A Literature Survey and Classifications on Data Deanonymisation

Dalal Al-Azizy[1,2], David Millard[1], Iraklis Symeonidis[3], Kieron O'Hara[1]
Nigel Shadbolt[1]

[1]Web & Internet Science, School of Electronics & Computer Science,
University of Southampton, UK
{daaa1g09,dem,kmo,nrs}@ecs.soton.ac.uk
[2]University of Tabuk, Saudi Arabia
[3]ESAT/COSIC, KU Leuven and iMinds
iraklis.symeonidis@esat.kuleuven.be

Abstract. The problem of disclosing private anonymous data has become increasingly serious particularly with the possibility of carrying out deanonymisation attacks on publishing data. The related work available in the literature is inadequate in terms of the number of techniques analysed, and is limited to certain contexts such as Online Social Networks. We survey a large number of state-of-the-art techniques of deanonymisation achieved in various methods and on different types of data. Our aim is to build a comprehensive understanding about the problem. For this survey, we propose a framework to guide a thorough analysis and classifications. We are interested in classifying deanonymisation approaches based on type and source of auxiliary information and on the structure of target datasets. Moreover, potential attacks, threats and some suggested assistive techniques are identified. This can inform the research in gaining an understanding of the deanonymisation problem and assist in the advancement of privacy protection.

Keywords: deanonymisation, re-identification, privacy

1 Introduction

Increasingly, the Web has become a world of interconnected information particularly in terms of social practices. People participate in different social platforms and post and publish their personal information and activities that are occasionally, and based on the context, set to be anonymous. They also consume other information from other platforms and applications such as government and commercial services through their social profiles. Governments and other institutional bodies are now releasing their data on the Web for transparency, accountability, economic, and service improvement purposes, among other reasons this extensive growth of the Web is continuously accompanied by increasing concerns about new threats to privacy, and this also applies to all other networking environments that depend on their private intranets or isolated databases. The key fact here about the threats to privacy is the enormous amounts of data that are increasingly available in different contextual forms.

Here, we study the problem of data re-identification resulting from the rich availability of data from heterogeneous resources. We conduct a survey of the state-of-the-art techniques of deanonymisation attacks. Our work extends similar work found in the literature; brief survey [1] on online social networks and another

systematic review [2] of re-identification attacks on health data. In our survey, we include - on a large scale - recent techniques from various fields to provide a broad overview of the deanonymisation problem.

1.1 Research Motivation

This research is motivated by a number of issues. First, publishing data in the Web including personal data is becoming more advanced and growing immensely with a corresponding growth in publishable data links. Second, using online social networks, which include personal and private data. Third, mashup applications are also an issue in terms of collecting information from different platforms for third party interests. Fourth, the increasing concern about data leakage [3] [4]. Fifth, recent research has questioned the ability of anonymisation techniques to successfully preserve privacy [5-7]. Sixth, publishing open government data and the concern of jigsaw re-identification problem regarding private data (transparent government may lead to transparent citizens in terms of their private personal information) [8]. Finally, Political and commercial interests for advancing open data and linked open data with low interests in discussing privacy issues and challenges about publishing data that may affect their business.

Therefore, in considering all these issues that affect one context such as online social networks, it can be strongly inferred that the deanonymisation problem is more serious when considering all other different contexts.

1.2 Survey Purpose

The purpose of this research is to explore the deanonymisation problem and provide a broader conceptual view and understanding of how it - and all other issues related to this problem - occurs. In particular, it is important to understand how an adversary can exploit any data to use as background knowledge to achieve the deanonymisation attacks and what the possible threats are that result from that. This survey aims to provide analysis and classification for deanonymisation to include the broader audience of non-technical practitioners from legal and governmental fields. This is to help those interested to make sense of and understand the deanonymisation problem and how it might affect their decisions in terms of publishing data and protecting privacy rights.

This survey includes large scale of deanonymisation techniques to provide a broad overview of deanonymisation. We conduct the following steps: First, we design a deanonymisation framework for analysing several techniques in different environments. Second, we classify deanonymisation approaches based on the appropriate methodology that the adversary follows to enable them to exploit auxiliary information with target data structure.

1.3 Limitations

This research does not study the information gained as background knowledge by the adversary from real world information. For instance, information gathered from work environments such as colleagues' personal information, or neighborhood data regarding houses and population, any data that might be used and combined with other data to give meaningful information. That being the case, it is worth studying

such scenarios. However, this is beyond of scope of our survey that focuses on published data.

The rest of this paper is organised as follows. In section two we review the related work. In section three, we explain our research methodology for this survey. In section four, we introduce our deanonymisation framework for analysis, followed by classification of approaches and other related issues with more details in section five. In section six, we discuss our insights and present arguments, and then the paper is concluded in section seven.

2 Related Work

In this section, we first briefly overview data anonymisation and data deanonymisation to provide an outlook on privacy solutions and their possible violations. Subsequently, we present the existing survey studies on deanonymisation that are similar to our work and then we state the features that distinguish this paper from existing work in the remarks section.

