, and improving the overall anonymization process. Taken together, these connections highlight the importance of the NAFA algorithm in optimizing social network anonymization techniques and protecting user privacy.

Another intriguing pattern in the provided evolution map is the "Artificial Intelligence (AI)" keyword, which first emerged as a theme within the subperiod of 2011–2014. This theme exhibits a connection to the "Machine Learning" keyword, indicating a strong relationship between these two concepts. Additionally, the "Artificial Intelligence (AI)" theme displays external connections to other themes in the field. Specifically, there are links to the keywords "Anonymization," "K-anonymity," and "Clustering Algorithms," suggesting potential areas of intersection between these keywords. This pattern provides valuable insight into the evolving trends and themes within the field, highlighting the emergence of AI as a central concept in the 2011–2014 subperiod. The connections between "Artificial Intelligence (AI)" and "Machine Learning," as well as other key themes, suggest potential avenues for further exploration and research in the field.

In the subsequent subperiod, 2015–2018, the "Artificial Intelligence (AI)" keyword has already established itself as a prominent theme. Notably, this keyword exhibits connections to several other keywords, including "Adaptive Random Walk," "Supervised Machine Learning Algorithms," "Friendship Attack," and "DBSCAN Algorithm." These connections suggest potential overlap and intersection between different subthemes within the field. Specifically,

the link to "Adaptive Random Walk" highlights its potential utility as an AI-based approach in the context of social network anonymization. The connection to "Supervised Machine Learning Algorithms" points to the ongoing importance of machine learning in developing AI-based privacy-preserving methods. Additionally, the links to "Friendship Attack" and "DBSCAN Algorithm" suggest potential research areas related to privacy and security in the context of social networks.

Moreover, the "Machine Learning" keyword is placed within the "Adversarial Machine Learning" theme, reflecting the growing importance of security and privacy in developing machine learning algorithms for social network anonymization. This keyword exhibits connections to several other keywords, including "Privacy Preservation," "Supervised Machine Learning," and "Social Relationships." The links between "Machine Learning" and these keywords provide valuable insights into the ongoing trends and themes within the field of social network anonymization. They highlight the importance of protecting social network users' privacy through anonymization techniques that utilize machine learning algorithms. Additionally, it is crucial to keep user data private when analyzing social networks using machine learning algorithms. The connections also emphasize the importance of supervised learning algorithms in developing machine learning algorithms for social network anonymization.

In the next subperiod, 2019–2022, the "Artificial Intelligence (AI)" keyword is placed within the "Structured Data" theme. This placement suggests a shift toward using AI techniques to analyze and manage structured data in the context of social network anonymization. Within this theme, the "Artificial Intelligence (AI)" keyword has connections to several other keywords, including "Assortativity," "Sequential Clustering," and "Node Degree." These connections suggest potential research areas related to the use of AI techniques for identifying patterns and structures within social network data. Furthermore, the "Artificial Intelligence (AI)" keyword has external connections to several other themes in the field, including "Hierarchical Clustering," "K-degree," "Information Loss," "Anonymization," "Differential Privacy," and "Clustering Algorithms." These connections suggest that there is growing interest in using AI techniques for addressing issues related to privacy and security in the context of social network anonymization, as well as developing new clustering and anonymization algorithms.

Moreover, the term "Machine Learning" is positioned within the "Matrix Algebra" theme. This keyword is associated with concepts like "Adaptive Random Walk," "Generative Adversarial Network (GAN)," "Feature Learning," "Random Projection Algorithm," and "Random Perturbation" within the specified theme. Additionally, it has external connections to "Differential Privacy," "Structural Properties," "Synthetic Graph Generation," "Graph Generation Model," and "Randomization Approach." The connection among machine learning techniques and matrix algebra, adaptive random walks, GANs, feature learning, random projection algorithms, and Random Perturbation shows that the researchers used them together to protect user privacy while preserving the overall graph structure. These techniques have external connections to concepts like differential privacy, structural properties, synthetic graph generation, graph generation models, and randomization approaches, which together contribute to creating synthetic social network graphs that maintain privacy and data utility for analysis.

In the current subperiod, the "Adversarial Machine Learning" keyword is placed within the "Singular Value Decomposition (SVD)" theme. It connects to "Structural Attacks," "Matrix Decomposition," and "Markov Clustering (MCL)" keywords within this theme. The mentioned connections can be interpreted as follows. In the context of social network anonymization, understanding structural attacks helps researchers develop strategies to counteract them using adversarial machine learning approaches, thereby strengthening the privacy

of the anonymized network. In adversarial machine learning for social network anonymization, matrix decomposition techniques like SVD can be employed to process and analyze complex graph data while mitigating the impact of adversarial perturbations. Besides, in the context of adversarial machine learning, MCL can help develop robust privacy-preserving methods that can withstand attacks aimed at exploiting community structures in social networks.

Additionally, the "Adversarial Machine Learning" links externally to keywords, such as "Graph Modification Approach," "Differential Privacy," "Clustering Algorithms," "Graph Isomorphism," "Fuzzy Set Theory," and "Neuro Fuzzy." These connections demonstrate the key role of adversarial machine learning in protecting user privacy and maintaining data utility for analysis in social network anonymization.

It is noteworthy that the term "Supervised Machine Learning" emerged as a distinct theme during the 2019–2022 subperiod. This concept is intrinsically linked to the "Health Information" keyword, highlighting its applicability in the privacy-preserving analysis of health-related social networks. Furthermore, it exhibits external connections with keywords like "Clustering Algorithms," "Artificial Neural Network (ANN)," "Anonymization," "K-anonymity," "Graph Neural Network (GAN)," and "Backpropagation Algorithm." These connections demonstrate the interdisciplinary nature of supervised machine learning in addressing the challenges of privacy preservation and data utility within social network anonymization. In other words, these external connections highlight the versatility of supervised machine learning and its ability to contribute to various aspects of social network anonymization, ranging from privacy preservation and data utility to advanced modeling and optimization.

Alluvial diagram In this subsection, an alluvial diagram is presented to show the flow or transition of keywords between different categories. In the context of this study on social network anonymization, the alluvial diagram demonstrates how the authors' keywords have changed over time and how they relate to each other. The diagram consists of a series of vertical bars or columns, each representing one of the four subperiods that was defined previously. The columns are connected by horizontal lines, and the width of the lines represents the number of keywords that are shared between adjacent subperiods. The PageRank score of each keyword is calculated in each block, demonstrating the importance of nodes (keywords) within the network. (All the keywords in each module, along with their PageRank scores, are presented in Appendix C.) By analyzing the PageRank scores of the keywords within each block, we gained insight into the most influential themes and concepts in the social network anonymization field, as well as an understanding the relationships and connections between these themes. It is worth noting that by analyzing the alluvial diagram in Fig. 27, insight into the evolution of the focus of research on social network anonymization over time is gained. It helps us understand which keywords have remained consistently popular throughout all four subperiods that have become less common over time and which have emerged as new focus areas in more recent years. Also, it is helpful in the way that how different keywords are related to each other and how they cluster together into more prominent themes or topics.

From the depicted alluvial diagram, some of the most prominent keywords of the social network anonymization field, along with their PageRank scores, are presented in Table 7. In the following paragraphs, these keywords are analyzed regarding their positions and relations in the alluvial diagram.

From Fig. 27, it is clear that the keyword "Social Network" consistently remains in module 1 throughout all four subperiods, though with fluctuations in its PageRank score. In the 2007–2010 subperiod, it had a PageRank score of 0.00206 in 1 out of 14 modules, indicating

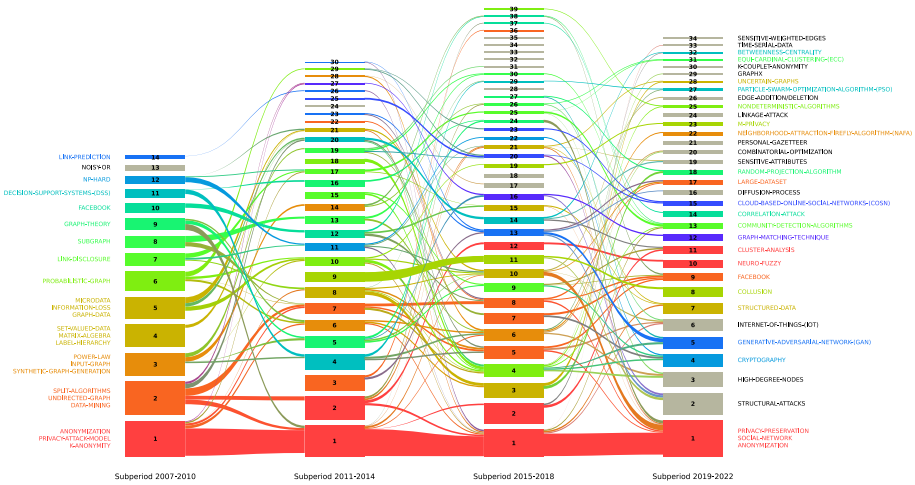


Fig. 27 The alluvial diagram of the social network anonymization field

relevance but limited influence. By 2011–2014, the score rose significantly to 0.0201 in 1 out of 30 modules, reflecting its increasing importance due to the growing awareness of social network privacy concerns related to the rapid growth of platforms, such as Facebook and Twitter, during the period. In the 2015–2018 subperiod, the score slightly declined to 0.0152 in 1 out of 39 modules, likely due to the emergence of newer privacy concerns such as machine learning. In the final subperiod, 2019–2022, the PageRank further dropped to 0.0137 in 1 out of 34 modules, signaling ongoing relevance but decreasing prominence. Throughout all subperiods, "Social Network" consistently appears in module 1, illustrating its sustained importance in the social network anonymization field, with a notable rise in influence during the second subperiod followed by a gradual decline. Nevertheless, it remains the base of the field and serves as a central theme in social network anonymization research. This demonstrates that while new techniques emerge, the foundational concept of social networks continues to anchor much of the work in this domain.

Furthermore, the alluvial diagram shows that the keyword "Privacy Preservation" consistently remained in module 1 across all four subperiods, with slight fluctuations in its PageRank score. In the 2007–2010 subperiod, it had a PageRank score of 0.0121 in 1 out of 14 modules, indicating significant influence as privacy concerns became central in social network research. By 2011–2014, its PageRank increased to 0.0191 in 1 out of 30 modules, reflecting the growing focus on safeguarding user data in response to evolving threats such as re-identification attacks and structural vulnerabilities in social networks. In the 2015–2018 subperiod, the score slightly decreased to 0.0147 in 1 out of 39 modules, signaling its continued relevance but reduced centrality, possibly due to the diversification of privacy research toward more specialized techniques, such as Differential Privacy and Graph Modification. In the final subperiod, 2019–2022, the PageRank marginally increased to 0.0152 in 1 out of 34 modules, suggesting renewed interest as privacy concerns grew more complex with the rise of decentralized platforms and AI-driven attacks. Overall, "Privacy Preservation" consistently remained in module 1 throughout all subperiods, underscoring its pivotal role in social network anonymization research despite minor fluctuations in its prominence.

Table 7 Some of the most prominent keywords' module placements and PageRank scores regarding the alluvial diagram

Keyword name	2007–2010		2011–2014		2015–2018		2019–2022	
	Module	PageRank	Module	PageRank	Module	PageRank	Module	PageRank
Social network	1	0.0026	1	0.0201	1	0.0152	1	0.0137
Privacy preservation	1	0.0121	1	0.0191	1	0.0147	1	0.0152
Anonymization	1	0.0307	1	0.0150	1	0.0112	1	0.0109
K-anonymity	1	0.0197	1	0.0101	1	0.00802	1	0.00616
Clustering algorithms	2	0.00	2	0.00987	1	0.00689	1	0.00608
Differential privacy	3	0.00	5	0.00979	24	0.00607	1	0.00719
Generalization approach	2	0.0166	2	0.00830	21	0.00401	1	0.00436
L-diversity	–	–	20	0.00526	14	0.00579	11	0.00503
Graph modification approach	3	0.00	10	0.00516	10	0.00532	1	0.00714
Randomization approach	7	0.00410	8	0.00750	3	0.00569	18	0.00449
K-degree	–	–	11	0.00442	15	0.00478	7	0.00310
Perturbation techniques	–	–	7	0.0102	8	0.00666	1	0.00623
Uncertain graphs	–	–	1	0.00170	21	0.00361	28	0.00325
Machine learning	–	–	30	0.00143	13	0.00568	5	0.00797
Genetic algorithm (GA)	–	–	1	0.00382	6	0.00485	27	0.00307
Artificial intelligence (AI)	–	–	30	0.00249	1	0.00691	7	0.00538
Optimization	–	–	6	0.00683	6	0.00602	3	0.00864

Besides, the keyword "Anonymization" consistently appears in module 1 throughout the four subperiods, though its PageRank score gradually declines over time. In the 2007–2010 subperiod, it had a PageRank score of 0.0307 in 1 out of 14 modules, indicating its significant influence and centrality in the field during the early stages of social network anonymization research. This reflects the field's initial focus on foundational anonymization techniques, which were pivotal as privacy concerns began to rise with the increased usage of social networks. By 2011–2014, the PageRank score decreased to 0.0150 in 1 out of 30 modules, likely due to the shift toward more specific methods like L-diversity and Differential Privacy, which addressed particular challenges in anonymization leading to a more granular focus on specialized solutions rather than the broader, overarching concept of anonymization. In the 2015–2018 subperiod, the keyword remained in module 1 out of 39 modules, though its PageRank further dropped to 0.0112, indicating a continued reduction in prominence as research diversified into more specialized privacy solutions. By the 2019–2022 subperiod, the PageRank score declined slightly to 0.0109 in 1 out of 34 modules, reflecting the growing complexity of privacy issues, such as AI-driven attacks and synthetic data risks. Despite this steady decline in influence, the keyword "Anonymization" remains consistently present in module 1, highlighting its lasting relevance and connection to key concepts in social network anonymization research.

