2023 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

The Impact of Linkability On Privacy Leakage

Ahmad Hassanpour
Dept. of Information Security
Norwegian University of Science and Technology (NTNU)
Gjovik, Norway
Ahmad.Hassanpour@ntnu.no

Masrur Masqub Utsash
Dept. of Information Security
Norwegian University of Science and Technology (NTNU)
Gjovik, Norway
Masrur.M.Utsash@ntnu.no

Bian Yang
Dept. of Information Security
Norwegian University of Science and Technology (NTNU)
Gjovik, Norway
Bian.Yang@ntnu.no

***Abstract*—Online Social Networks are responsible for disclosing a large amount of sensitive information. Often, users unknowingly disclose vast amounts of sensitive and potentially (un)related data, oblivious to the associated privacy risks. Our research provides a comprehensive evaluation of the linkability between user profiles and shared content across various OSNs, a factor that has considerable implications for privacy leakage. We introduce a novel method for quantifying the linkability between profiles across multiple networks, based on key features and metrics that capture profile similarities. We applied this methodology to a dataset of user profiles across three online social networks named Flickr, Facebook, and Twitter. Our approach includes examining both structured and unstructured data related to user profiles, enabling us to offer a valuable understanding of linkability trends and identify potential privacy risks. Through our findings, we aim to inform the development of privacy-enhancing technologies and contribute to improving the current privacy landscape within OSNs. Our research underscores the critical need for robust privacy measures in the face of the growing interconnectedness of user data across different social networks.**

***Index Terms*—Linkability, Online Social Networks, Privacy Leakage.**

I. Introduction

The rise of the World Wide Web has significantly changed the fundamentals of human interaction because of the increasing use of information communication technologies in the modern digital society. Online social networks (OSNs) (e.g., Facebook, Twitter, LinkedIn, Reddit) provide an environment through which individuals may interact, share knowledge, express their emotions, and establish and preserve relationships with other online users [28]. This advancement of technology is accompanied by huge privacy concerns as most of the users tend to publish a lot of valuable information in the form of both *structured* (e.g., name, phone number, address, workplace, school) and *unstructured* data (e.g., text, image, video) [33] without even knowing them consciously [6]. Therefore, OSNs serve as a crucial platform for exposing personal information by enabling users to share their activities and engage with others through different means which can lead to violation of users' privacy in various aspects [16].

Protecting users' privacy in OSNs is a multifaceted challenge that requires consideration of all dimensions of privacy, including personal, contextual, and societal factors [34]. Although OSNs offer policies and privacy settings to regulate user profiles and posts [7], but the used language is complex and difficult to understand, making users vulnerable to privacy breaches [29]. Moreover, OSN providers collect, process, and analyze user data, and may also sell this data to third parties for advertising and marketing purposes [24]. Consequently, researchers have investigated privacy from various perspectives, including social, legal, and technical, in order to prevent privacy breaches and improve privacy protections in OSNs.

Previous experimental findings have revealed conflicts between privacy controls and the functionalities offered by OSNs which allows a range of privacy exploits such as indicating a misalignment between users' desired level of privacy control and the actual outcomes achieved [19]. Moreover, users now a days tend to use multiple OSNs for separate purposes as the primary capabilities differ from one another and users disclose different types of private information within those platforms. As a result, being able to link one user's multiple OSN profiles can lead to increase privacy leakage because of the access to more diverse private information [5] [1]. Cross-linking multiple OSN platforms thus can facilitate profile and data correlation, leading to inadvertent information sharing and privacy breaches. In the context of measuring privacy leakage, several previous experiments [21] [26] [11] [22] suggested that the privacy quotient is calculated based on *sensitivity* (the level of confidentiality and potential harm if disclosed [23]) and

This work was supported by the Project Privacy Matters (PRIMA) under Grant H2020-MSCA-ITN-2019-860315.

ASONAM '23, November 6-9, 2023, Kusadasi, Turkey

ACM ISBN 979-8-4007-0409-3/23/11…$15.00
http://dx.doi.org/10.1145/3625007.3627XXX

visibility (the extent to which data can be accessed, viewed, or shared by other entities [14]). Inspired from there, Hassanpour et al. [15] proposed an adaptive privacy leakage calculating framework where an additional and important metric called *linkage* has been introduced which can be used alongside sensitivity and visibility in order to obtain a more accurate privacy leakage score calculation.