2.1 Data Anonymisation

This technique is the most common method used to preserve data privacy by removing personal identifiable information (PII). This ensures eliminating the risk of private data disclosure while these anonymous data are being processed or transferring between systems or networks [9]. A number of anonymisation models were developed for protecting privacy such as k-anonymity [10-12], l-diversity [13] , [14] and t-closeness [15]. However, k-anonymity model is argued to ensure privacy [16].

2.2 Data Deanonymisation

Narayanan et al [17] researched the problem of deanonymisation where disclosure of anonymous data succeeded. They could achieve structural and computational attacks exploiting auxiliary information from other data sources. A number of studies [18] and [19] have investigated this problem and proved their theoretical and practical results about threats of re-identification and linkability. Deanonymisation becomes serious and practical when the structure of the target dataset is known and where large amounts of auxiliary data can be exploited.

2.3 Existing Surveys on Deanonymisation

There are two studies in the literature that systematically reviewed a number of deanonymisation techniques in detail. The first study is a brief survey [1] of five deanonymisation studies in online social networks. The purpose of surveying this problem is to technically show how it is both possible and practical to deanonymise released data in online social networks. In this brief survey, Ding et al [1] unified the models of deanonymisation based on the feature matching between the data released in online social networks and the background knowledge acquired by the adversary. They also classified the five attacks of deanonymisation into two categories based on matching direction between released data D in online social networks and background

knowledge K: mapping-based methods (match K against D) and guessing-based methods (match D against K).

The second study of related work [2] is a systematic review of re-identification attacks on health data and this was undertaken for slightly different purposes. It focused on re-identification attacks' representation among health data and compared these attacks with other deanonymisation attacks on other types of data. Furthermore, the review computed the total proportion of all successfully re-identified records in these attacks and assessed whether this highlights shortcomings in existing de-identification techniques. El Emam et al [2] reviewed 14 articles of deanonymisation attacks for evaluation against health regulations and de-identification standards.

2.4 Remarks

Our research is more focused on the methodology by which deanonymisation can be achieved. In this paper, we proposed a framework to guide an in-depth analysis of the current deanonymisation techniques in the literature. Moreover, we classify the deanonymisation approaches and provide lists of attacks, threats and assistive techniques.

The major features that distinguish our work from existing surveys are as follows. We extends the brief survey [1] for a more systematic and thorough review in terms of number of deanonymisation techniques and various contexts to include, and this review also generates differently based classifications.

Our survey is more comprehensive in that it includes a large-scale number from the state-of-the-art investigation of the deanonymisation problem including their five attacks. Moreover, we cover not only online social networks, but we also include open data, databases, traces data and mobile sensors, and other heterogeneous networks and datasets. Pertaining to the classification method, our work is much broader and takes into account the deanonymisation methodology of exploiting the background knowledge of the structure of the target dataset with auxiliary information acquired by the adversary and then, based on that, following a certain approach to achieve the attack. We also provide details of types of attacks to commit deanonymisation and resultant threats. Regarding the second review study [2], only four studies are included in our survey as the remaining studies do not have the data needed for our research framework for analysis and or do not satisfy and technical level that we are looking for. For instance, Bender et al [20] used cluster analysis for deanonymising register data of statistical agencies using survey data related to scientific purposes. Some few studies are reports relate to specific healthcare and legal bodies.

There are also more studies in the literature that researched deanonymisation [21-27] but these are not included in our survey as they are for analysis and evaluation purposes.

3 Research Methodology

Here, we explain our methodology in conducting this survey pertaining to selecting research papers of deanonymisation techniques and the way we analyse them.

3.1 Searching for Papers

We followed certain criteria to search for papers and articles to include in our survey. We searched for academic literature from several online conference and journal databases: IEEE, ACM, Springer and Google Scholar. Key words used for searching are: deanonymisation, re-identification, anonymisation and privacy risks. As our survey is focused on technical mechanisms in which deanonymisation is achieved, we only consider papers that report novel methods for successful deanonymisation attacks and that provide practical and theoretical results.

3.2 Methods of Analysis and Classifications

Our survey aims to achieve a comprehensive understanding and conceptualising of deanonymisation. To achieve this objective, we provide a systematic review for a large scale of various types of deanonymisation techniques in heterogeneous environments. Hence, we propose a deanonymisation framework to analyse the state-of-the-art techniques to understand all their related issues. The components of the framework will extract specific details from each technique to guide the analysis in reasoning how the deanonymisation is achieved in a specific way. Then we classify the deanonymisation approaches based on the analysis. We also provide classification for all possible attacks that practically perform the deanonymisation and assistive techniques that can help in achieving more effective deanonymisation.

4 Deanonymisation Framework

To survey the studies in the literature, a framework of all identified aspects of deanonymisation is designed for analysis and classifications. This framework is used as a guideline to understand the problem of deanonymisation and how it works in different contexts. The framework consists of major components as shown in figure 1; each of them reflects an essential stage of the deanonymisation process.

In this section, we present our classifications on deanonymisation. The major part of our research in this survey is to understand how the adversary can use type and source of the auxiliary information with the structure of target dataset to form sufficient background knowledge, which in turn enables them to follow the most effective approach to achieve a successful deanonymisation attack. We also classify the type of attacks in the literature and the threats resulted from that. Moreover, we introduce all the assistive techniques that are used to boost the deanonymisation to be more effective. And we finally identify some of the adversary capabilities.