Moreover, the alluvial diagram shows that the keyword "K-anonymity" consistently appeared in module 1 throughout the four subperiods, though its PageRank score steadily declined. In the 2007–2010 subperiod, "K-anonymity" was in module 1 out of 14 modules with a PageRank score of 0.0197, reflecting its significant influence as a primary method for ensuring data privacy during the early stages of social network anonymization research. By 2011–2014, the PageRank score dropped to 0.0101 in 1 out of 30 modules, as newer techniques, such as L-diversity, began to address privacy challenges that K-anonymity could not, such as vulnerability to homogeneity and background knowledge attacks. In the 2015–2018 subperiod, its score further decreased to 0.00802 in 1 out of 39 modules, indicating a reduced role but still serving as a baseline comparison for more advanced techniques. In the 2019–2022 subperiod, K-anonymity remained in module 1 with a slightly lower score of 0.00616 in 1 out of 34 modules, maintaining its relevance despite its declining influence. Overall, "K-anonymity" continues to be a foundational concept in the field of social network anonymization, recognized for its historical significance, though more advanced techniques have taken precedence.

It is worth mentioning that, the keyword "Clustering Algorithms" shows significant shifts in both module placement and PageRank score across the four subperiods, reflecting its evolving role in social network anonymization research. In the 2007–2010 subperiod, it was in module 2 out of 14 modules with a PageRank score of 0.0, indicating minimal influence and relevance at this early stage. Its low score suggests that the technique had yet to gain traction or establish strong connections with the primary themes of that period, likely because more fundamental approaches to social network anonymization were still being explored. In the 2011–2014 subperiod, the keyword remained in module 2 but saw a notable increase in its PageRank score to 0.00987, signaling its growing importance and influence within the field. This rise suggests that clustering algorithms had begun to gain recognition as valuable tools for anonymization, likely due to their ability to group similar nodes and reduce the granularity of network data, thereby helping to protect privacy. By 2015–2018, "Clustering Algorithms" transitioned to module 1 out of 39 modules, though its PageRank score slightly declined to 0.00689, reflecting its alignment with more central research themes, despite the slight reduction in influence. In the 2019–2022 subperiod, the keyword stayed in module 1 but with a marginally lower PageRank score of 0.00608, suggesting that while newer

methods emerged, "Clustering Algorithms" remained relevant. The consistent presence of this keyword across subperiods demonstrates its lasting utility, especially in approaches that focus on grouping data while preserving privacy. Overall, "Clustering Algorithms" has grown from a peripheral to a central technique, maintaining its significance in the evolving research landscape despite minor fluctuations in prominence.

Also, the keyword "Differential Privacy" demonstrates significant shifts in both module placement and PageRank score across the four subperiods, reflecting its evolving role and growing relevance in social network anonymization research. In the 2007–2010 subperiod, it was positioned in module 3 out of 14 modules with a PageRank score of 0.0, indicating that it had not yet emerged as a prominent concept, as foundational approaches dominated the field. By the 2011–2014 subperiod, "Differential Privacy" moved to module 5 out of 30 modules, with a notable rise in its PageRank score to 0.00979, signaling its growing recognition as an important tool for protecting privacy, particularly as researchers began to adopt more mathematically rigorous solutions in response to increasing privacy concerns. In the 2015–2018 subperiod, "Differential Privacy" shifted to module 24 out of 39 modules, with a slightly lower PageRank score of 0.00607. This decline in score and change in module placement reflect a thematic shift, possibly due to the integration of differential privacy with other advanced methods, such as machine learning and AI, which began to dominate the research landscape. In the 2019–2022 subperiod, "Differential Privacy" returned to module 1 out of 34 modules, with an increased PageRank score of 0.00719, highlighting its renewed prominence as a central concept in response to emerging data privacy challenges. This final shift suggests that differential privacy had become part of a more central cluster of keywords, reflecting its growing importance in the evolving research landscape. Overall, "Differential Privacy" has transitioned from a peripheral concept to a central pillar in social network anonymization, showcasing its enduring relevance and adaptability in addressing the field's increasingly complex privacy concerns.

Further, the keyword "Generalization Approach" has experienced notable shifts in both module placement and PageRank score across the four subperiods, reflecting changes in its relevance within the social network anonymization field. In the 2007–2010 subperiod, it was positioned in module 2 out of 14 modules with a PageRank score of 0.0166, indicating that generalization was a relatively important approach during the early stages of research. Its prominent role at this time stemmed from the widespread use of k-anonymity methods. In the 2011–2014 subperiod, the keyword remained in module 2 out of 30 modules, but its PageRank score dropped to 0.00830, suggesting a decline in influence as newer, more robust methods, such as differential privacy, began to take center stage, though generalization continued to be relevant. By the 2015–2018 subperiod, the keyword moved to module 21 out of 39 modules, with a further decrease in PageRank to 0.00401, signaling a continued decline in centrality. The generalization approach had become more peripheral, possibly due to its limitations, such as reduced data utility and vulnerability to sophisticated attacks, such as background knowledge attacks. In the 2019–2022 subperiod, "Generalization Approach" reappeared in module 1 out of 34 modules, with a slightly increased PageRank score of 0.00436, indicating a modest resurgence in relevance, though it did not regain its earlier prominence. Its placement in module 1 suggests that it remains connected to certain research clusters, possibly in combination with other modern techniques. In summary, the "Generalization Approach" has seen a decline in influence over time, with decreasing PageRank scores and module shifts reflecting its reduced prominence. However, the slight increase in the final subperiod suggests a renewed, albeit modest, interest in the approach, potentially through integration with more advanced anonymization methods.

The alluvial diagram also shows that the keyword "L-diversity" emerged and underwent changes in both module placement and PageRank score across the four subperiods. In the 2011–2014 subperiod, it first appeared in module 20 out of 30 modules with a PageRank score of 0.00526, indicating its relevance but not as a central concept. This period marked the initial rise of L-diversity as an improvement over K-anonymity, particularly in addressing background knowledge attacks by ensuring diversity within sensitive attributes in anonymized groups. In the 2015–2018 subperiod, "L-diversity" shifted to module 14 out of 39 modules with a slightly higher PageRank score of 0.00579, reflecting a modest increase in influence as its applications were further explored. However, its limitations, particularly with complex privacy risks, became more apparent during this time, leading to a focus on more advanced methods. By the 2019–2022 subperiod, the keyword moved to module 11 out of 34 modules, with a slightly lower PageRank score of 0.00503, signaling continued relevance but declining prominence. This decline was likely driven by the growing adoption of t-closeness, which offered stronger protection against background knowledge attacks by ensuring that sensitive attribute distributions in anonymized groups resembled the overall data distribution. Overall, while "L-diversity" has remained relevant throughout the subperiods, its influence has gradually declined, particularly as more robust alternatives like t-closeness emerged to address its limitations.

With respect to the "Graph Modification Approach," it is apparent that across the four subperiods, the module placement and PageRank score of this keyword underwent considerable changes. The keyword "Graph Modification Approach" experienced minimal influence, as reflected by its placement in module 3 out of 14 modules with a PageRank score of 0.0 during the 2007–2010 subperiod. At this early stage, the approach had not gained much traction in social network anonymization. However, by the 2011–2014 subperiod, the keyword moved to module 10 out of 30 modules, with its PageRank score rising to 0.00516, signaling growing interest as researchers began recognizing its potential in handling privacy threats like re-identification attacks. In the 2015–2018 subperiod, the keyword remained in module 10 out of 39 modules, with a slight PageRank increase to 0.00532, indicating sustained relevance, especially as more advanced variations, such as Constrained Perturbation and noise addition were developed. By the 2019–2022 subperiod, the keyword transitioned to module 1 out of 34 modules, with its PageRank climbing to 0.00714, placing it at the center of research due to its flexibility and effectiveness in balancing privacy and utility in large-scale, dynamic networks. Overall, the "Graph Modification Approach" evolved from a peripheral concept to a central role in addressing modern privacy challenges, as reflected in its increasing prominence across the subperiods.

Additionally, the alluvial diagram shows that the keyword "Randomization Approach" experienced significant shifts in both module placement and PageRank score across the four subperiods. In the 2007–2010 subperiod, it was positioned in module 7 out of 14 modules, with a PageRank score of 0.00410, indicating moderate influence. While not a dominant theme, it remained relevant as an early privacy-preserving technique. In the 2011–2014 subperiod, the keyword moved to module 8 out of 30 modules, with an increased PageRank score of 0.00750, reflecting its growing importance as randomization gained attention for anonymizing social networks with relatively simple techniques that preserved data utility. By the 2015–2018 subperiod, "Randomization Approach" shifted to module 3 out of 39 modules, though its PageRank score slightly decreased to 0.00569. This shift to a more central cluster of keywords indicates that while the method remained relevant, its influence began to decline with the emergence of more sophisticated techniques, such as differential privacy, which offered stronger protection against re-identification attacks. Despite this, randomization continued to be valued for its simplicity and ease of implementation. In the

2019–2022 subperiod, the keyword transitioned to module 18 out of 32 modules, with a further reduced PageRank score of 0.00449. This decline reflects the decreasing prominence of randomization-based techniques as more robust anonymization methods gained traction. The move to a lower-ranked module suggests that randomization, while still useful in specific scenarios, has become less central due to its limitations, such as vulnerability to sophisticated attacks and weaker privacy guarantees. Overall, the "Randomization Approach" experienced fluctuations in relevance, peaking in the early 2010s before gradually declining as the field advanced. Nonetheless, it remains a topic of interest due to its simplicity and effectiveness in certain cases.

In addition, the alluvial diagram illustrates that the "K-degree" keyword experienced significant changes in both module placement and PageRank score over the four subperiods. In the 2007–2010 subperiod, "K-degree" had not yet emerged, indicating its absence as a prominent theme in the social network anonymization domain. However, in the 2011–2014 subperiod, the keyword surfaced in module 11 out of 30 modules, with a PageRank score of 0.00442. This suggests that "K-degree" became relevant as researchers began focusing on node degree methods for anonymization, although it had not yet gained substantial influence given its lower-ranked placement. During the 2015–2018 subperiod, the "K-degree" keyword transitioned to module 15 out of 39 modules, with a slightly higher PageRank score of 0.00478. This indicates growing attention to K-degree approaches, likely due to their simplicity and computational efficiency for anonymizing social networks. In the 2019–2022 subperiod, the keyword moved to module 7 out of 32 modules, reflecting a higher-ranked position, but with a reduced PageRank score of 0.00310. This decline in influence may stem from the recognition of the limitations of K-degree-based anonymization, such as its challenges in handling larger, more complex networks and its vulnerability to advanced attacks, such as inference or structural attacks. Overall, while the "K-degree" keyword has remained a relevant topic since its emergence, its influence has fluctuated due to its relative simplicity and limitations compared to more advanced techniques that offer better scalability and protection.

Furthermore, the alluvial diagram reveals that the keyword "Perturbation Techniques" first emerged in the 2011–2014 subperiod, showing shifts in module placement and PageRank score across the four subperiods. Initially, it was located in module 7 out of 30 with a PageRank score of 0.0102, indicating that it became an influential theme within its cluster upon its introduction. Its position in a relatively high-ranked module highlights the immediate relevance of perturbation techniques, as researchers focused on anonymization approaches that protect user identities by subtly modifying graph data. In the 2015–2018 subperiod, "Perturbation Techniques" shifted to module 8 out of 39, with a decreased PageRank score of 0.00666. This slight decline suggests a minor reduction in its prominence, possibly due to the emergence of alternative methods that address limitations of perturbation, such as maintaining data utility. Despite this, its placement in a high-ranked module highlights its ongoing relevance as a practical anonymization approach. By the 2019–2022 subperiod, the keyword moved to module 1 out of 32, with a slightly lower PageRank score of 0.00623, showing a thematic shift to the highest-ranked module. This move reflects the technique's integration with other influential themes, including Graph Modification, potentially indicating broader applicability and increased acceptance as a foundational approach for safeguarding privacy in complex networks. The "Perturbation Techniques" keyword has thus consistently held a place of importance, adapting to meet evolving needs and anchoring itself within more central clusters as it gained prominence across the field.