Our work is shown the impact of linkability on privacy leakage. We evaluate the degree of connection between identities and published content on distinct OSNs. In the context of OSN, the linkability score is particularly important for privacy leakage score calculation due to the vast amount of personal information that is shared and interconnected across different platforms. For example, if a user's Facebook and Twitter accounts are linked, it may be possible to infer additional information about the user based on their activity across both platforms which might increase the privacy leakage score. On the other hand, the disclosure of user interests on one OSN may be accompanied by the publication of conflicting or unrelated information on the same or other OSNs, resulting in a potential fluctuation or decrease in the privacy score due to the contradicting nature of the newly shared information.

The objective of this paper is to evaluate the degree of linkability between user profiles across multiple OSNs which can be referred as *user profiling* and also evaluate the linkability of users' published information among different OSNs. User profiling is the process of constructing a thorough description of a specific user based on that person's actions, preferences, interests, and other crucial characteristics, thereby gathering insights into their behaviors and preferences [12] [17]. To address this challenge, we propose a method for measuring linkability between profiles across multiple networks, based on a set of features and metrics that capture the similarity between the profiles. Additionally, we conducted an analysis to establish connections between users' posts across multiple online social networks (OSNs) with the aim of identifying the linkability of users' shared information. Our method provides a quantitative assessment of linkability, enabling the identification of potential privacy risks and informing the development of privacy-enhancing technologies. We apply our method to a dataset of user profiles across multiple social networks (i.e., Flickr, Facebook, Twitter) and present our findings, demonstrating the utility of our approach in measuring linkability between profiles and among published posts.

In order to accomplish our research objective, we used two main methods to evaluate linkability. The first strategy concentrated on looking at user profiles' related structured data, whereas the second strategy looked at unstructured data. We sought to obtain a thorough grasp of linkability trends by methodically examining both types of data. We then used the results from these two methodologies to make a firm judgement on the degree of linkability between user IDs. We were able to examine and quantify the linkability of profiles across various OSNs using this integrated technique.

II. Background and Related Works

For calculating the privacy leakage score, numerous scholars have suggested multitude of methodologies among which two major approaches are significant namely *statistical-based* and *machine learning based* models. The first method relies on two intuitive properties which are sensitivity and visibility. This statistical-based method works on Dichotomous variables (takes only one of two possible values) or Polytomous variables (having more than two possible categories, either ordered or unordered) or a combination of them. On the other hand, ML-based models mostly try to measure the privacy of unstructured data (e.g., text, image, video).

One of the pioneering efforts towards the development of a privacy metric for online social networks was put forward by Maximilian et al. [21] in 2009 where the authors have proposed a formulation (1) to calculate the privacy score based on the sensitivity β_i and the visibility $v(i,j)$ of profile items $i \in \{1,\ldots,n\}$ of user j in a social network.

$$PR(i) = \sum_i PR(i,j) = \sum_i \beta_i \times v(i,j) \tag{1}$$

On the other hand, over the past decade, there has been growing interest in the analysis of linkability between user profiles on online social networks (OSNs) [2]. Linkability refers to the ability to link or associate various pieces of information or activities to a specific individual or entity, even if that information or activity is intended to be anonymous or separate [32]. A number of previous studies have explored various aspects of this issue, such as, identification of common patterns in user behavior across multiple platforms for measuring linkability scores [20]. These efforts have contributed to a greater understanding of the privacy risks associated with social media use and have laid the groundwork for the development of more effective privacy protection measures. In order to identify the linkability among two different sources, Chandok S. [5] proposed two separate methods which can be used against identities drawn from same or separate OSNs. First, *weighted sum method*, where the linkability score is calculated based on the similarity of feature using a function of feature and metrics. In this method, Computation of linkability score is performed in two steps namely *feature similarity indicator* and *linkability score calculator*. Once the weights have been assigned, the weighted sum score is calculated for each pair of profiles. The score represents the degree of linkability between the profiles, with higher scores indicating a greater risk of privacy leakage. Second, *probabilistic method*, where the intuitive idea is that the linkability score relies probabilistically on the feature similarity values. This approach defines linkability score as the probability of discovering two identities that are identical based on how similar their attributes are. Goga et al. [13] found that it is possible to link a user across multiple OSNs using information inherited from posted content. The researchers focused on components such as geo-location, post timestamp, and writing style to analyze their approach, using Yelp, Twitter, and Flickr as examples. Labitzke et al. [18] had shown the possibility of

profile correlations based on extracted friends lists even under legal and technical constraints.