4.1 Target Dataset

This refers to the data source that the adversary is interested in attacking and is usually protected by anonymisation techniques. To achieve the deanonymisation attack, the adversary needs to pay attention to the following:

a) *Anonymisation techniques:* to defeat it properly.
b) *Anonymised data*: that they usually have interests to disclose it.
c) *Dataset structure:* to configure the proper methodology and approaches of deanonymisation and what auxiliary information is needed for that.

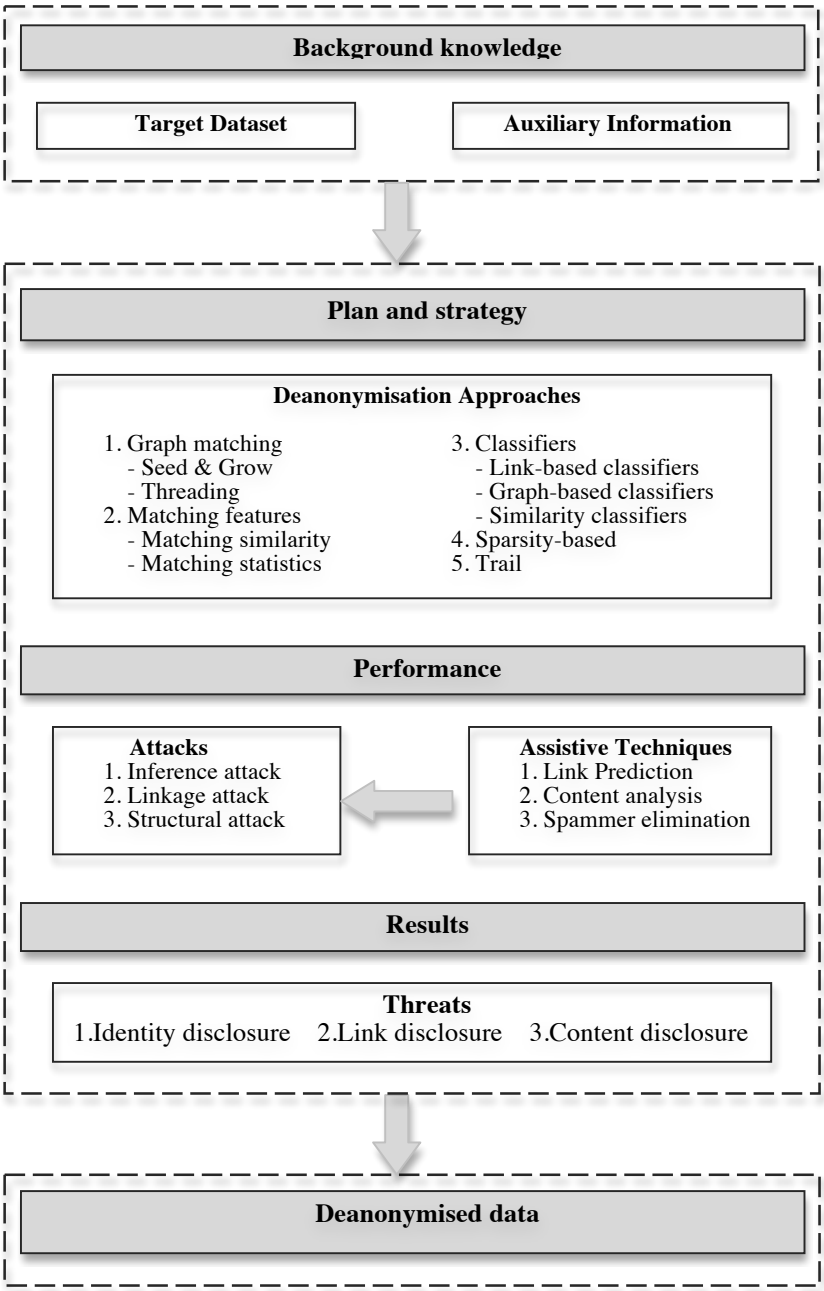

Fig. 1. Deanonymisation framework for analysis and classifications.

d) *Data sparsity:* which also has a relationship with data structure. Sparsity refers to the way in which data are populated through the Web and their relation with other data.

In some cases this target dataset can be used as a source of auxiliary information that assists the adversary to gain a foundation from which to achieve the deanonymisation.

4.2 Auxiliary Information

The adversary usually needs background information as a requirement to start from a strong base in order to achieve the deanonymisation attack. Sometimes auxiliary information is extracted from the same target datasets where the attacker can manipulate it computationally.

4.3 Deanonymisation Approaches

When the adversary finds that they have enough background knowledge to carry out the attack, then next stage is to choose a proper approach. We analyse a methodology that the adversary intend to find and follow based on the background knowledge that they could acquire from auxiliary information and the structure of target dataset. As a result, the adversary decides the most appropriate approach to commit the deanonymisation. We emphasise on this part of the framework as it shows how much publishing certain data helps the adversary to find a way to exploit these data. Furthermore, this specifies how available data can be combined with other data to form meaningful data that reveal many values that are not taken into account during publishing. A traditional approach of deanonymisation is matching data from different resources to disclose the anonymised parts in one of these resources. Some approaches can be combined for complex deanonymisation attacks. More details about these approaches are presented in the next section.