Besides, it can be understood from the alluvial diagram that the "Uncertain Graph" keyword emerged in the 2011–2014 subperiod and has since experienced shifts in both module

placement and PageRank score. Initially appearing in module 1 out of 30 with a low PageRank score of 0.00170, "Uncertain Graph" entered the field as a relevant approach but not yet a prominent theme. This early presence suggests that researchers were beginning to explore the potential of uncertain graph structures to address privacy concerns in social networks, although its influence remained limited. In the 2015–2018 subperiod, the keyword moved to module 21 out of 39 modules, with an increased PageRank score of 0.00361, indicating a rise in importance. This shift reflects a growing interest in using uncertain graphs for privacy-preserving purposes, particularly as the complexity of social network data increased, requiring approaches that could manage ambiguities in relationships while protecting sensitive information. However, its placement in a lower-ranked module suggests that, despite gaining traction, it had not become a dominant theme. By the 2019–2022 subperiod, "Uncertain Graph" transitioned to module 28 out of 32 modules, with a marginally lower PageRank score of 0.00325. This decline in module placement and a slight decrease in PageRank indicate a reduction in prominence, likely due to the rise of more versatile methods like differential privacy, which offer stronger theoretical guarantees for similar privacy challenges. Nonetheless, "Uncertain Graph" has maintained consistent relevance, highlighting its niche application in scenarios where traditional deterministic models are insufficient, and underscoring its value in handling uncertainties in anonymized social networks.

Also, as illustrated in the alluvial diagram, the "Machine Learning" keyword first appeared in the 2011–2014 subperiod, positioned in module 30 out of 30 with a low PageRank score of 0.00143. This initial placement at the lowest-ranked module and with a modest PageRank score suggests that, although relevant, machine learning had not yet become a central theme within social network anonymization during this period. Its limited influence likely reflects early-stage research exploring machine learning applications in the domain without established methodologies or widespread adoption. By the 2015–2018 subperiod, "Machine Learning" transitioned to module 13 out of 39, with its PageRank score increasing to 0.00568. This movement to a mid-ranked module with a higher PageRank score indicates that machine learning had gained greater significance, driven by its potential to enhance traditional anonymization techniques. Researchers were likely leveraging machine learning's pattern recognition capabilities to improve privacy-preserving methods, particularly in addressing complex data structures and identifying sensitive information more effectively. In the 2019–2022 subperiod, "Machine Learning" advanced to module 5 out of 32 with a further increased PageRank score of 0.00797, placing it in a more prominent and central cluster of keywords. This rise in module rank and PageRank reflects machine learning's growing integration into social network anonymization, as it became a critical component for adapting privacy techniques to handle large-scale, high-dimensional social network data. The keyword's trajectory highlights a clear shift from exploratory research to a recognized foundational tool, underlining machine learning's transformative impact on the field by enabling adaptive, data-driven privacy solutions.

Regarding the "Genetic Algorithm (GA)" keyword, the alluvial diagram indicates that it emerged in the social network anonymization field during the 2011–2014 subperiod, appearing in module 1 out of 30 with a PageRank score of 0.00382. This initial placement reflects GA's early relevance, particularly as a meta-heuristic optimization algorithm applied to Clustering approaches to achieve k-anonymity by identifying the most optimal clusters for balancing privacy and data utility. GA's adaptability made it an effective tool for navigating the trade-offs inherent in anonymization, which contributed to its prominence during this period. By the 2015–2018 subperiod, GA transitioned to module 6 out of 39, with a slightly increased PageRank score of 0.00485. This rise suggests GA's continuing importance in optimizing k-anonymity solutions, although its position in a mid-ranked module

points to increasing research diversification and the emergence of other methodologies. In the 2019–2022 subperiod, GA moved to module 27 out of 32, with a decreased PageRank score of 0.00307, reflecting a decline in prominence. This shift likely corresponds with the rise of newer meta-heuristic methods, such as particle swarm optimization (PSO) and firefly algorithm (FA), which offered comparable optimization capabilities while potentially being more efficient or better suited to complex network structures. GA's trajectory from initial relevance to diminished influence illustrates the field's evolution as researchers increasingly turned toward algorithms that address dynamic and large-scale privacy challenges.

Furthermore, the alluvial diagram highlights that the "Artificial Intelligence (AI)" keyword first emerged in the social network anonymization field during the 2011–2014 subperiod. Initially positioned in module 30 out of 30, with a PageRank score of 0.00249, AI's presence was minimal, indicating its early, exploratory stage within this field. This limited influence reflects the broader research landscape of the time, where AI applications in social network anonymization were still largely unexplored. In the 2015–2018 subperiod, AI experienced significant growth, advancing to module 1 out of 39 with an increased PageRank score of 0.00691. This sharp rise in both ranking and score suggests that AI had quickly transitioned into a foundational concept, likely driven by the increasing adoption of machine learning for privacy tasks such as pattern recognition and data anonymization. As AI techniques became more robust, their ability to adaptively anonymize social networks while preserving data utility made them a valuable asset, transforming AI into a central theme in the field. In the 2019–2022 subperiod, AI shifted to module 7 out of 32 with a PageRank score of 0.00538, indicating a minor decline in prominence. This slight decrease may be attributed to the emergence of more specialized AI-driven methods and newer privacy-preserving technologies that addressed specific anonymization challenges. While AI's centrality in module placement declined, its continued relevance demonstrates its adaptability and enduring impact on the field. AI remains an indispensable component of social network anonymization research, underscoring the field's progression toward more sophisticated, AI-integrated solutions to meet evolving privacy needs.

Additionally, the alluvial diagram illustrates that "Optimization" gained prominence in the social network anonymization field beginning in the 2011–2014 subperiod, where it was positioned in module 6 out of 30 modules with a PageRank score of 0.00683. This suggests that optimization techniques became relevant during this time, especially the meta-heuristic optimization algorithms for refining anonymization methods. These optimization techniques improved clustering-based anonymization by finding the most suitable clusters to meet privacy standards, such as k-anonymity, while balancing data utility. In the 2015–2018 subperiod, "Optimization" remained in module 6 out of 39 modules with a slightly reduced PageRank score of 0.00602. Despite the minor drop, its placement in a similar-ranking module reflects the ongoing relevance of optimization in supporting scalable and effective anonymization. In the 2019–2022 subperiod, "Optimization" rose to module 3 out of 32 modules, with its PageRank score increasing to 0.00864, reflecting its peak importance in the field. This increase suggests that optimization became central in tackling the challenges of anonymizing large-scale networks. Researchers in this period likely explored novel and advanced meta-heuristic algorithms to improve the efficiency and adaptability of anonymization methods. Overall, the "Optimization" keyword has progressed from an emerging technique to a core theme, consistently used as meta-heuristic optimization algorithms, to refine privacy-utility balances in social network anonymization approaches.

3.5 Social network anonymization approaches

In this section, we present a new taxonomy for social network anonymization approaches. This classification is derived from our review of prior surveys discussed in the related work section, combined with the provided bibliometric analyses.

Reviewing the conducted surveys in the social network anonymization field was incredibly beneficial in proposing a new taxonomy. It helped us understand the existing taxonomies and their strengths and weaknesses. Besides, the previous surveys helped us understand the evolution of the field and the trends that have shaped it.

Furthermore, the analyses of the authors' keywords have revealed significant insights into the current trends and key focus areas in the field of social network anonymization. These analyses allowed us to identify the patterns, relationships, and clusters among keywords in the context of the existing literature and research in social network anonymization. Additionally, they provided an understanding of the evolving landscape of various themes, identifying if they are emerging or disappearing, specialized or peripheral, as well as whether they are motor themes or fundamental (basic) themes within this domain.

After interpreting the provided findings, the identified patterns and trends have indeed facilitated our understanding of the prevailing strategies and methodologies within the field. This has effectively allowed us to discern the main approaches utilized in social network anonymization. Hence, in the current study, a taxonomy for social network anonymization is presented. The proposed classification scheme incorporates the discovered trends, central areas of focus, and themes, thereby offering an updated and more comprehension of the field.

In order to accomplish this, we primarily focused on the keywords and themes specifically associated with certain approaches employed in the domain of social network anonymization. Based on the outcomes gleaned from our investigations, we identified the following key techniques in this domain: "Graph Modification Approach," "Generalization Approach," "Differential Privacy," "Uncertain Graphs," and "Cryptography." These keywords and themes serve as the cornerstones of this field. Additionally, our review of existing surveys revealed that numerous studies have combined these methods, leading to the development of hybrid strategies. Consequently, we have included an additional category labeled "Hybrid" to the classification scheme, which is illustrated in Fig. 28.

As depicted in Fig. 28, the primary approaches for social network anonymization comprise "Graph Modification," "Generalization/Clustering," "Differential Privacy," "Uncertain Graphs," "Cryptography," and "Hybrid." We will delve into each of these approaches, along with their respective subcategories, in the subsequent paragraphs.

3.5.1 Graph modification

This approach involves altering the structure of the social network graph to obscure the identities, relationships, attributes, or all for individuals while preserving the overall network structure. This can be done through various techniques like edge modification (addition, deletion or altering links), node modification (addition, deletion or altering nodes), or a combination of both [19, 21]. Graph Modification remains a central approach to social network anonymization because it allows for controlled distortion of the data that balances privacy with usability. Graph modification can handle various privacy challenges in both small and large networks. For instance, edge and node modifications directly tackle the risks posed by attackers seeking to exploit structural vulnerabilities in social graphs. These modifications make it more difficult for adversaries to accurately identify individuals or uncover relationships while still enabling meaningful analysis of the anonymized network. Additionally,

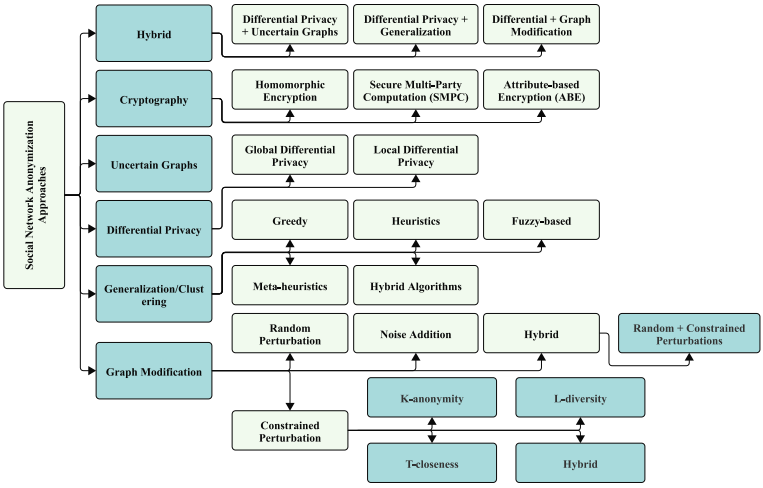


Fig. 28 Social network anonymization approaches

its flexibility allows it to be applied across different types of networks, such as directed, undirected, weighted, or unweighted graphs.

Based on the keyword and theme analyses conducted so far, the "Graph Modification" approach can be categorized into four subcategories, including "Random Perturbation," "Constrained Perturbation," "Noise Addition," and "Hybrid." It is worth mentioning that while techniques like random walk-based anonymization fall under this category, they have not been widely adopted due to scalability issues and their limited focus on specific privacy attacks. This method initially gained attention for its ability to introduce random movements through the graph, altering paths and making it difficult for attackers to follow specific connections between nodes. However, over time, the method's limitations became apparent. This anonymization method is computationally expensive, particularly when applied to large-scale, dynamic networks, which has made it less favorable compared to more adaptable techniques. Additionally, it is primarily effective against inference attacks, which are only one of many types of privacy risks in social networks. Consequently, random walk-based anonymization has not become a central or enduring method in the field, as it cannot effectively address the broader spectrum of privacy threats that other Graph Modification methods, such as Random Perturbation or Constrained Perturbation, can handle.

The "Random Perturbation" inherently comes from randomization and introduces random modification into the social network data and structure. This could be achieved by employing random modification using the mentioned edge modification, node modification, or both. There are several papers in the literature that developed social network anonymization methods based on the "Random Perturbation" method [18, 38, 53, 54, 60, 83, 85, 101, 108, 111, 113, 137, 145, 189, 197, 198, 208].

"Constrained Perturbation" is another variant of the "Graph Modification" approach in social network anonymization, where changes to the graph's structure are restricted by certain rules or constraints. During Constrained Perturbation, modifications such like adding, deleting, or changing edges or nodes are performed, but these changes must adhere to predefined constraints to maintain certain structural properties of the original network. After examining the keywords and themes, we identified that various well-established data anonymization

models, specifically "K-anonymity," "L-diversity," "T-closeness," and even hybrid models, are widely used in the field of social network anonymization [2, 7, 8],Bhattacharya & Mani, 2015, [12, 20, 22–25, 28–31, 41, 44, 45, 56, 57, 62, 66, 69, 72, 76–78, 82, 90–92, 94–96, 99, 100, 103, 104, 106, 109, 110, 112, 114, 118, 121, 125–127, 129–134, 138–141, 149, 150, 156, 160, 162, 163, 168–170],Wang, Shih, et al., 2013, [173–175],S.-L. [177, 178],Y. [178],W. [179–181, 183, 186, 190–194],Zhang, Li, et al., 2021, [199–202, 205, 209, 213]. These models are often employed alongside graph modification techniques. Given the inherent constraints these methods have in achieving anonymization, it is justifiable to classify these studies under the "Constrained Perturbation" subcategory.