However, upon thorough review of existing literature pertaining to the computation of privacy leakage scores in association of the consideration of linkability, limited references were identified. Hassanpour et al. [15] suggested that the linkability between posts or actions can significantly impact privacy, leading to an increase or decrease in the Privacy Leakage Score (PLS). He proposed a formula for calculating the privacy leakage score considering linkage score along with visibility and sensitivity.

$$PLS = sensitivity \times linkage \times visibility \tag{2}$$

The scarcity of relevant studies in this domain suggests a research gap in comprehensively addressing this particular aspect. The paucity of prior work underscores the need for further exploration and development of methodologies for accurately quantifying privacy leakage scores. By acknowledging this knowledge gap, our study contributes to the existing body of research by proposing novel approaches and methodologies in the calculation of privacy leakage score.

III. Design Experiment

This section expounds on our preliminary computations for the linkability score metric, which serves as a measure of the degree of association between two distinct online identities belonging to the same user on different online social networks (OSNs). Specifically, we delve into the derivation of features from user provided information and activity, which are subsequently used to construct activity profiles. The efficacy of the linkability score is evaluated by examining how accurately it estimates the extent to which two given identities are linkable. This evaluation helps us understand the usefulness of the linkability score in identifying potential privacy breaches and mitigating them.

A. Targeted OSNs

In our study, we sought to test the efficacy of our proposed linkability score metric across multiple OSNs. To accomplish this, we created datasets from three distinct OSNs, Facebook, Twitter and Flickr. By employing diverse datasets, our study aims to transcend platform-specific and user-specific limitations, enabling broader generalizability of our findings across various user populations and online contexts. Here, we overview three OSNs used in our study.

Facebook is a social networking platform that allows users to create a personal profile, share text, photos and videos, connect with friends and family, join interest groups, and engage in various activities such as playing games and participating in online events. The Facebook users can decide to whom he wants to limit the information that he is publishing through customized option for privacy. As of 2022, it has the highest number of monthly active users among all online social networks worldwide, with approximately 2.96 billion users [10]. Every minute, about 400 new users register on Facebook. Simultaneously, over 510,000 comments, 293,000 status updates, and 136,000 photos are posted, with 4 million posts being liked [27].

Twitter is a micro-blogging OSN where registered users (known as tweeters) post short messages (called tweets), which can be include text, photo, or videos. Some Twitter users choose to make their tweets public, making them accessible to anybody, even those without a Twitter account. Whereas others only allow their so-called followers, or Twitter users who have specifically asked for and received access to their tweets. Political figures, journalists, sportsmen, and other celebrities have all joined Twitter, making it one of the most widely used and diversified OSNs today. As of 2022, Twitter has a monthly active user base of around 450 million, showing an audience growth of over 40% since 2018 [31].

Flickr is an online social network and cloud storage provider, specializing in the sharing of multimedia content, specifically photographs and videos. This platform allows users to annotate their multimedia content with text, which enhances the user experience by providing additional context to the content. In order to post or view restricted content on Flickr, an account is generally required. However, public content can be viewed by anyone without an account. Additionally, Flickr has a unique feature known as *contacts*, which is similar to the concept of friends or connections on Facebook and Twitter. As of 2022, the registered user base of Flickr exceeds 112 million, with 60 million being categorized as active users, defined as those who access the platform at least once a month [4].

B. Collected Data

Our experiment was conducted using a dataset that consisted of both structured and unstructured data, which was obtained from Facebook, Twitter, and Flickr. As a measure to ensure compliance with the General Data Protection Regulation (GDPR), we exclusively extracted publicly posted information of users from these platforms. However, the initial challenge that we encountered was to identify users who had accounts across all three of the aforementioned OSNs. In order to overcome this obstacle, we leveraged the feature on Flickr that allowed users to mention their associated user accounts on other online social networks. This feature was instrumental in identifying our target users, which served as the *ground truth* for the dataset that we acquired.