We classify deanonymisation approach as shown in figure 1 based on the way that the adversary uses to exploit the auxiliary information with the structure of target dataset to achieve the attack. The typical approach for deanonymisation is matching data in which auxiliary information is used for mapping with an anonymised target dataset. Therefore, any overlapping between information will act as a guide in revealing the hidden elements. The following approaches are taking unique form to perform the matching mission.

Now, we explain each approach separately providing definitions, examples and comparisons.

Graph matching: this approach constitutes the common form among online social networks as these networks consist of nodes and edges. As a large number of deanonymisation studies were undertaken on online social networks, graph matching represents the most common approach introduced here in this research. This approach, in general, focuses on two graphs from the same target domain or perhaps from two different domains. Both graphs are used for mapping tests between the shared anonymised nodes and measuring overlaps. Zhang et al [28] utilised this approach in heterogeneous networks.

Graph topology in [18] which built on [29], refers to the graph structure as a major information to achieve deanonymisation. The study argues that privacy can be breached once its structure is revealed, as a standard technique of re-identification can be developed based on that. Here, graph topology that makes use of structural properties of social graphs adopts the following steps:

a) Find the largest common sub-graph between a pair of ego nets;
b) Focus on common nodes of two graphs that are in distance of 1-hop from the centre; and
c) Utilise the degree distribution of nodes of social networks and observe the 1-hop neighbourhood degree distribution of the common nodes, storing each node's degrees in a list as a signature if matching signatures are found between a pair of nodes from the two graphs, those nodes are considered to be the same. It does not work with graphs where their common nodes are at distance of 2-hop from the centre (one node in a pair, or both nodes in a pair) as in [30] which covers n-hop and therefore can deanonymise more data.

Other studies [31] and [32] rely on aggregating networks from Twitter and Flickr to build up an auxiliary background and therefore can perform the graph matching.

Other studies [33] and [34] also use graph matching for deanonymisation.

Seed & grow: graph matching in some studies starts with a complex process called seeding to plant a node such as a user account in a social graph and then make it building up links with other nodes. This expansion process is called seed and grow in [35] and [36] . Narayanan et al [17] described their seeding method in two steps: seed and propagation. They combined that with link prediction to strengthen their approach and can manage the graph evolution. That could deliver high accuracy and coverage for their attack. They used graph matching by crawling Flickr with scrubbed user identities. The first step is seed identification where they deanonymise a small number of nodes. These nodes are used in the second step, which is called propagation, as anchors to breed the deanonymisation to a large number of nodes. This study used auxiliary information from the same target domain, crawling Flickr, but generating an evolved version. Also these studies [31], [32], [37], [38] used seeding based deanonymisation techniques. Another study [39] presents a new graph matching algorithm that relies on smaller seeds than other methods.

Threading: is another starting point for complex graph matching. Ding et al [40] utilised the threading method for correlating the sequential releases due to the rapid growth in dynamic social networks. Then they match these releases with eachother.

Matching similarity: is an approach where attacks rely on similar features between the target dataset and auxiliary information to perform the matching. Gambs et al [41] performed their attack on geolocated databases. They used mobility traces to find distance similarities and then match them with the help of statistical predictors. Likewise, Ji et al [42] match similarity between social data and mobility traces data. Another study [43] used text similarity between resume and tweets. Moreover, some techniques use node similarity to match graphs in social networks [44].

Matching statistics: in this approach attacks depend on using statistics to map datasets. In [45] the attack relies on the unique features of users' data to perform matching statistics. All these features in this approach led to users being identified.

Link-based classifiers: the second type of classifiers is link-based and group-based which use friendship and group membership to identify some private attributes as in [46].

Group-based classifiers: the second type of classifiers is group-based which use friendship and group membership to identify some private attributes as in [46]. Also the group-based classifier was combined with auxiliary information from the browsing history to identify users in social networks in [47].

Similarity classifiers: The Classifiers approach depends on local features such as temporal activity, text, geographic, and social features to form similarity classifiers as in [48]. These classifiers predict whether or not two accounts from two different social platforms are belonging to the same individual by deciding on similarities between them.

Sparsity-based: this technique is utilised to attack sparse data either high dimensional micro data or specified with certain relations such as location. Sparse data share comparatively few relationships. Lane et al [49] used activity stream of the target within anonymised sets of streams gathered from other users. They designed a two-stage deanonymisation framework: the first stage follows activity relationship mining, which consist of rules that link various aspects of two captured activity types and can help to decide if two activities were probably achieved by the same individual or not. The second stage uses the SCORE algorithm: the adversary input the auxiliary info and a stream of one or more activity types of a single anonymised user. This attack focused on the risks of sharing data of inference-based representation. Another study [50] considers a hypothesis that deanonymisation improves due to database sparsity. It also considers another hypothesis that states if the auxiliary information consists of values matching with rare attributes of a target database, then the deanonymisation achieved is greater. Both hypotheses are observed and motivated by the heuristic deanonymisation of the study [19] that targets high-dimensional micro-data. Merener [50] suggests that to achieve a high level of deanonymisation in a less sparse database, using large auxiliary information is needed. Frankowski et al [51] shows that re-identification is possible in sparse related spaces when matching data. They test data available in movie mentions with another data in movie rating. They found that data relate to a specific item in both datasets can lead to deanonymising users participating in such datasets.