"Noise Addition" is another class of the "Graph Modification" approach to anonymize social networks. The "Noise Addition" and "Random Perturbation" methods do have similarities, specifically in the context of graph modification for social network anonymization; however, they might differ in the extent of modifications made and the type of randomness introduced. The "Noise Addition" method focuses on introducing a sufficient level of random noise to protect privacy. There is only one study in the literature that specifically used this method to add "noise" or randomness to the data using addition, deletion, or modifying edges or nodes in the graph. In this paper, the authors proposed a privacy-preserving method called seamless privacy for the preservation of the users' privacy and their sensitive information by adding noise to the output graph [39].

Moreover, various research works in the field have devised methods for anonymizing social networks by integrating the previously mentioned techniques. For instance, some researchers merged the "Random Perturbation" and "Constrained Perturbation" strategies to create innovative methods for anonymizing social networks [49, 55, 97, 98, 122, 167, 188, 207].

3.5.2 Generalization/clustering

The generalization approach, also recognized as the clustering-based method, works by grouping nodes and edges into clusters, called super-node and super-edge, respectively. Each of these clusters is then anonymized, which publishes collective information regarding the structural properties of its constituent nodes. This method, as outlined by Hay et al., effectively conceals detailed information about individual entities within the network [58, 59].

One of the key reasons for the persistent use of this method in the social network anonymization field is its adaptability and scalability, particularly in large, complex social networks. Clustering methods strike a balance between anonymizing individual nodes and maintaining the overall structural integrity of the network, ensuring that the anonymized network still reflects meaningful patterns for analysis. This is particularly valuable in research and practical applications where maintaining utility while anonymizing data is critical. Moreover, clustering-based methods can be customized depending on the network's size, structure, and sensitivity of the data involved. By grouping nodes, this approach helps to mitigate risks associated with re-identification attacks while allowing researchers to analyze network properties at a higher level of abstraction. This adaptability has made clustering methods central in social network anonymization research, as they can be applied across a range of social networks, from small-scale community networks to large-scale global networks.

From the insights derived through our analyses, we classified the "Generalization/Clustering" studies [6, 9, 16], Gangarde, Pawar, et al., 2021, [46–48, 58, 63, 64, 71, 74, 75, 86, 102, 116, 117, 119, 125, 136, 144, 146–148, 158, 159, 161, 164, 172, 187, 204] within the social network anonymization literature into five categories. This categorization is

based on the type of algorithms these studies have developed to cluster the nodes. The five categories include: “Greedy Algorithms,” “Fuzzy-based Algorithms,” “Heuristic Algorithms,” “Meta-heuristic Algorithms,” and “Hybrid Algorithms.”

3.5.3 Differential privacy

As mentioned earlier, Differential Privacy, introduced by Dwork in 2006, is a rigorous mathematical framework for preserving privacy in data analysis and a key approach to social network anonymization. This approach concentrates on the process through which data are released rather than the raw data itself. This is accomplished by introducing a precise amount of randomness into the data publishing process. The central aim of differential privacy is to facilitate the sharing of valuable statistical data insights without compromising the privacy of the individuals within the dataset.

The findings of this study reveal that differential privacy has steadily emerged as a leading approach in social network anonymization since its introduction. A key factor behind its growing prominence is its adaptability to large-scale, dynamic networks, where user relationships are constantly shifting. While traditional anonymization techniques often struggle to keep pace with the evolving structure of social networks, differential privacy’s mathematical framework offers reliable privacy protection, regardless of network growth or complexity. Moreover, differential privacy’s ability to withstand linkage attacks, where adversaries exploit external data to re-identify users, sets it apart. By introducing carefully calibrated noise to the data, it ensures that individual identities and relationships remain secure, even in the presence of external datasets. This robust defense against re-identification is a crucial reason why differential privacy has become increasingly favored in the field of social network anonymization.

Furthermore, based on the analyses of the keywords and themes, we found that the studies related to the “Differential Privacy” approach can be categorized into “Global Differential Privacy” and “Local Differential Privacy” methods.

The “Global Differential Privacy” adds noise at the time of data queries on the entire dataset [26, 27, 50, 68, 89, 93, 105, 211, 212]. The curator of the social network has access to the original unaltered graph (nodes and edges). When information is requested from this network (like the total number of nodes, average number of connections per user, or other aggregate statistics), noise is added to the result to maintain privacy. This approach allows for more accurate results since noise is optimized across the whole dataset. However, because the original network is unaltered, there may be a higher potential risk to privacy, particularly if the curator is compromised.

On the other hand, “Local Differential Privacy” adds noise to each individual data point (such as individual nodes or edges) before it is added to the database [51, 52, 67, 73, 165, 206]. This means that the curator never sees the raw data of any individual node or edge. This approach provides more robust privacy protection as it ensures that the information about any specific individual is obfuscated. However, because noise is added to each data point independently, this method may lead to less accurate aggregate statistics or network analyses.

3.5.4 Uncertain graphs

Our analysis results reveal “Uncertain Graphs” as another approach within the realm of social network anonymization. This approach refers to the technique of introducing uncertainty

into the structure of the social network graph to protect user privacy. In a certain graph, the existence of each edge is known with certainty. However, in an uncertain graph, each edge is assigned a probability that indicates the likelihood of its existence. This uncertainty can help to anonymize the data by obscuring the exact relationships between individuals. In other words, the probabilities could be generated to obfuscate the true relationships and make it more difficult for an adversary to re-identify individuals or infer sensitive information.

We identified some studies conducted using the "Uncertain Graphs" approach in the social network anonymization field [11, 123, 124, 166, 182, 184, 185]. One of the core reasons for the increasing adoption of uncertain graphs is that their probabilistic nature significantly complicates inference and re-identification attacks. By introducing uncertainty, this approach provides an additional layer of privacy protection, which makes it difficult for attackers to accurately map out relationships in the network. Additionally, uncertain graphs are particularly well suited to social networks where relationships may not be strictly binary or static, often, the strength or likelihood of connections between users fluctuates. The uncertain graph model can capture this fluidity, making it more reflective of real-world social interactions. This flexibility allows the technique to mask the variability in relationships while still preserving the network's overall structure for analysis.

3.5.5 Cryptography

Another keyword or theme that is detected as an approach in the field of social network anonymization is "Cryptography." This approach plays an essential role in preserving privacy and security within social networks, including anonymization. Drawing from the insights gathered through our analyses, we highlight several cryptography-oriented techniques commonly employed for anonymizing social networks:

(1) Homomorphic Encryption is a cryptographic technique that enables computations on encrypted data, yielding results as if the operations were performed on the original, unencrypted data. This method has significant implications for anonymizing social networks (P. [89, 93, 214]. It can enhance user data privacy by allowing calculations on encrypted data, thereby safeguarding users' personal information. Homomorphic encryption also facilitates secure data sharing within the network, allowing sensitive information to remain encrypted even in-use, that is, accessible only to authorized entities. Furthermore, it enhances user privacy when interacting with third-party applications on the network by enabling these services to compute encrypted data, preventing them from accessing raw, unencrypted user information.

(2) Secure Multi-Party Computation (SMPC) is a cryptographic protocol used in social network anonymization [195, 196]. It enables multiple parties to collectively compute functions on their private inputs without revealing the inputs to each other, preserving individual data privacy. SMPC has various applications in social networks, including privacy-preserving data analysis, collaborative filtering for personalized recommendations, secure integration of anonymized data from multiple sources, and privacy-preserving analysis of social graph structures. These applications allow for data analysis, recommendations, data linkage, and network analysis while maintaining privacy and confidentiality. As a result, the SMPC protocols employ cryptographic techniques to ensure secure computations and simultaneously preserve individual privacy.

(3) Attribute-Based Encryption (ABE) is another cryptographic scheme that enables precise access control based on attributes, benefiting social network anonymization by enhancing privacy and access control for sensitive data [84]. ABE enables data encryption and decryption based on attributes instead of specific identities. Users and data are associated with attributes like roles or affiliations, and encryption policies are defined accordingly. Only users possessing the required attributes can decrypt and access the data. ABE encryption in social networks offers several applications. Selective Data Access allows users with specific attributes or roles to decrypt and access relevant information, ensuring that sensitive data are viewed only by authorized individuals. Privacy-Preserving Sharing enables the secure sharing of sensitive information within a social network by encrypting data based on attributes and restricting decryption and access to users with matching attributes. Anonymous Attribute-Based Access supports controlled data sharing while maintaining user anonymity, ensuring privacy, and granting access based on attributes.

3.5.6 Hybrid

Furthermore, our analyses revealed that numerous studies in the field of social network anonymization have employed a combination of the aforementioned approaches. For instance, some of the authors combined the "Uncertain Graphs" and "Differential Privacy" approaches to preserve the privacy of users in a social network [65]. Other authors used the "Differential Privacy" approach along with "Generalization/Clustering" to ensure user privacy in social networks [68]. In another study, the authors tried to use the "Differential Privacy" and "Graph Modification" approaches to propose an efficient social network anonymization model [143]. Besides, "Differential Privacy" and "Cryptography" were combined in another study to achieve a privacy-preserving social network publishing scheme [203].

4 Future research trends in social network anonymization

Based on the conducted analyses in the current study, we identified some of the most important research trends in the social network anonymization field. These trends are not only of academic interest but also have significant practical implications for researchers and practitioners. Researchers can use these trends to guide their future work, exploring new methods and enhancing existing ones. For practitioners, particularly those involved in data privacy and security, these trends offer opportunities to adopt advanced techniques that ensure compliance with regulatory standards, protect user privacy, and maintain the utility of social network data for analytical purposes.

(1) The advent of AI and machine learning models in developing new and innovative anonymization approaches has opened up a new frontier in data privacy for social networks. For instance, suppose there is a social network that encompasses a vast user base with diverse demographic attributes. In such a scenario, traditional anonymization methods, which are based on K-anonymity or L-diversity, might struggle to maintain effective privacy protections due to the diversity and volume of data. AI and machine learning models, however, can efficiently manage such diversity and scale by recognizing patterns and dynamically anonymizing user data to uphold user privacy. These AI-driven approaches can provide a more efficient and scalable solution compared to

traditional methods, especially when handling high-dimensional datasets with numerous attributes associated with each user. It is worth noting that deep generative models, such as GANs and Variational Autoencoders (VAE), can be beneficial in social network anonymization. These models can be used to create synthetic representations of original social network data. This synthesized data, which mimics the original data's structure and relationships without revealing individual identities, can be used for analysis or research without privacy concerns. Additionally, these deep learning models can be applied to alter the network structure subtly, making it difficult to identify individual nodes (users) while preserving the overall graph's structure. GANs, for instance, can generate synthetic nodes and edges that obscure the original identities but maintain the global network structure. Besides, AI and machine learning have a unique advantage. They can learn and improve over time. They can adapt to new privacy challenges, which is particularly important as social networks and privacy threats continue to evolve. For instance, if a new form of privacy violation emerges, these models can be quickly trained to recognize and counteract it. It is worth noting that practitioners can implement AI-driven methods for real-time data privacy protection, especially in industries such as healthcare, finance, or social media, where rapid and accurate anonymization is necessary to handle vast amounts of user data.

(2) In today's world of interconnected digital networks, there is an urgent need for the development of new and novel social network anonymization approaches, especially those based on innovative meta-heuristic algorithms. As the complexity and diversity of online interactions increase, traditional anonymization techniques might not be sufficient to maintain privacy and security. Meta-heuristic algorithms, such as Genetic Algorithm (GA), Particle Swarm Optimization Algorithm (PSO), and Firefly Algorithm (FA), due to their inherent flexibility and adaptability, can offer a potential solution. They can be designed to automatically adapt to the changing topologies of dynamic social networks, effectively anonymizing data in real-time and ensuring that users' identities and personal information remain protected, regardless of how the network evolves. Furthermore, the utilization of meta-heuristic algorithms in anonymization processes can significantly enhance the efficiency and accuracy of privacy-preserving techniques. These algorithms are designed to optimize solutions for complex problems and can help in determining the most effective ways to anonymize data without losing the underlying structure and utility of the data. This could directly benefit organizations managing large datasets by offering scalable, efficient solutions. Practitioners will also be able to apply these advanced techniques to enhance the effectiveness of anonymization methods while preserving data utility for business and research purposes.