Out of a random selection of 5,473 Flickr users, we were able to sort out 45 users who possessed registered accounts across all our targeted OSNs. The structured data that we collected consisted of users' name, location (both current and hometown), and user name, occupation and some other publicly available information. Alongside this, we obtained unstructured data that included bio, texts, images, and image captions. Our efforts in acquiring this data were aimed at analyzing the linkability score across diverse user pairs and evaluating the effectiveness of the score in estimating the likelihood of linkability between two given identities.

Moreover, we gathered a total of 2264, 1684, and 693 images from Flickr, Facebook, and Twitter, respectively, sourced from 45 users across these platforms.

C. Methodology

To ascertain the linkability between users' profiles, we employed two distinct approaches. The first approach involved user profiling, where we leveraged the structured data openly shared by the users. By analyzing attributes such as name, user name, and location, we aimed to link profiles that capture the essence of each user. The second approach focused on identifying the content correlation within an individual's data. This involved examining the relationships between published image files, to unveil patterns and associations that contribute to the linkability between profiles. By combining these approaches, we aimed to gain a comprehensive understanding of the linkability dynamics present in users' online profiles.

1) User Profiling Measuring: In order to calculate the similarity of the attributes from different OSNs, we used the *bert-base-nli-mean-tokens* model [25]. BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained language model developed by Google [9] that uses a bidirectional transformer architecture to create deep contextual representations of words in text. It has shown state-of-the-art performance in various natural language processing tasks, such as text classification and sentence similarity [35]. 'Bert-base-nli-mean-tokens' stands for BERT base model for natural language inference using mean pooling of the token embeddings which is a fine-tuned version of the original BERT language model.

Initially, we performed similarity calculations on the users belonging to the same OSN. We used cosine similarity metrics to calculate similarity. These similarity calculations were performed on various attributes of the users such as their name, location, and user name. Once we had obtained the similarity values for each attribute, we combined them to obtain an overall similarity score. To achieve this, we took the average of the similarity scores across all attributes for each pair of users. After calculating the similarity values for users within the same OSN and obtaining the average of the similarity values based on selected attributes, we proceeded to calculate the similarity for cross-OSN users. For this, we employed the same method of using the BERT-based embedding model to obtain the vector representations of the user attributes. Next, we computed the cosine similarity (3) between the vector representations of each pair of users from different OSNs. This resulted in a similarity score for each pair of users that belonged to different OSNs. These similarity scores were then normalized to obtain a value between 0 and 1. From the matrix of calculated values, we used Top-1 (identifying the single best choice or outcome among multiple options) and Top-3 approach (considering 3 best choices instead of just one among multiple options) to identify the best linkability among the entities. For Top-1 approach, only the single best choice or outcome among multiple options would be considered, whereas for a Top-3 approach the three best choices would be included instead of just the best one [3].

Overall, this approach allowed us to estimate the linkability between users across different OSNs by leveraging the similarity between their attributes, as captured by the BERT-based embedding model.

2) Content Linkage Measuring: In this work, we compared and measure the linkilibity of posts' content between Flickr, Facebook, and Twitter for a specific user. We consider the image modality since Flickr manily are being used for posting pictures. To measure the linkilibilty between each pair of images, we first extract a representation vector for each image using a deep learning model called EfficientNetV2 [30] which is trained on ImageNet dataset [8]. Then, to calculate the similarity between each pair of images, we utilize cosine similarity as below:

$$similarity = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} \tag{3}$$

where $\mathbf{A}$ and $\mathbf{B}$ are the extracted representation vectors for the first (from Flickr) and the second (from Twitter or Facebook) images, respectively. For each image in Flickr, we found the most similar image in Facebook (or Twitter when we are comparing posts' content in Flickr and Twitter).

IV. Results

This section provides an overview of the results obtained from the proposed linkability scoring methods. In particular, we present the computed linkability scores for the selected approaches and analyze the outcomes. These results serve as a basis for evaluating the effectiveness of the methods and their potential for accurate identity and content linkage across multiple OSNs.

A. Data Analysis

In order to gain a better understanding of the collected data for the 45 shortlisted users, we conducted some statistical analyses. In our dataset we found 28.89% female and 71.11% male population who were selected totally randomly (Table I). In contemporary times, online social networks have become integral to daily life, and individuals frequently disclose their personal information, including contact details, birthdays, relationship statuses, and political and religious affiliations. While some may inadvertently reveal sensitive information, others may do so intentionally.