Trail: The trail re-identification approach was developed for linking genomic data to their identified users whose records are publicly available in various databases. This method exploits the unique features of visited location and utilises visit pattern in distributed environment such as healthcare as in [52], [53]. Moreover, the concept of trail re-identification approach also includes deanonymising users by collecting network data and then mapping their anonymised traces with IP addresses [54]. Likewise, in [55] matching IP addresses with Tor hidden services using traffic analysis. Another technique of trail approach is tracking users by detecting their fingerprints in web browsers [56]. This method relies on visited web pages in the browser's history.
Table 1 summarises features that used to form each approach.

Table 1. Features utilised for exploitation in each approach.

Deanonymisation approach	Features to exploit	Reference
Graph matching	Network structure	[18], [28],[29], [30-34]
Seed & grow	Growing links into the target graph	[17], [31],[32], [35-39]
Threading	Correlation between sequential release of dynamic OSN	[40]
Similarity matching	Similarity of specific features for mapping	[41-44]
Statistical matching	Statistics of unique features	[45]
Link-based classifiers	Friendship membership for classification	[46]
Graph-based classifiers	Group membership for classification	[47]
Similarity classifiers	Similarity of local features for classification	[48]
Sparcity-based	How data are sparse, high dimensional, microdata	[49-51]
Trail	Unique features of visited location	[52-56]

4.4 Deanonymisation Attacks

Deanonymisation attacks take different forms based on the method that the adversary finds is most effective to exploit background information. Attacks can be inference, linkage, structural, and predictive, active or passive. One attack can be described as more than one of these forms. Some attacks are supported with assistive techniques for more effective deanonymisation.

Deanonymisation attacks listed below are described in terms in the way it is committed. An attack may convey more than one description listed here.

Inference attacks: is a major form of committing deanonymisation. In this attacks, the adversary tends to use some available information such as friendship or group relationship to infer sensitive properties that carry hidden values or behaviours [46], [41], [57]. Some techniques [58] use algorithms to infer about customers' transactions using auxiliary information about them with temporal changes of recommender systems. Danezis and Troncoso [59] use Bayesian inference to have knowledge about communication patterns and profiles information of users.

Linkage attacks: is another common form of practical deanonymisation attacks. In this attack, adversary can link auxiliary information of certain users with their anonymised records in a dataset [50], [22].

Structural attacks: it is the typical method to deanonymise graph using its structure [24] ,[28], [60]. Structural based deanonymisation attacks are proved theoretically and empirically by quantification [61] that exploiting social network structures is enough and even more powerful than other attacks that use seeding knowledge. A novel technique called Grasshopper [38] shows also higher effectiveness of deanonymisation than other state-of-the-art algorithms of structural attacks using seeding knowledge.

4.5 Deanonymisation Threats

According to the survey study [1], there are two particular deanonymisation threats; identification and linkability. The first threat, identification, leads to re-identifying records, IDs, users, nodes, links and some private attributes. The second threat, linkability, leads to re-identifying some relationships between users or accounts from different networks belonging to a certain user that are meant to be anonymised. Sharma et al [3] list the threats of privacy disclosure as identity, link and content disclosures.

The threats resulted from deanonymisation attacks are related to three major sets: identity, links and attributes. According to the survey study [1], there are two particular deanonymisation threats; identification and linkability. The first threat, identification, refers to any data disclosed to identify an identity of a user or any entity and also any attribute relates to them. Linkability refers to linking entities with links that meant to be anonymised or entities from different datasets. For instance, linking two accounts of a user from different resources with each other. There is another more specific classification [3] for privacy disclosure. We summarise this classification as follows.

Identity disclosure: which refers to deanonymising a user identity that was anonymised by removing the PII or by assigning a pseudonym.

Link disclosure: which refers to deanonymising relationships between users that meant to be hidden. This may be achieved by inference attack using observed links or users' attributes.

Content disclosure: which refers to revealing any anonymised data related to the target user such as address, email, etc.

4.6 Assistive Techniques

This part is critical as it shows how much some techniques can contribute to boost the deanonymisation attacks for more effectiveness.

Link predictions: is a method that the adversary uses to bridge the gaps between auxiliary information and target dataset if they are from different resources and have less in common for matching [1], [17].

Content analysis: is a method suggested in [1] for analysing a user's content in online social networks. This content may display usefulness in terms of providing information about the structure or some attributes to advance the deanonymisation attacks.

Spammer elimination: in online social networks there are users whom their accounts are created for spamming. Those accounts tend to make random links with other users. These developing edges may lower the effectiveness of deanonymisation attacks. Therefore, discovering and removing these spamming nodes may improve deanonymisation [1].