(3) As dynamic social networks continually evolve, they become more complex and interconnected, heightening the need for advanced social network anonymization approaches. These approaches are crucial in preserving users' anonymity in an ever-changing environment and mitigating risks related to cybercrimes. Besides, they should preserve the users' data privacy and ensure the ethical use of AI and machine learning in analyzing dynamic social network data, preventing misuse. Hence, for researchers, this opens up a field of exploration for developing adaptive models that can anonymize networks in real-time, providing relevant, up-to-date privacy protections. Practitioners will also benefit from these advanced techniques as they can better safeguard user data in ever-changing environments and tackle real-time privacy risks.

(4) Developing new and novel social network anonymization approaches using hybrid methods, such as combining Differential Privacy with Graph Modification and integrating Uncertain Graphs with Differential Privacy, is becoming increasingly vital. Hybrid

approaches have the potential to offer superior privacy protection by harnessing the strengths of different techniques. For example, Differential Privacy provides a mathematical guarantee of privacy by adding a carefully calculated amount of noise to the data. When combined with Graph Modification, which alters the structure of the network to obscure individual identities, it can create a multi-layered approach that enhances overall privacy while still maintaining the utility of the data. Similarly, combining Uncertain Graphs and Differential Privacy could offer another powerful hybrid solution. Uncertain Graphs capture the inherent uncertainty in social relationships, thereby providing an additional layer of anonymization. When coupled with Differential Privacy's mathematical safeguards, this hybrid approach can result in a more comprehensive anonymization strategy, providing robust protection against both current and emerging de-anonymization techniques. For practitioners, these hybrid approaches can offer enhanced flexibility, allowing organizations to meet the specific privacy needs of different use cases, such as anonymizing datasets for research while maintaining user trust and compliance with data protection regulations.

(5) Developing social network anonymization approaches based on hyper-heuristics is vital due to their adaptability, efficiency, and flexibility. These high-level problem-solving frameworks can work dynamically within graph theory or clustering algorithms to offer solutions that not only preserve user privacy but also retain information loss at a minimum. Their ability to handle large-scale, high-dimensional data can outperform traditional methods, enhancing the quality of anonymization. Moreover, hyper-heuristics can be customized to various types of social networks and anonymization models, providing a universally applicable tool for data privacy. It is worth mentioning that these techniques could be used by practitioners working with large-scale, complex datasets to tailor privacy solutions to different types of social networks, maintaining privacy without sacrificing data utility.

(6) Developing strong privacy-enhancing techniques to provide resilience against AI-powered attacks and tools is an emerging and very hot research area. As AI becomes increasingly sophisticated in identifying patterns and re-identifying individuals within anonymized datasets, privacy mechanisms must evolve accordingly. Large Language Models (LLMs), in particular, could pose a potential threat as the attackers are able to use them to analyze anonymized social network data and extract sensitive information that was meant to be hidden. Future research could explore the development of social network anonymization techniques that can counteract AI-powered adversarial attacks, including those leveraging LLMs, ensuring that user data remains private even in the face of rapidly advancing technological threats.

(7) The development of privacy-preserving approaches that can incorporate user preferences is an important avenue for future research. In this context, users would have the ability to specify which aspects of their data are most sensitive and require stronger privacy protections. For instance, users could determine how their attributes (e.g., location, relationships, or activity history) should be handled, ensuring that more stringent privacy mechanisms are applied to these specific elements. Such a system would empower users by giving them more control over their personal information and aligning privacy preservation with individual concerns. This personalized anonymization approach would also enhance user trust and satisfaction, as it provides transparency in how data is anonymized and processed in social network environments. Additionally, integrating user preferences into anonymization methods could lead to more efficient data handling, where stronger privacy mechanisms are focused on the most critical areas while maintaining overall data utility. For instance, clustering algorithms or graph modification

techniques can be adapted to focus stronger protections on selected areas while maintaining data utility. These methods could be implemented by the practitioners to enhance transparency and trust with users. Offering privacy control options could improve user satisfaction and ensure compliance with emerging data protection laws.

(8) As decentralized social platforms, such as blockchain-based social networks, gain traction, traditional anonymization methods may fall short in addressing the unique challenges posed by these platforms. Future research could explore how existing anonymization techniques, such as Differential Privacy and Graph Modification, might be adapted or combined to decentralized environments. These networks are inherently different due to their distributed nature, lack of a central authority, and peer-to-peer architecture. Therefore, it is crucial to develop new anonymization techniques that can preserve privacy while maintaining the integrity of the decentralized data-sharing model. Additionally, approaches, such as SMPC and Homomorphic Encryption, could be further explored to ensure privacy in decentralized networks, particularly when multiple parties are involved in data processing without central control. Practitioners working on decentralized platforms will benefit from such techniques to maintain user privacy without compromising the integrity of decentralized, peer-to-peer networks.

(9) Recently, synthetic data-generation methods have posed a new threat to online social network users' privacy by creating data that closely resembles real user data. These generated datasets can be used to infer sensitive information or mimic user behavior in ways that compromise privacy. Therefore, developing privacy-preserving approaches that provide resilience against the risks posed by synthetic data-generation methods is becoming increasingly critical. Future research could explore how to detect and mitigate these threats by enhancing existing anonymization techniques, such as Graph Modification, Randomization, Differential Privacy, and Uncertain Graphs, or by devising new strategies specifically aimed at countering synthetic data risks.

5 Conclusion

In the era of big data, social network anonymization has become increasingly critical, as the vast amount of personal information shared across digital platforms demands robust privacy safeguards. This study contributes to this essential field by conducting a comprehensive bibliometric analysis of research published from 2007 to 2022, offering a systematic overview of its evolution. We collected and preprocessed data from Scopus using the Elsevier Developer Portal, carefully analyzing publications to identify core themes, frequent keywords, and prominent research trajectories. Using VOSviewer, we created network visualizations of keywords to reveal the field's most central terms, followed by statistical and network analyses integrating growth, T-score, and weighted node strength to quantify keyword importance. Afterward, through a co-word analysis conducted in SciMAT, we identified and tracked the development of prominent themes, allowing us to explore the thematic shifts over time within social network anonymization. Based on these findings, we proposed a novel taxonomy of anonymization approaches, mapping both established and emerging techniques.

Future research directions can focus on developing several key areas within social network anonymization. Researchers could expand AI-driven and machine learning-based approaches, such as deep generative models, to create scalable, adaptive anonymization solutions capable of handling diverse, high-dimensional datasets. Additionally, exploring novel meta-heuristic algorithms, beyond traditional ones like Genetic Algorithms (GA) and

Particle Swarm Optimization (PSO), could provide a dynamic tool to balance privacy and utility in complex, large-scale networks. Integrating techniques, such as Differential Privacy with Graph Modification, may also lead to significant advancements, combining the strengths of different methods for more comprehensive privacy protection. Another promising avenue involves hyper-heuristic approaches that adapt to context-sensitive privacy requirements, along with developing resilience-enhancing strategies against AI-powered adversarial attacks, particularly those using Large Language Models (LLMs). Furthermore, researchers could investigate personalized anonymization methods, allowing users to specify privacy preferences for certain attributes, and could innovate new strategies suited to decentralized networks, including blockchain-based platforms. Finally, in light of privacy threats posed by synthetic data-generation methods, creating anonymization techniques capable of counteracting these new risks would be valuable.

While our study primarily utilizes Scopus data, future work could expand upon this by incorporating additional sources, including conference papers and preprints, for a more comprehensive analysis. Ultimately, we hope that this work will serve as a catalyst for academics and practitioners to foster further advancements in social network anonymization and address the ever-growing complexity of online privacy.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10115-024-02289-y.

Acknowledgements We would like to acknowledge the support of the Digital Finance CRC, which is backed by the Cooperative Research Centers program, an Australian Government initiative.

Author contributions In developing and completing this paper, N.Y. was primarily responsible for conceptualization, methodology, performing literature research, data analysis using bibliometric software, and writing the main manuscript. H.Y. significantly contributed to data collection, statistical analysis, and revision of the manuscript, while H.G., M.S.K., and M.R. played a critical role in drawing the figures, restructuring the manuscript's flow and revising the content. B.A. provided key assistance in the revision process, helping to incorporate reviewer feedback, expand qualitative analyses, and enhance the clarity and coherence of the revised manuscript. A.H.G., as the corresponding author, supervised the project, ensured its coherence, and provided expert guidance throughout its various stages. All authors have read, reviewed, and approved the final manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. No funding was received to assist with preparing this manuscript.

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

Informed consent All of the authors consent to participate.

References

1. Abawajy JH, Ninggal MIH, Herawan T (2016) Privacy preserving social network data publication. IEEE Commun Surv Tutorials 18(3):1974–1997
2. Alavi, A., Gupta, R., & Qian, Z. (2019). When the attacker knows a lot: The gaga graph anonymizer. Information Security: 22nd International Conference, ISC 2019, New York City, NY, USA, September 16–18, 2019, Proceedings 22,
3. Ali AS, Zaaba ZF, Singh MM (2024) The rise of "security and privacy": bibliometric analysis of computer privacy research. Int J Inf Secur 23(2):863–885
4. Ali I, Balta M, Papadopoulos T (2023) Social media platforms and social enterprise: bibliometric analysis and systematic review. Int J Inf Manage 69:102510
5. Amiri B, Karimianghadim R, Yazdanjue N, Hossain L (2021) Research topics and trends of the hashtag recommendation domain. Scientometrics 126:2689–2735
6. Babu KS, Jena SK, Hota J, Moharana B (2013) Anonymizing social networks: a generalization approach. Comput Electr Eng 39(7):1947–1961
7. Baktha K, Tripathy B (2018) Alpha anonymization in social networks using the lossy-join approach. Trans Data Priv 11(1):1–22
8. Bensimessaoud S, Badache N, Benmeziane S, Djellalbia A (2016) An enhanced approach to preserving privacy in social network data publishing. In: 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST)
9. Bhagat S, Cormode G, Krishnamurthy B, Srivastava D (2009) Class-based graph anonymization for social network data. Proceed VLDB Endow 2(1):766–777
10. Bhattacharya, M., & Mani, P. (2015). Preserving privacy in social network graph with K-anonymize degree sequence generation. In: 2015 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)
11. Boldi, P., Bonchi, F., Gionis, A., & Tassa, T. (2012). Injecting uncertainty in graphs for identity obfuscation. *arXiv preprint* arXiv:1208.4145.
12. Bonchi F, Gionis A, Tassa T (2014) Identity obfuscation in graphs through the information theoretic lens. Inf Sci 275:232–256
13. Cahlik T (2000) Comparison of the maps of science. Scientometrics 49(3):373–387
14. Callon M, Courtial JP, Laville F (1991) Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemsitry. Scientometrics 22:155–205
15. Callon M, Rip A, Law J (1986) Mapping the dynamics of science and technology: sociology of science in the real world. Macmillan, Basingstoke
16. Campan, A., & Truta, T. M. (2008). Data and structural k-anonymity in social networks. International Workshop on Privacy, Security, and Trust in KDD,
17. CampanA, T. (2008). Aclusteringapproachfordataand structuralanonymityinsocialnetworks. *2ndACM SIGKDDInternationalWorkshoponPrivacy, Security, and TrustinKDD (PinKDD'08)*, r54.
18. Casas-Roma, J. (2014). Privacy-preserving on graphs using randomization and edge-relevance. Modeling Decisions for Artificial Intelligence. In: 11th International Conference, MDAI 2014, Tokyo, Japan, October 29–31, 2014. Proceedings 11,
19. Casas-Roma J (2019) An evaluation of vertex and edge modification techniques for privacy-preserving on graphs. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-019-01363-6
20. Casas-Roma J, Herrera-Joancomartí J, Torra V (2017) k-Degree anonymity and edge selection: improving data utility in large networks. Knowl Inf Syst 50:447–474
21. Casas-Roma J, Herrera-Joancomartí J, Torra V (2017) A survey of graph-modification techniques for privacy-preserving on networks. Artif Intell Rev 47:341–366
22. Casas-Roma J, Salas J, Malliaros FD, Vazirgiannis M (2019) k-degree anonymity on directed networks. Knowl Inf Syst 61:1743–1768
23. Chakraborty S, Tripathy B (2016) Alpha-anonymization techniques for privacy preservation in social networks. Soc Netw Anal Min 6:1–11
24. Chakraborty S, Tripathy B (2016) Privacy preserving anonymization of social networks using eigenvector centrality approach. Intell Data Anal 20(3):543–560
25. Chen K, Zhang H, Wang B, Yang X (2013) Protecting sensitive labels in weighted social networks. 2013 10th Web Information System and Application Conference,
26. Chen L, Zhu P (2015a). Preserving network privacy with a hierarchical structure approach. 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD),
27. Chen L, Zhu P (2015b). Preserving the privacy of social recommendation with a differentially private approach. 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity),