TABLE I
Gender distribution in dataset

Gender	Percentage
Male	71.11%
Female	28.89%

In our obtained dataset, users disclosed their personal information to various extend. Such as, the ratio of location (hometown and/or current location) disclosure is 75.55% for

Facebook, 93.33% for Twitter and 97.77% for Flickr (Table II). Among all, 73.33% of Flickr users shared their occupation

TABLE II
LOCATION DISCLOSURE RATE

OSN	Location Disclosure
Facebook	75.55%
Twitter	93.33%
Flickr	97.77%

and among the Facebook users, 66.66% disclosed that (Table III). Users also shared *bio* which typically refers to a short

TABLE III
OCCUPATION DISCLOSURE

OSN	Work Info Disclosure
Facebook	66.66%
Flickr	73.33%

written description or summary that a user includes on their profile to provide information about themselves. We found that 97.77% Flickr users, 88.88% Twitter users and 60% Facebook users shared this type information on their profile which can directly cause privacy leakage if displayed to unintended audience (Table IV). We also had access to some

TABLE IV
SHORT USER INTRODUCTION RATE

OSN	Bio Disclosure
Facebook	60%
Twitter	88.88%
Flickr	97.77%

other sensitive information as the users shared those publicly. Such as, 35.55% users shared their relationship status, 6.66% users disclosed their date of birth and 20% users mentioned their email address within the OSNs (Table V). Among the

TABLE V
OTHER SENSITIVE INFO

Sensitive Info	Disclosure
Relationship status	35.55%
Date of birth	6.66%
Email address	20%

unstructured data, every user shared images in their profile (at least one) but there was some variations for text data they shared. We also found 4.44% users who intentionally chose to protect their shared content from the mass public in Twitter.

B. Profiling Linkage performance

After conducting our experiment on the acquired dataset, we found that a significant portion of the users were linkable using our proposed method. Specifically, our approach was able to identify connections between users across multiple online social networks with a high degree of accuracy. However, it is important to note that accuracy may vary depending on the specific attributes and features used in the analysis. In order to fully understand the effectiveness of our approach, we analyzed the accuracy rate that we obtained during the experiment.

During the evaluation process, we meticulously compared each online social network with the others and discovered varying accuracy rates for each pair of OSN. Such as, while calculating the linkability for Facebook against Flickr we found 91.12% accuracy and against Twitter 80% accuracy based on similarity score.

TABLE VI
FACEBOOK VS. FLICKR & TWITTER

OSN Name	**Accuracy against Facebook**	
	Top-1	***Top-3***
Flickr	88.88%	91.12%
Twitter	71.11%	80%

For the analysis of Flickr against Twitter we obtained 95.56% accuracy and against Facebook 97.78%.

TABLE VII
FLICKR VS. TWITTER & FACEBOOK

OSN Name	**Accuracy against Flickr**	
	Top-1	***Top-3***
Twitter	88.88%	95.56%
Facebook	95.55%	97.78%

During the calculation of Twitter against Flickr we get 97.78% accuracy for and Facebook we get 91.12% accuracy.

TABLE VIII
TWITTER VS. FLICKR & FACEBOOK

OSN Name	**Accuracy against Twitter**	
	Top-1	***Top-3***
Flickr	91.11%	97.78%
Facebook	91.11%	91.12%

C. Content Linkage Performance

After linking user profiles, we generate separate poll of images for each user in different OSNs. Thus, using similarity score, the most similar image for each image in Flickr has been found in Facebook for each user. We done the same process between Flickr and Twitter polls. The distribution similarity score for both cases (i.e., Flickr-Facebook and Flickr-Twitter) have been shown in Figures 1 and 2. A similarity score between 0.9-1 shows a high degree of similarity between the images being compared. In this context, we operate under the

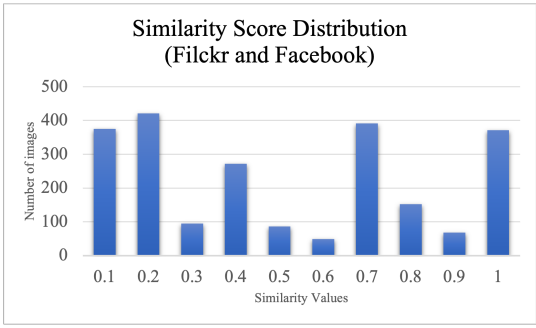

Fig. 1. Similarity score distribution between Flickr (including 2264 images) and Facebook (including 1684 images) images. The x and y axis show the cosine similarity values and number of images, respectively.