Adversary capabilities: feature matching task is a major capability for an adversary to think of committing a deanonymisation attack [1]. Other technical capabilities include prediction and inference, which require computational, statistical or probabilistic skills to compute the feasibility of deanonymisation.

4.7 Deanonymised Data

There are no specific methods or standard metrics for measuring the disclosed data resulting from deanonymisation. Most of the studies we found in the literature report their results in statistics compared to the sample of data they tested with precision and recall. Therefore, it is hard to think about deanonymisation approaches in terms of their effectiveness in a comparable way. The surveys [1] and [2], although following a systematic process, do not agree over a standard metric that can be used to measure the effectiveness of deanonymisation approaches.

5 Discussion

The aim of this research is to understand the problem of deanonymisation comprehensively in different contexts. We proposed a framework for analysis and

classifications to survey a number of studies in order to understand the deanonymisation problem comprehensively. We classified the approaches in which the deanonymisation is committed based on the available background. This shows a significant fact about how might trivial data from a resource can contribute to form feasible approaches to attack another anonymised data in totally different resource. That is why we conduct this survey to extend similar work in the literature in different contexts such as the Semantic Web. Among different approaches found in the literature, matching methodology represents the traditional way in which deanonymisation is carried out. It exploits different features in order to achieve the mapping and overlapping process and that basically depends on the structure of the target dataset and the auxiliary information. What we found in the literature that is some published data can be combined with totally different contextual data from other resources. Therefore, this shows the risk behind that which we do not expect. And that must affect the way we think in terms of widening our thinking about possible scenarios of attacks and threats. Which in turns affect the decision making about publishing data and the privacy protection that must accompany such data.

It is appealing to compare deanonymisation methods in terms of effectiveness and reported results. However, this not reasonable as every study has its own measures and the whole deanonymisation process depends on chosen methodology, methods, and features used from the auxiliary domain into the target domain. Approaches are very specific to the situation where these attacks achieved. Additionally, it is expected to be high expensive computationally if we think to test different parameters in deanonymisation experiments.

Finally, we envisage the deanonymisation problem is more challenging with the evolution of the Web to include Linked Data, open data, and big data. This is due to the fact that the technologies advancing the Web to be more discoverable, linked and open is actually forming the requirements of deanonymisation attacks to be more possible and practical or even automated.

6 Conclusions and Future Work

We analysed the problem of deanonymisation and we classified its approaches, attacks and threats for a comprehensive understanding. The techniques in the literature show evidence that this problem may become more serious than we thought. This stimulated by advancement in technology from two angles. Firstly, the Web has become the environment for publishing, storing and sharing information from various resources covering and serving different fields. Secondly, recent researches argued that the challenge for preserving privacy is getting greater as the data leakage is highly possible in the Web. Therefore, matching data from heterogeneous resources is feasible and thus deanonymisation can be achieved successfully. Also, in isolated environments such as health care, which hold sensitive information, can be deanonymised using data from the Web such as open government data.

In future work, we aim to model the attack patterns of deanonymisation. Also understanding how the changing context is affecting the effectiveness of deanonymisation. And if there are dead ends where deanonymisation cannot be achieved. More importantly, how to balance between providing a value from publishing data while protecting privacy from possible threats. This will advance research in data disclosure control and policy support.

Acknowledgments. This research is funded by University of Tabuk in Saudi Arabia and supported by Saudi Arabian Cultural Bureau in London.