28. Cheng J, Fu A, Liu J. (2010). K-isomorphism: privacy preserving network publication against structural attacks. Proceedings of the 2010 ACM SIGMOD International Conference on Management of data,
29. Chester S, Gaertner J, Stege U, Venkatesh S (2012). Anonymizing subsets of social networks with degree constrained subgraphs. 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining,
30. Chester S, Kapron BM, Ramesh G, Srivastava G, Thomo A, Venkatesh S (2011) k-Anonymization of social networks by vertex addition. ADBIS 2(789):107–116
31. Clarkson, K. L., Liu, K., & Terzi, E. (2010). Toward identity anonymization in social networks. *Link Mining: Models, Algorithms, and Applications*, 359–385.
32. Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F (2011) An approach for detecting, quantifying, and visualizing the evolution of a research field: a practical application to the fuzzy sets theory field. J Informet 5(1):146–166
33. Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F (2012) SciMAT: a new science mapping analysis software tool. J Am Soc Inform Sci Technol 63(8):1609–1630
34. Cormode G, Srivastava D, Yu T, Zhang Q (2008) Anonymizing bipartite graph data using safe groupings. Proceed VLDB Endow 1(1):833–844
35. Cormode G, Srivastava D, Yu T, Zhang Q (2010) Anonymizing bipartite graph data using safe groupings. VLDB J 19(1):115–139
36. Coulter N, Monarch I, Konda S (1998) Software engineering as seen through its research literature: A study in co-word analysis. J Am Soc Inf Sci 49(13):1206–1223
37. Courtial J (1994) A coword analysis of scientometrics. Scientometrics 31(3):251–260
38. Dev H. (2014). Privacy preserving social graphs for high precision community detection. Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data,
39. Ding X, Wang W, Wan M, Gu M (2013). Seamless privacy: Privacy-preserving subgraph counting in interactive social network analysis. 2013 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery,
40. Ding Y, Chowdhury GG, Foo S (2001) Bibliometric cartography of information retrieval research by using co-word analysis. Inf Process Manag 37(6):817–842
41. Djomo R, Djotio Ndie T. (2021). Towards Preventing Neighborhood Attacks: Proposal of a New Anonymization's Approach for Social Networks Data. Big Data Technologies and Applications: 10th EAI International Conference, BDTA 2020, and 13th EAI International Conference on Wireless Internet, WiCON 2020, Virtual Event, December 11, 2020, Proceedings 10,
42. Dwork C, McSherry F, Nissim K, Smith A (2006). Calibrating noise to sensitivity in private data analysis. Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4–7, 2006. Proceedings 3,
43. Dwork C, McSherry F, Nissim K, Smith A (2016) Calibrating noise to sensitivity in private data analysis. J Priv Confid 7(3):17–51
44. Elabd E, Abdul-Kader H, Ead W (2019) Securely publishing social network data. Int Arab J Inf Technol 16(4):694–702
45. Fu Y, Wang W, Fu H, Yang W, Yin D (2018). Privacy preserving social network against dopv attacks. Web Information Systems Engineering–WISE 2018: 19th International Conference, Dubai, United Arab Emirates, November 12–15, 2018, Proceedings, Part I 19,
46. Gangarde R, Pawar A, Sharma A (2021). Comparisons of Different Clustering Algorithms for Privacy of Online Social Media Network. 2021 IEEE Pune Section International Conference (PuneCon),
47. Gangarde R, Sharma A, Pawar A (2022). Clustering Approach to Anonymize Online Social Network Data. 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS),
48. Gangarde R, Sharma A, Pawar A, Joshi R, Gonge S (2021) Privacy preservation in online social networks using multiple-graph-properties-based clustering to ensure k-anonymity, l-diversity, and t-closeness. Electronics 10(22):2877
49. Gao T, Li F (2019a). Privacy-preserving sketching for online social network data publication. 2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON),
50. Gao T, Li F (2019b). Sharing social networks using a novel differentially private graph model. 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC),
51. Gao T, Li F, Chen Y, Zou X (2017). Preserving local differential privacy in online social networks. Wireless Algorithms, Systems, and Applications: 12th International Conference, WASA 2017, Guilin, China, June 19–21, 2017, Proceedings 12,
52. Gao T, Li F, Chen Y, Zou X (2018) Local differential privately anonymizing online social networks under hrg-based model. IEEE Trans Comput Soc Syst 5(4):1009–1020

53. Gong W, Jin R, Li Y, Yang L, Mei J (2021) Privacy protection method for sensitive weighted edges in social networks. KSII Trans Int Inf Syst 15(2):540–557
54. Guo Y, Liu Z, Zeng Y, Wang R, Ma J (2018). Preserving privacy for hubs and links in social networks. 2018 International Conference on Networking and Network Applications (NaNA),
55. Hajian S, Tassa T, Bonchi F (2016) Individual privacy in social influence networks. Soc Netw Anal Min 6:1–14
56. Hamideh Erfani, S., & Mortazavi, R. (2019). A Novel Graph-modification Technique for User Privacy-preserving on Social Networks. *Journal of Telecommunications and Information Technology*.
57. Hamzehzadeh S, Mazinani SM (2019) ANNM: A new method for adding noise nodes which are used recently in anonymization methods in social networks. Wireless Pers Commun 107(4):1995–2017
58. Hay M, Miklau G, Jensen D, Towsley D, Li C (2010) Resisting structural re-identification in anonymized social networks. VLDB J 19:797–823
59. Hay M, Miklau G, Jensen D, Towsley D, Weis P (2008) Resisting structural re-identification in anonymized social networks. Proceed VLDB Endow 1(1):102–114
60. Hay, M., Miklau, G., Jensen, D., Weis, P., & Srivastava, S. (2007). Anonymizing social networks. *Computer science department faculty publication series*, 180.
61. He Q (1999) Knowledge discovery through co-word analysis. Libr. Trends 48:133–159
62. He, X., Vaidya, J., Shafiq, B., Adam, N., & Atluri, V. (2009). Preserving privacy in social networks: A structure-aware approach. 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology,
63. He X, Vaidya J, Shafiq B, Adam N, Atluri V (2012) Structure-aware graph anonymization. Web Intell Agent Syst Int J 10(2):193–208
64. Hoang AT, Carminati B, Ferrari E (2019). Cluster-based anonymization of directed graphs. 2019 IEEE 5th International Conference on Collaboration and Internet Computing (CIC),
65. Hu J, Shi W, Liu H, Yan J, Tian Y, Wu Z (2017). Preserving Friendly-correlations in uncertain graphs using differential privacy. 2017 International Conference on Networking and Network Applications (NaNA),
66. Hu Q (2015). (kl)-Anonymity for Social Networks based on k-Neighborhood Anonymity. Fourth International Conference on Information Science and Cloud Computing (ISCC2015),
67. Huang H, Yang Y, Li Y (2021). PSG: Local privacy preserving synthetic social graph generation. International Conference on Collaborative Computing: Networking, Applications and Worksharing,
68. Huang H, Zhang D, Xiao F, Wang K, Gu J, Wang R (2020) Privacy-preserving approach PBCN in social network with differential privacy. IEEE Trans Netw Serv Manage 17(2):931–945
69. Huang, K., Hu, H., Zhou, S., Guan, J., Ye, Q., & Zhou, X. (2022). Privacy and efficiency guaranteed social subgraph matching. *The VLDB Journal*, 1–22.
70. Ji S, Mittal P, Beyah R (2016) Graph data anonymization, de-anonymization attacks, and de-anonymizability quantification: a survey. IEEE Commun Surv Tutor 19(2):1305–1326
71. Jiang H (2015). A novel clustering-based anonymization approach for graph to achieve privacy preservation in social network. Proc. Int. Conf. on Adv. in Mechan. Engin. and Indust. Inform. AMEII 2015,
72. Jiao J, Liu P, Li X (2014). A personalized privacy preserving method for publishing social network data. Theory and Applications of Models of Computation: 11th Annual Conference, TAMC 2014, Chennai, India, April 11–13, 2014. Proceedings 11,
73. Ju X, Zhang X, Cheung WK (2019). Generating synthetic graphs for large sensitive and correlated social networks. 2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW),
74. Kaur I, Bhardwaj V (2019) K-anonymity enhancement for privacy preservation with hybridization of cuckoo search and neural network using clustering. Int J Innov Technol Explor Eng 8(10):1189–1196. https://doi.org/10.35940/ijitee.J8792.0881019
75. Kaveri VV, Maheswari V (2015) Cluster based anonymization for privacy preservation in social network data community. J Theor Appl Inf Technol 73(2):269–274
76. Kavianpour S, Ismail Z, Mohtasebi A (2011). Effectiveness of using integrated algorithm in preserving privacy of social network sites users. Digital Information and Communication Technology and Its Applications: International Conference, DICTAP 2011, Dijon, France, June 21–23, 2011, Proceedings, Part II,
77. Kiabod M, Dehkordi MN, Barekatain B (2019) TSRAM: A time-saving k-degree anonymization method in social network. Expert Syst Appl 125:378–396
78. Kiabod M, Dehkordi MN, Barekatain B (2021) A fast graph modification method for social network anonymization. Expert Syst Appl 180:115148
79. Kiranmayi M, Maheswari N (2021) A review on privacy preservation of social networks using graphs. J Appl Sec Res 16(2):190–223

80. Kreso I, Kapo A, Turulja L (2021) Data mining privacy preserving: research agenda. Wiley Interdiscip Rev Data Min Knowl Discov 11(1):e1392
81. Külcü Ö, Henkoğlu T (2014) Privacy in social networks: an analysis of facebook. Int J Inf Manage 34(6):761–769
82. Kumar S, Kumar P (2017) Upper approximation based privacy preserving in online social networks. Expert Syst Appl 88:276–289
83. Kumar S, Kumar P (2021) Privacy preserving in online social networks using fuzzy rewiring. IEEE Trans Eng Manag 70:2071–2079
84. Kumaran U (2021) A secure and privacy-preserving approach to protect user data across cloud based online social networks. Research anthology on artificial intelligence applications in security. IGI Global, Hershey, pp 560–585
85. Lan L (2015). Preserving weighted social networks privacy using vectors similarity. In: 2015 8th International Conference on Biomedical Engineering and Informatics (BMEI)
86. Langari RK, Sardar S, Mousavi SAA, Radfar R (2020) Combined fuzzy clustering and firefly algorithm for privacy preserving in social networks. Expert Syst Appl 141:112968
87. Law J, Bauin S, Courtial J, Whittaker J (1988) Policy and the mapping of scientific change: A co-word analysis of research into environmental acidification. Scientometrics 14(3–4):251–264
88. Lee B, Jeong Y-I (2008) Mapping Korea's national R&D domain of robot technology by using the co-word analysis. Scientometrics 77(1):3–19
89. Li A, Fang J, Jiang Q, Zhou B, Jia Y (2020). A graph data privacy-preserving method based on generative adversarial networks. Web Information Systems Engineering–WISE 2020: 21st International Conference, Amsterdam, The Netherlands, October 20–24, 2020, Proceedings, Part II 21
90. Li C, Aggarwal CC, Wang J (2011). On anonymization of multi-graphs. In: Proceedings of the 2011 SIAM International Conference on Data Mining
91. Li C, Amagasa T, Kitagawa H, Srivastava G (2014). Label-bag based graph anonymization via edge addition. In: Proceedings of the 2014 International C* Conference on Computer Science & Software Engineering
92. Li M, Liu Z, Dong K (2016) Privacy preservation in social network against public neighborhood attacks. 2016 IEEE Trustcom/BigDataSE/ISPA
93. Li P, Zhou F, Xu Z, Li Y, Xu J (2020). Privacy-preserving graph operations for social network analysis. Security and Privacy in Social Networks and Big Data: 6th International Symposium, SocialSec 2020, Tianjin, China, September 26–27, 2020, Proceedings 6
94. Li Y, Shen H (2010). On identity disclosure in weighted graphs. In: 2010 International Conference on Parallel and Distributed Computing, Applications and Technologies
95. Liu K, Terzi E (2008). Towards identity anonymization on graphs. Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data
96. Liu L, Liu J, Zhang J, Wang J (2010). Privacy preservation of affinities in social networks. Proceedings of the International Conference on Information Systems
97. Liu P, Bai Y, Wang L, Li X (2017) Partial k-anonymity for privacy-preserving social network data publishing. Int J Softw Eng Knowl Eng 27(01):71–90
98. Liu P, Cui L, Li X (2014). A hybrid algorithm for privacy preserving social network publication. Advanced Data Mining and Applications: 10th International Conference, ADMA 2014, Guilin, China, December 19–21, 2014. Proceedings 10
99. Liu P, Li X (2013). An improved privacy preserving algorithm for publishing social network data. 2013 IEEE 10th International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing,
100. Liu X, Li J, Zhou D, An Y, Xia X (2016). Preserving the d-Reachability When Anonymizing Social Networks. Web-Age Information Management: 17th International Conference, WAIM 2016, Nanchang, China, June 3–5, 2016, Proceedings, Part II
101. Liu X, Li M, Xia X, Li J, Zong C, Zhu R (2018). Spatio-temporal features based sensitive relationship protection in social networks. Web Information Systems and Applications: 15th International Conference, WISA 2018, Taiyuan, China, September 14–15, 2018, Proceedings
102. Liu X, Yang X (2011). A generalization based approach for anonymizing weighted social network graphs. Web-Age Information Management: 12th International Conference, WAIM 2011, Wuhan, China, September 14–16, 2011. Proceedings 12
103. Maag ML, Denoyer L, Gallinari P (2014). Graph anonymization using machine learning. 2014 IEEE 28th International Conference on Advanced Information Networking and Applications
104. Macwan K, Patel S (2019). Privacy preserving approach in dynamic social network data publishing. Information Security Practice and Experience: 15th International Conference, ISPEC 2019, Kuala Lumpur, Malaysia, November 26–28, 2019, Proceedings 15