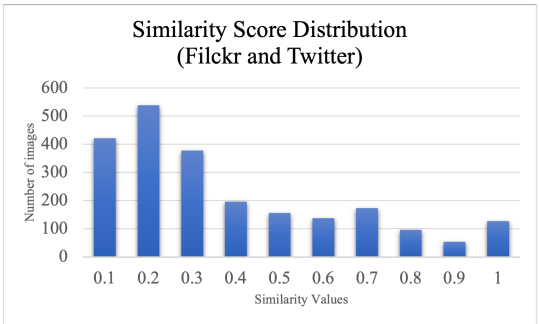

Fig. 2. Similarity score distribution between Flickr (including 2264 images) and Twitter (including 693 images) images. The x and y axis show the cosine similarity values and number of images, respectively.

assumption that a similarity score exceeding 0.8 (estimated empirically) indicates nearly identical images, distinguished by only a minimal degree of variation. Considering this threshold, as depicted in Figure 1, approximately 450 images that were shared on Flickr have also been published on Facebook.

V. Discussion

The first part of our experiment aimed at linking profiles across various Online Social Networks (OSNs) using minimal profile attributes, specifically, name, user name, and location. Evidently, the accuracy of profile linkage would potentially increase if additional information was incorporated into the process. As can be deduced from Tables VI, VII, and VIII, taking into account the top-1 accuracy, the most accurate results were achieved when information from Flickr profiles was used to establish a link with a corresponding profile on Facebook. This suggests a high degree of similarity or overlap between the information disclosed on Flickr and Facebook profiles. In simpler terms, it appears that users tend to share closely matching information across these two platforms. Therefore, if you have access to a user's Flickr profile, there's a higher likelihood of accurately linking it to the same user's Facebook profile compared to Twitter. This observation emphasizes the impact and value of shared data across multiple social media platforms in enhancing the accuracy of profile linkage.

Furthermore, the latter segment of our study indicates that a significant number of images shared on Flickr do not appear on other Online Social Networks (OSNs). This situation can result in a heightened risk of privacy breaches if profiles across different OSNs are linked. Take, for example, a sample of 2264 images uploaded on Flickr. Our findings reveal that roughly 20 percent of these images were also found on Facebook, and a smaller portion, about 8 percent, surfaced on Twitter. These statistics suggest a lower likelihood of successfully establishing a link between profiles on Flickr and Twitter due to the reduced overlap in shared content. However, it's essential to note that while the probability of linking is lower, the potential for privacy leakage escalates dramatically. The reason being, the content shared on these two platforms is distinctly different. Therefore, if a link is established, it would expose a broader range of the user's information, potentially revealing aspects of their personal lives that they intended to keep separate on these individual platforms. This underscores the critical need for users to be conscious of the data they share across different social networks, given the potential risks associated with profile linkage across multiple platforms.

It is important to clarify that our analysis on content linkage is currently focused solely on the image modality. However, this does not limit the application of our techniques. They can indeed be adapted for other unstructured data types, such as text and video. This would entail using appropriate deep learning models to extract representative vectors from these data types, and then leveraging cosine similarity as a measure of distance between these vectors, much like we have done with images.

VI. Conclusion

Our research on Online Social Networks (OSNs) focuses on the privacy implications arising from linkability of user profiles and shared content across different platforms. We developed a method to quantify this linkability, utilizing key attributes (i.e., name, user name, location) to determine profile similarities. Applying this to profiles and content from Flickr, Facebook, and Twitter, we examined both structured and unstructured data, offering a valuable view of linkability trends and potential privacy risks. Our findings highlight that minimal profile attributes can significantly enhance the accuracy of profile linkages, particularly between platforms like Flickr and Facebook where data overlap is significant. However, we also found a substantial number of images shared on Flickr do not appear on other OSNs, reducing the likelihood of profile linkage but paradoxically increasing potential privacy leakage. This discovery underscores the need for robust privacy measures given the increased interconnectedness of user data across OSNs. Our research emphasizes the importance of user consciousness in data sharing across different OSNs, considering the potential privacy risks of profile linkage. Our work aims to inform the development of privacy-enhancing technologies and strategies to better protect user privacy in OSNs.