References

1. Ding, X., Zhang, L., Wan, Z., Gu, M.: A Brief Survey on De-anonymization Attacks. In Online Social Networks, in International Conference on Computational Aspects of Social Networks. (2010) 611–615.
2. El Emam, K., Jonker, E., Arbuckle, L., Malin, B.: A systematic review of re-identification attacks on health data. PLoS One, Vol. 6. No. 12. p. e28071. Jan (2011)
3. Sharma, S., Gupta, P., Bhatnagar, V.: Anonymisation in social network: a literature survey and classification, International J. Social Network, Vol. 1. No. 1. pp. 51–66 (2012)
4. Toch, E., Wang, Y., Cranor, L. F.: Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems. User Model. User-adapt. Interact., Vol. 22. No. 1–2. pp. 203–220. Mar (2012)
5. Ohm, P.: Broken promises of privacy: Responding to the surprising failure of anonymization. UCLA Law Rev. (2010)
6. Alexin, Z.: Does fair anonymization exist?. Int. Rev. Law, Comput. Technol., Vol. 28. No. 1. pp. 21–44. Jan (2014)
7. Dwork, C., Naor, M.: On the Difficulties of Disclosure Prevention in Statistical Databases or The Case for Differential Privacy. J. Priv. Confidentiality, Vol. 2. No. 1. pp. 93–107. (2008)
8. O'Hara, K.: Transparent Government, Not Transparent Citizens: A Report on Privacy and Transparency for the Cabinet Office. (2011)
9. Sun, X., Wang, H., Zhang, Y.: On the identity anonymization of high-dimensional rating data. No. March (2011). pp. 1108–1122 (2012)
10. Sweeney, L.: k-ANONYMITY: A MODEL FOR PROTECTING PRIVACY. Int. J. Uncertainty, Fuzziness Knowledge-Based Syst., Vol. 10. No. 05. pp. 557–570. Oct (2002)
11. Bayardo, R. J., Agrawal, R.: Data Privacy through Optimal k-Anonymization. 21st Int. Conf. Data Eng. pp. 217–228.
12. Li, N.: Provably Private Data Anonymization: Or , k -Anonymity Meets Differential Privacy. (2010)
13. Machanavajjhala, A., Kifer, D., Gehrke, J., Venkitasubramaniam, M.: L-Diversity: Privacy Beyond k-Anonymity. ACM Trans. Knowl. Discov. Data, Vol. 1. No. 1, p. 3–es. Mar (2007)
14. Zhou, B., Pei, J.: The k-anonymity and l-diversity approaches for privacy preservation in social networks against neighborhood attacks. Knowl. Inf. Syst., Vol. 28. No. 1. pp. 47–77. Jun (2010)
15. Li, N.: t -Closeness: Privacy Beyond k -Anonymity and -Diversity. In ICDE, Vol. 7. pp. 106-115. (2007)
16. Domingo-Ferrer, J., Torra, V.: A Critique of k-Anonymity and Some of Its Enhancements. In Third Int. Conf. Availability, Reliab. Secur. pp. 990–993. Mar (2008)
17. Narayanan, A., Shi, E., Rubinstein, B. I. P.: Link Prediction by De-anonymization: How We Won the Kaggle Social Network Challenge. In Neural Networks (IJCNN). (2011)
18. Sharad, K., Danezis, G.: De-anonymizing D4D Datasets. In Workshop on Hot Topics in Privacy Enhancing Technologies. (2013)
19. Narayanan, A., Shmatikov, V.: Robust De-anonymization of Large Sparse Datasets. In IEEE Symposium on Security and Privacy. pp. 111–125. (2008)
20. Bender, S., Brand, R., Bacher, J.: Re-identifying register data by survey data: An empirical study. Stat. J. United Nations ECE, Vol. 18. No. 00311. pp. 373–381. (2001)
21. Gulyás, G., Imre, S.: Analysis of identity separation against a passive clique-based de-anonymization attack. Infocommunications J. pp. 1–10. (2011)
22. Torra, V., Stokes, K.: A formalization of re-identification in terms of compatible probabilities. CoRR, Vol. abs/1301.5. pp. 1–20. (2013)
23. Datta, A., Sharma, D., Sinha, A.: Provable De-anonymization of Large Datasets with Sparse Dimensions. In Principles of Security and Trust (2012)
24. Gulyas, G. G., Imre, S.: Measuring Importance of Seeding for Structural De-anonymization Attacks in Social Networks. In The Sixth IEEE Workshop on SECurity and SOCial Networking. pp. 610–615. (2014)
25. Hay, M., Miklau, G., Jensen, D.: Resisting structural re-identification in anonymized social networks. Proceedings of the VLDB Endowment 1.1 (2008) 102-114.
26. Dankar, F. K., El Emam, K., Neisa, A., Roffey, T.: Estimating the re-identification risk of clinical data sets. BMC medical informatics and decision making. 12.1 (2012) 66.