105. Macwan K, Patel S (2021) Privacy preservation approaches for social network data publishing. Artificial intelligence for cyber security: methods, issues and possible horizons or opportunities. Springer, Cham, pp 213–233
106. Macwan KR, Patel SJ (2018) k-NMF anonymization in social network data publishing. Comput J 61(4):601–613
107. Majeed A, Lee S (2020) Anonymization techniques for privacy preserving data publishing: a comprehensive survey. IEEE access 9:8512–8545
108. Masoumzadeh A, Joshi J (2010). Preserving structural properties in anonymization of social networks. In: 6th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2010),
109. Masoumzadeh A, Joshi J (2012) Preserving structural properties in edge-perturbing anonymization techniques for social networks. IEEE Trans Depend Secure Comput 9(6):877–889
110. Mauw S, Trujillo-Rasua R, Xuan B (2016). Counteracting active attacks in social network graphs. Data and Applications Security and Privacy XXX: 30th Annual IFIP WG 11.3 Conference, DBSec 2016, Trento, Italy, July 18–20, 2016. Proceedings 30
111. Medforth N, Wang K (2011). Privacy risk in graph stream publishing for social network data. In: 2011 IEEE 11th International Conference on Data Mining
112. Medková J (2020). High-degree noise addition method for the $ k $-degree anonymization algorithm. 2020 Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems (SCIS-ISIS)
113. Milani Fard A, Wang K (2015) Neighborhood randomization for link privacy in social network analysis. World Wide Web 18:9–32
114. Minello G, Rossi L, Torsello A (2020) k-anonymity on graphs using the Szemerédi regularity lemma. IEEE Trans Netw Sci Eng 8(2):1283–1292
115. Mittal, P., Papamanthou, C., & Song, D. (2012). Preserving link privacy in social network based systems. *arXiv preprint* arXiv:1208.6189.
116. Mohapatra D, Patra MR (2017) A level-cut heuristic-based clustering approach for social graph anonymization. Soc Netw Anal Min 7:1–13
117. Mohapatra D, Patra MR (2019) Anonymization of attributed social graph using anatomy based clustering. Multim Tools Appl 78:25455–25486
118. Mohapatra D, Patra MR (2019b). Graph Anonymization Using Hierarchical Clustering. Computational Intelligence in Data Mining: Proceedings of the International Conference on CIDM 2017
119. Mohapatra D, Patra MR (2020). Cluster-Based Anonymization of Assortative Networks. Computational Intelligence in Data Mining: Proceedings of the International Conference on ICCIDM 2018,
120. Mongeon P, Paul-Hus A (2016) The journal coverage of Web of Science and Scopus: a comparative analysis. Scientometrics 106:213–228
121. Mortazavi R, Erfani S (2020) GRAM: an efficient (k, l) graph anonymization method. Expert Syst Appl 153:113454
122. Nagaraj K, Sridhar A, Sharvani G (2017). Identification of Network Communities and Assessment of Privacy Using Hybrid Algorithm. 2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS),
123. Nguyen HH, Imine A, Rusinowitch M (2014). A maximum variance approach for graph anonymization. International Symposium on Foundations and Practice of Security
124. Nguyen HH, Imine A, Rusinowitch M (2015). Anonymizing social graphs via uncertainty semantics. Proceedings of the 10th ACM symposium on information, computer and communications security,
125. Ni W, Sun F, Weng G, Xu L (2013). A Hellinger Distance Based Anonymization Method for Weighted Social Networks. 2013 10th Web Information System and Application Conference,
126. Ninggal MIH, Abawajy JH (2015) Utility-aware social network graph anonymization. J Netw Comput Appl 56:137–148
127. Nobari S, Karras P, Pang HH, Bressan S (2014). L-opacity: Linkage-aware graph anonymization.
128. Pratomo AB, Mokodenseho S, Aziz AM (2023) Data encryption and anonymization techniques for enhanced information system security and privacy. West Sci Inf Syst Technol 1(01):1–9
129. Puttaswamy KP, Sala A, Zhao BY (2009). Starclique: Guaranteeing user privacy in social networks against intersection attacks. Proceedings of the 5th international conference on Emerging networking experiments and technologies
130. Qing-jiang K, Xiao-hao W, Jun Z (2011). The (P, α, K) anonymity model for privacy protection of personal information in the social networks. 2011 6th IEEE Joint International Information Technology and Artificial Intelligence Conference
131. Rajabzadeh S, Shahsafi P, Khoramnejadi M (2020) A graph modification approach for k-anonymity in social networks using the genetic algorithm. Soc Netw Anal Min 10:1–17

132. Rajaei M, Haghjoo MS, Miyaneh EK (2015) Ambiguity in social network data for presence, sensitive-attribute, degree and relationship privacy protection. PLoS ONE 10(6):e0130693
133. Rajaei M, Haghjoo MS, Miyaneh EK (2015) An anonymization algorithm for (α, β, γ, δ)-social network privacy considering data utility. J Univers Comput Sci 21(2):268–305
134. Ren X, Jiang D, Jain DK (2022) A personalized α, β, l, k-anonymity model of social network for protecting privacy. Wirel Commun Mob Comput 2022:11. https://doi.org/10.1155/2022/7187528
135. Ronda-Pupo GA, Guerras-Martin LÁ (2012) Dynamics of the evolution of the strategy concept 1962–2008: a co-word analysis. Strateg Manag J 33(2):162–188
136. Ros-Martin M, Salas J, Casas-Roma J (2019) Scalable non-deterministic clustering-based k-anonymization for rich networks. Int J Inf Secur 18:219–238
137. Rousseau F, Casas-Roma J, Vazirgiannis M (2018) Community-preserving anonymization of graphs. Knowl Inf Syst 54:315–343
138. Safia B, Yacine C (2018). Privacy Preservation in Social Networks Sequential Publishing. 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA),
139. Saharkhiz A, Shahriari HR (2011). A Method for Preserving Privacy in Published Multi-relational Social Networks. KMIS,
140. Sarah AK, Tian Y, Al-Rodhaan M (2018). A novel (k, x)-isomorphism method for protecting privacy in weighted social network. 2018 21st Saudi Computer Society National Computer Conference (NCC),
141. Sargolzaei E, Khazali MJ, Keikha F (2016) Privacy preserving approach of published social networks data with vertex and edge modification algorithm. Indian J Sci Technol 9:12
142. Sathiya Devi S, Indhumathi R (2019). A study on privacy-preserving approaches in online social network for data publishing. Data Management, Analytics and Innovation: Proceedings of ICDMAI 2018, Volume 1,
143. Shakeel S, Anjum A, Asheralieva A, Alam M (2021) k-NDDP: An efficient anonymization model for social network data release. Electronics 10(19):2440
144. Shishodia MS, Jain S, Tripathy B (2013). GASNA: Greedy algorithm for social network anonymization. Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining,
145. Siddaramappa HK, Maradithaya S, Kumar S (2019). Secure Analysis of Social Media Data. 2019 International Conference on Automation, Computational and Technology Management (ICACTM),
146. Siddula M, Cai Z, Miao D (2018). Privacy preserving online social networks using enhanced equicardinal clustering. 2018 IEEE 37th International Performance Computing and Communications Conference (IPCCC),
147. Siddula M, Li Y, Cheng X, Tian Z, Cai Z (2019) Anonymization in online social networks based on enhanced equi-cardinal clustering. IEEE Trans Comput Soc Syst 6(4):809–820
148. Sihag, V. K. (2012). A clustering approach for structural k-anonymity in social networks using genetic algorithm. Proceedings of the CUBE international information technology conference,
149. Singh A, Bansal D, Sofat S (2018) What about privacy of my OSN data? Cybern Syst 49(1):44–63
150. Skarkala ME, Maragoudakis M, Gritzalis S, Mitrou L, Toivonen H, Moen P (2012). Privacy preservation by k-anonymization of weighted social networks. 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining,
151. Small H (1973) Co-citation in the scientific literature: a new measure of the relationship between two documents. J Am Soc Inf Sci 24(4):265–269
152. Small H, Griffith BC (1974) The structure of scientific literatures I: Identifying and graphing specialties. Sci Stud 4(1):17–40
153. Sood SK, Rawat KS (2021) A scientometric analysis of ICT-assisted disaster management. Nat Hazards 106(3):2863–2881
154. Sood SK, Rawat KS, Kumar D (2023) Scientometric analysis of ICT-assisted intelligent control systems response to COVID-19 pandemic. Neural Comput Appl 35(26):18829–18849
155. Sood SK, Rawat KS, Sharma G (2022) Role of enabling technologies in soft tissue engineering: a systematic literature review. IEEE Eng Manag Rev 50(4):155–169
156. Srivastava G, Citulsky E, Tilbury K, Abdelbar A, Amagasa T (2016). The effects of ant colony optimization on graph anonymization. *GSTF J Comput JoC*,*5*(1).
157. Stegmann J, Grohmann G (2003) Hypothesis generation guided by co-word clustering. Scientometrics 56(1):111–135
158. Stokes K (2011). On some clustering approaches for graphs. 2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011),
159. Stokes K (2013). Graph k-Anonymity through k-Means and as Modular Decomposition. Secure IT Systems: 18th Nordic Conference, NordSec 2013, Ilulissat, Greenland, October 18–21, 2013, Proceedings 18,

160. Stokes K, Torra V (2012) Reidentification and k-anonymity: a model for disclosure risk in graphs. Soft Comput 16:1657–1670
161. Su J, Cao Y, Chen Y, Liu Y, Song J (2021) Privacy protection of medical data in social network. BMC Med Inform Decis Mak 21:1–14
162. Sun C, Philip SY, Kong X, Fu Y (2013). Privacy preserving social network publication against mutual friend attacks. 2013 IEEE 13th International Conference on Data Mining Workshops
163. Tai C-H, Yu PS, Yang D-N, Chen M-S (2011). Privacy-preserving social network publication against friendship attacks. Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,
164. Thompson B, Yao D (2009). The union-split algorithm and cluster-based anonymization of social networks. Proceedings of the 4th International Symposium on Information, Computer, and Communications Security,
165. Tian H, Zheng X, Zhang X, Zeng DD (2021) ϵ-k anonymization and adversarial training of graph neural networks for privacy preservation in social networks. Electron Commer Res Appl 50:101105
166. Tian Y, Yan J, Hu J, Wu Z (2018). A privacy preserving model in uncertain graph mining. 2018 International Conference on Networking and Network Applications (NaNA),
167. Torra V, Salas J (2019). Graph perturbation as noise graph addition: A new perspective for graph anonymization. Data Privacy Management, Cryptocurrencies and Blockchain Technology: ESORICS 2019 International Workshops, DPM 2019 and CBT 2019, Luxembourg, September 26–27, 2019, Proceedings 14,
168. Tripathy B, Panda G (2010). A new approach to manage security against neighborhood attacks in social networks. 2010 International Conference on Advances in Social Networks Analysis and Mining,
169. Tsai Y-C, Wang S-L, Hong T-P, Kao H-Y (2015). Extending [K 1, K 2] Anonymization of Shortest Paths for Social Networks. Multidisciplinary Social Networks Research: Second International Conference, MISNC 2015, Matsuyama, Japan, September 1–3, 2015. Proceedings 2,
170. Tsai Y-C, Wang S-L, Kao H-Y, Hong T-P (2012). Confining edge types in k-anonymization of shortest paths. 2012 Third International Conference on Innovations in Bio-Inspired Computing and Applications,
171. Van Eck N, Waltman L (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 84(2):523–538
172. Wang R, Zhang M, Feng D, Fu Y (2015). A clustering approach for privacy-preserving in social networks. Information Security and Cryptology-ICISC 2014: 17th International Conference, Seoul, South Korea, December 3–5, 2014, Revised Selected Papers 17,
173. Wang S-L, Shih C-C, Ting I-H, Hong T-P (2013). Edge Selection for Degree Anonymization on K Shortest Paths. The 3rd International Workshop on Intelligent Data Analysis and Management,
174. Wang S-L, Tsai Y-C, Kao H-Y, Hong T-P (2010). Anonymizing set-valued social data. 2010 IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing,
175. Wang S-L, Tsai Y-C, Kao H-Y, Ting I-H, Hong T-P (2013) Shortest paths anonymization on weighted graphs. Int J Softw Eng Knowl Eng 23(01):65–79
176. Wang, S.-L., Tsai, Z.-Z., Hong, T.-P., Ting, I.-H., & Tsai, Y.-C. (2011). Anonymizing Multiple K-anonymous Shortest Paths For Social Graphs. 2011 Second International Conference on Innovations in Bio-inspired Computing and Applications,
177. Wang S-L, Tsai Z-Z, Ting I, Hong T-P (2014) K-anonymous path privacy on social graphs. J Intell Fuzzy Syst 26(3):1191–1199
178. Wang Y, Xie L, Zheng B, Lee KC (2011). Utility-oriented k-anonymization on social networks. Database Systems for Advanced Applications: 16th International Conference, DASFAA 2011, Hong Kong, China, April 22-25, 2011, Proceedings, Part I 16
179. Wu W, Xiao Y, Wang W, He Z, Wang Z (2010). K-symmetry model for identity anonymization in social networks. Proceedings of the 13th international conference on extending database technology,
180. Wu X, Ying X, Liu K, Chen L (2010). A survey of privacy-preservation of graphs and social networks. *Managing and mining graph data*, 421–453.
181. Xia H (2018). An Efficient Algorithm in Anonymous Social Network with Reachability Preservation. 2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC),
182. Xiao D, Eltabakh MY, Kong X (2018). Sharing uncertain graphs using syntactic private graph models. 2018 IEEE 34th International Conference on Data Engineering (ICDE),
183. Xie Y, Zheng M, Liu L (2016). A personalized sensitive label-preserving model and algorithm based on utility in social network data publishing. Human Centered Computing: Second International Conference, HCC 2016, Colombo, Sri Lanka, January 7–9, 2016, Revised Selected Papers 2,

184. Yan J, Zhang L, Shi W, Hu J, Wu Z (2017). Uncertain graph method based on triadic closure improving privacy preserving in social network. 2017 International Conference on Networking and Network Applications (NaNA),
185. Yan J, Zhang L, Tian Y, Wen G, Hu J (2018). An uncertain graph approach for preserving privacy in social networks based on important nodes. 2018 International Conference on Networking and Network Applications (NaNA),
186. Yang J, Wang B, Yang X, Zhang H, Xiang G (2014) A secure K-automorphism privacy preserving approach with high data utility in social networks. Sec Commun Netw 7(9):1399–1411
187. Yazdanjue N, Fathian M, Amiri B (2020) Evolutionary algorithms for k-anonymity in social networks based on clustering approach. Comput J 63(7):1039–1062
188. Ying X, Pan K, Wu X, Guo L (2009). Comparisons of randomization and k-degree anonymization schemes for privacy preserving social network publishing. Proceedings of the 3rd workshop on social network mining and analysis,
189. Ying X, Wu X (2008). Randomizing social networks: a spectrum preserving approach. proceedings of the 2008 SIAM International Conference on Data Mining,
190. Yu D, Zhao H, Wang L, Liu P, Li X (2019). A hierarchical k-anonymous technique of graphlet structural perception in social network publishing. Mobile, Secure, and Programmable Networking: 4th International Conference, MSPN 2018, Paris, France, June 18–20, 2018, Revised Selected Papers 4,
191. Yu L, Wang Y, Wu Z, Zhu J, Hu J, Chen Z (2014). Edges protection in multiple releases of social network data. Web-Age Information Management: 15th International Conference, WAIM 2014, Macau, China, June 16–18, 2014. Proceedings 15,
192. Yu L, Yang T, Wu Z, Zhu J, Hu J, Chen Z (2013). Sensitive edges protection in social networks. Web-Age Information Management: 14th International Conference, WAIM 2013, Beidaihe, China, June 14–16, 2013. Proceedings 14,
193. Yu L, Zhu J, Wu Z, Yang T, Hu J, Chen Z (2012). Privacy Protection in Social Networks Using l-Diversity. Information and Communications Security: 14th International Conference, ICICS 2012, Hong Kong, China, October 29–31, 2012. Proceedings 14,
194. Yuan M, Chen L, Philip SY, Yu T (2011) Protecting sensitive labels in social network data anonymization. IEEE Trans Knowl Data Eng 25(3):633–647
195. Yuan M, Chen L, Yu PS, Mei H (2013). Privacy preserving graph publication in a distributed environment. Web Technologies and Applications: 15th Asia-Pacific Web Conference, APWeb 2013, Sydney, Australia, April 4–6, 2013. Proceedings 15,
196. Yuan M, Chen L, Yu PS, Mei H (2015) Privacy preserving graph publication in a distributed environment. World Wide Web 18:1481–1517
197. Yue R, Li Y, Wang T, Jin Y (2018). An efficient adaptive graph anonymization framework for incremental data publication. 2018 5th International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC),
198. Zhang FQ, Dong L, Zhang SP, Li X, Chen GR (2014). Research of Privacy-preserving for Graph Mining of Government Affairs Office Automation Systems. Applied Mechanics and Materials,
199. Zhang H, Li X, Xu J, Xu L (2021). Graph Matching Based Privacy-Preserving Scheme in Social Networks. Security and Privacy in Social Networks and Big Data: 7th International Symposium, SocialSec 2021, Fuzhou, China, November 19–21, 2021, Proceedings 7,
200. Zhang H, Lin L, Xu L, Wang X (2021) Graph partition based privacy-preserving scheme in social networks. J Netw Comput Appl 195:103214
201. 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 Comput Sci 147:643–646
202. Zhang M, Qin S, Guo F (2017). Satisfying link perturbation and k-out anonymous in social network privacy protection. 2017 IEEE 17th International Conference on Communication Technology (ICCT),
203. Zhang M, Zhou J, Zhang G, Cui L, Gao T, Yu S (2022) APDP: attribute-based personalized differential privacy data publishing scheme for social networks. IEEE Trans Netw Sci Eng. 10(2):922–933
204. Zhang R, Qu B (2010). Preserving Privacy in Social Networks against Homogeneity Attack. 2010 International Conference on Internet Technology and Applications,
205. Zhang Y, Ma T, Cao J, Tang M (2016) K-anonymisation of social network by vertex and edge modification. Int J Embedded Syst 8(2–3):206–216
206. Zhang Y, Wei J, Zhang X, Hu X, Liu W (2018). A two-phase algorithm for generating synthetic graph under local differential privacy. Proceedings of the 8th International Conference on Communication and Network Security,
207. Zhao Y, Li Z (2019). Privacy management in social network data publishing with community structure. Proceedings of the 2nd International Conference on Healthcare Science and Engineering 2nd,

208. Zheleva E, Getoor L (2007). Preserving the privacy of sensitive relationships in graph data. International workshop on privacy, security, and trust in KDD,
209. Zhou B, Pei J (2011) The k-anonymity and l-diversity approaches for privacy preservation in social networks against neighborhood attacks. Knowl Inf Syst 28(1):47–77
210. Zhou B, Pei J, Luk W (2008) A brief survey on anonymization techniques for privacy preserving publishing of social network data. ACM SIGKDD Explor Newsl 10(2):12–22
211. Zhu H, Zuo X, Xie M (2019) DP-FT: A differential privacy graph generation with field theory for social network data release. IEEE access 7:164304–164319
212. Zhu, T., Li, G., Zhou, W., Yu, P. S., Zhu, T., Li, G., Zhou, W., & Yu, P. S. (2017). Differentially private social network data publishing. *Differential Privacy and Applications*, 91–105.
213. Zou L, Chen L, Özsu MT (2009) K-automorphism: a general framework for privacy preserving network publication. Proceed VLDB Endow 2(1):946–957
214. Zuo X, Li L, Peng H, Luo S, Yang Y (2020) Privacy-preserving subgraph matching scheme with authentication in social networks. IEEE Trans Cloud Comput 10(3):2038–2049

Navid Yazdanjue earned his Master of Science in Information Technology in 2018 from the Iran University of Science and Technology (IUST) in Tehran, Iran. He was a Research and Teaching Assistant with IUST from 2019 to 2022. At present, he is a Digital Finance Cooperative Research Centre (DFCRC) Industry Ph.D. candidate at the Data Science Institute of the University of Technology Sydney (UTS). He collaborates with Cyber Intelligence House (CIH) company as an industry partner within the aforementioned DFCRC Industry Ph.D. program. His research interests are primarily focused on metaheuristic optimization techniques, evolutionary computation, machine learning models, social network analysis (SNA), and cybersecurity.

Hossein Yazdanjouei holds B.Sc. and M.Sc. degrees in electronics from Urmia University, Urmia, Iran. He is currently engaged in research at the intersection of artificial intelligence and optimization. His work focuses on machine learning, deep learning, image processing, fuzzy systems, and metaheuristics. He has contributed to numerous academic publications, including studies on anonymity in social networks, swarm intelligence for image processing, and ensemble metaheuristic algorithms.

Hassan Gharoun is a PhD student in the field of analytics at the University of Technology Sydney, New South Wales, Australia. He received a master's degree in Industrial Engineering from the University of Tehran, Iran. His doctoral research focuses on uncertainty-aware machine learning models. He applies data-driven approaches to address industrial challenges, particularly in predictive maintenance and maintenance planning optimization. Hassan has also served as a reviewer for esteemed journals such as IEEE Transactions on Neural Networks and Learning Systems and IEEE Internet of Things Journal. His research interests encompass uncertainty quantification in machine learning, probabilistic machine learning, and meta-learning.

Mohammad Sadegh Khorshidi is a PhD candidate in Information Systems at the University of Technology Sydney (UTS). His doctoral research focuses on applying advanced data analytics, machine learning, and genetic programming. He received his M.Sc. and B.Sc. degrees in Civil Engineering from Shiraz University, Shiraz, Iran in 2017 and 2014, respectively. Mohammad has been recognized for his contributions to the field with the DECRA PhD Scholarship from the Australian Research Council. He has published in various reputable journals and serves as a reviewer for journals such as Environmental Research Letters, Journal of Hydrology, and Water Resources Management.

Morteza Rakhshaninejad received the M.S. degree in Information Technology Engineering, majoring in E-Commerce (machine learning in fraud detection systems) from the Iran University of Science and Technology, Iran, in 2019. His research interests include applied machine learning, information systems, computational biology, bioinformatics, and big data.

Babak Amiri is an assistant professor at Iran University of Science and Technology. He did his PhD in Information Technology at The University of Sydney in 2014. His research interests are broad and interdisciplinary, encompassing areas such as artificial intelligence, big data analytics, and social computing, with a particular focus on the application of complex network theories to various business and engineering problems. Dr. Amiri has published extensively in highly regarded scientific journals and has contributed to the advancement of knowledge in complex systems, machine learning, and data science.

Amir H. Gandomi (Senior Member, IEEE) is a Professor of Data Science at the Faculty of Engineering and Information Technology, University of Technology Sydney. He is also affiliated with Obuda University, Budapest, as a Distinguished Professor. Prior to joining UTS, Prof. Gandomi was an Assistant Professor at the Stevens Institute of Technology and a distinguished research fellow at BEACON Center, Michigan State University. Prof. Gandomi has published over three hundred journal papers and 12 books, which have collectively been cited 60,000+ times. He has been named as one of the most influential scientific minds and received the Highly Cited Researcher award (top 1% publications and 0.1% researchers) from Web of Science for six years. In a recent study at Stanford University, released by Elsevier, Prof. Amir H Gandomi is ranked 24th most impactful researcher in the AI and Image Processing subfield in 2023. He also ranked 18th in GP bibliography among more than 17,000 researchers. He has received multiple prestigious awards for his research excellence and impact, such as the 2024 IEEE TCSC Award for Excellence in Scalable Computing (MCR), the 2023 Achenbach Medal, and the 2022 Walter L. Huber Prize, the highest-level mid-career research award in all areas of civil engineering. He has served as associate editor, editor, and guest editor in several prestigious journals, such as AE of IEEE Networks and IEEE IoTJ. Prof. Gandomi is active in delivering keynotes and invited talks. His research interests are global optimization and (big) data analytics using machine learning and evolutionary computations in particular.

Authors and Affiliations

Navid Yazdanjue[1] · Hossein Yazdanjouei[2] · Hassan Gharoun[1] · Mohammad Sadegh Khorshidi[1] · Morteza Rakhshaninejad[3] · Babak Amiri[3] · Amir H. Gandomi[1,4]

✉ Amir H. Gandomi
Gandomi@uts.edu.au

Navid Yazdanjue
navid.yazdanjue@gmail.com

Hossein Yazdanjouei
h.yazdanjouei@gmail.com

Hassan Gharoun
hassan.gharoun@student.uts.edu.au

Mohammad Sadegh Khorshidi
ms.khorshidi@student.uts.edu.au

Morteza Rakhshaninejad
m_rakhshaninejad@ind.iust.ac.ir

Babak Amiri
babakamiri@iust.ac.ir

1 Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia

2 Microelectronics Research Laboratory, Urmia University, Urmia 5756151818, Iran

3 School of Industrial Engineering, Iran University of Science and Technology, Narmak, Tehran 16846-13114, Iran

4 University Research and Innovation Center (EKIK), Óbuda University, Budapest 1034, Hungary