27. Cecaj, A., Mamei, M., Bicocchi, N.: Re-identification of anonymized CDR datasets using social network data. In The Third IEEE International Workshop on the Impact of Human Mobility in Pervasive Systems and Applications. (2014) 237–242.
28. Zhang, A., Xie, X., Chang, K. C.-C., Gunter, C. A., Han, J., Wang, X. F.: Privacy Risk in Anonymized Heterogeneous Information Networks. In EDBT (2014)
29. Pedarsani, P., Grossglauser, M.: On the privacy of anonymized networks. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11 (2011) 1235.
30. Zhu, T., Wang, S., Li, X., Zhou, Z., Zhang, R.: Structural Attack to Anonymous Graph of Social Networks. Math. Probl. Eng., Vol. (2013) 1–8
31. Narayanan, A., Shmatikov, V.: De-anonymizing Social Networks. In 30th IEEE Symposium on Security and Privacy. (2009) 173–187
32. Srivatsa, M., Hicks, M.: Deanonymizing Mobility Traces: Using Social Networks as a Side-Channel. Proceedings of the 2012 ACM conference on Computer and communications security. ACM (2012)
33. Sharad, K., Danezis, G.: An Automated Social Graph De-anonymization Technique. arXiv Prepr. arXiv1408.1276. (2014)
34. Nilizadeh, S., Kapadia, A., Ahn, Y.-Y.: Community-Enhanced De-anonymization of Online Social Networks. In CCS'14 (2014)
35. Peng, W., Li, F., Zou, X., Wu, J.: A Two-Stage Deanonymization Attack against Anonymized Social Networks. IEEE Trans. Comput., Vol. 63. No. 2. pp. 290–303 (2014)
36. Backstrom, L., Dwork, C., Kleinberg, J.: Wherefore Art Thou R3579X? Anonymized Social Networks, Hidden Patterns, and Structural Steganography. Proceedings of the 16th international conference on World Wide Web. ACM (2007)
37. Bringmann, K., Friedrich, T., Krohmer, A.: De-anonymization of Heterogeneous Random Graphs in Quasilinear Time. In ESA. (2014) 197–208
38. Simon, B., Gulyás, G. G., Imre, S.: Analysis of Grasshopper, a Novel Social Network De-anonymization Algorithm. Periodica Polytechnica Electrical Engineering and Computer Science, Vol. 58. No. 4. (2014) 161–173
39. Kazemi, E., Hassani, S. H., Grossglauser, M.: Growing a Graph Matching from a Handful of Seeds. In the 41st International Conference on Very Large Data Bases. (2015)
40. Ding, X., Zhang, L., Wan, Z., Gu, M.: De-Anonymizing Dynamic Social Networks. In IEEE Global Telecommunications Conference - GLOBECOM. (2011) 1–6
41. Gambs, S., Killijian, M.-O., Núñez del Prado Cortez, M.: De-anonymization attack on geolocated data. J. Comput. Syst. Sci., Vol. 80. No. 8. pp. 1597–1614. Dec (2014)
42. Ji, S., Li, W., Srivatsa, M., He, J. S., Beyah, R.: Structure based Data De-anonymization of Social Networks and Mobility Traces. (2014)
43. Okuno, T., Ichino, M., Kuboyama, T., Yoshiura, H.: Content-Based De-anonymisation of Tweets. In the Seventh International Conference on Intelligent Information Hiding and Multimedia Signal Processing. (2011) 53–56
44. FU, H., ZHANG, A., XIE, X.: Effective Social Graph De-anonymization based on Graph Structure and Descriptive Information. ACM Trans. Intell. Syst. Technol. (2008)
45. Unnikrishnan, J., Naini, F. M.: De-anonymizing Private Data by Matching Statistics. Allerton Conference on Communication, Control, and Computing. No. EPFL-CONF-196580 (2013)
46. Zheleva, E., Getoor, L.: To Join or Not to Join: The Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles. (2009) 531–540
47. Wondracek, G., Holz, T., Kirda, E., Kruegel, C.: A Practical Attack to De-anonymize Social Network Users. In IEEE Symposium on Security and Privacy. (2010) 223–238
48. Korayem, M., Crandall, D. J.: De-anonymizing Users Across Heterogeneous Social Computing Platforms. In Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media. (2013) 1–4
49. Lane, N. D., Xie, J., Moscibroda, T., Zhao, F.: On the feasibility of user de-anonymization from shared mobile sensor data. In Proceedings of the Third International Workshop on Sensing Applications on Mobile Phones - PhoneSense '12. (2012) 1–5
50. Merener, M. M.: Theoretical Results on De-Anonymization via Linkage Attacks. Transactions on Data Privacy 5.2 (2012) 377-402
51. Frankowski, D., Cosley, D., Sen, S., Terveen, L., Riedl, J.: You are what you say: privacy risks of public mentions. In Proceedings of the 29th SIGIR'06. (2006) 565–572

52. Malin, B., Sweeney, L.: How (not) to protect genomic data privacy in a distributed network: using trail re-identification to evaluate and design anonymity protection systems. Journal of biomedical informatics. 37.3 (2004) 179-192
53. Malin, B., Sweeney, L., Newton, E.: Trail Re-Identification: Learning Who You Are From Where You Have Been. Workshop on Privacy in Data (2003)
54. Foukarakis, M., Antoniades, D., Antonatos, S., Markatos, E. P.: On the anonymization and deanonymization of netflow traffic. In Proc. of FloCon. (2008)
55. Biryukov, A., Pustogarov, I., Weinmann, R.-P.: Trawling for Tor Hidden Services: Detection, Measurement, Deanonymization. In 2013 IEEE Symposium on Security and Privacy. (2013) 80–94
56. Pataky, M.: DE-ANONYMIZATION OF AN INTERNET USER BASED ON HIS WEB BROWSER. In CER Comparative European Research. (2014) 125–128
57. Danezis, G., Troncoso, C.: You Cannot Hide for Long: De-Anonymization of Real-World Dynamic Behaviour. In WPES'13. (2013) 49–59
58. Calandrino, J. A., Kilzer, A., Narayanan, A., Felten, E. W., Shmatikov, V.: You Might Also Like: Privacy Risks of Collaborative Filtering. Privacy Risks of Collaborative Filtering. IEEE Symposium on Security and Privacy. IEEE (2011)
59. Danezis, G., Troncoso, C.: Vida: How to use Bayesian inference to de-anonymize persistent communications. In Privacy Enhancing Technologies. (2009)
60. Ji, S., Li, W., Srivatsa, M., Beyah, R.: Structural Data De-anonymization: Quantification, Practice, and Implications. In CCS'14 (2014)
61. Ji, S., Li, W., Gong, N. Z., Mittal, P., Beyah, R.: On Your Social Network De-anonymizablity: Quantification and Large Scale Evaluation with Seed Knowledge. In The 2015 Network and Distributed System Security (NDSS) Symposium, San Diego, CA, US. (2015) 8–11.