REFERENCES

[1] Erfan Aghasian, Saurabh Garg, Longxiang Gao, Shui Yu, and James Montgomery. Scoring users' privacy disclosure across multiple online social networks. *IEEE access*, 5:13118–13130, 2017.

[2] Michael Backes, Pascal Berrang, Oana Goga, Krishna P Gummadi, and Praveen Manoharan. On profile linkability despite anonymity in social media systems. In *Proceedings of the 2016 ACM on Workshop on Privacy in the Electronic Society*, pages 25–35, 2016.

[3] Mathias Benedek, Caterina Mühlmann, Emanuel Jauk, and Aljoscha C Neubauer. Assessment of divergent thinking by means of the subjective top-scoring method: Effects of the number of top-ideas and time-on-task on reliability and validity. *Psychology of aesthetics, creativity, and the arts*, 7(4):341, 2013.

[4] Matic Broz. Flickr Statistics, User Count, amp; Facts (July 2023). 8 2022.

[5] Srishti Chandok and Ponnurangam Kumaraguru. *User identities across social networks: quantifying linkability and nudging users to control linkability*. PhD thesis, 2017.

[6] Jiayi Chen, Jianping He, Lin Cai, and Jianping Pan. Disclose more and risk less: Privacy preserving online social network data sharing. *IEEE Transactions on Dependable and Secure Computing*, 17(6):1173–1187, 2018.

[7] Sourya Joyee De and Abdessamad Imine. Choosing the right privacy settings. In *Privacy Risk Analysis of Online Social Networks*, pages 59–67. Springer, 2021.

[8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[10] S. Dixon. Facebook MAU worldwide 2023 — Statista, 5 2023.

[11] Josep Domingo-Ferrer. Rational privacy disclosure in social networks. In *Modeling Decisions for Artificial Intelligence: 7th International Conference, MDAI 2010, Perpignan, France, October 27-29, 2010. Proceedings 7*, pages 255–265. Springer, 2010.

[12] Magdalini Eirinaki and Michalis Vazirgiannis. Web mining for web personalization. *ACM Transactions on Internet Technology (TOIT)*, 3(1):1–27, 2003.

[13] Oana Goga, Howard Lei, Sree Hari Krishnan Parthasarathi, Gerald Friedland, Robin Sommer, and Renata Teixeira. Exploiting innocuous activity for correlating users across sites. In *Proceedings of the 22nd international conference on World Wide Web*, pages 447–458, 2013.

[14] Ralph Gross and Alessandro Acquisti. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80, 2005.

[15] Ahmad Hassanpour and Bian Yang. Prime: A novel privacy measuring framework for online social networks. In *2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 518–525. IEEE, 2022.

[16] Nadin Kökciyan and Pınar Yolum. P ri g uard: A semantic approach to detect privacy violations in online social networks. *IEEE Transactions on Knowledge and Data Engineering*, 28(10):2724–2737, 2016.

[17] Yasamin Kowsari. Identifying user profiles via user footprints. *arXiv preprint arXiv:2208.06251*, 2022.

[18] Sebastian Labitzke, Jochen Dinger, and Hannes Hartenstein. How i and others can link my various social network profiles as a basis to reveal my virtual appearance. In *4. DFN-Forum Kommunikationstechnologien*. Gesellschaft für Informatik eV, 2011.

[19] Yan Li, Yingjiu Li, Qiang Yan, and Robert H Deng. Privacy leakage analysis in online social networks. *Computers & Security*, 49:239–254, 2015.

[20] Siyuan Liu, Shuhui Wang, Feida Zhu, Jinbo Zhang, and Ramayya Krishnan. Hydra: Large-scale social identity linkage via heterogeneous behavior modeling. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, pages 51–62, 2014.

[21] E Michael Maximilien, Tyrone Grandison, Tony Sun, Dwayne Richardson, Sherry Guo, and Kun Liu. Privacy-as-a-service: Models, algorithms, and results on the facebook platform. In *Proceedings of Web*, volume 2, 2009.

[22] Raj Kumar Nepali and Yong Wang. Sonet: A social network model for privacy monitoring and ranking. In *2013 IEEE 33rd International Conference on Distributed Computing Systems Workshops*, pages 162–166. IEEE, 2013.

[23] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 75–84, 2007.

[24] Samia Oukemeni, Helena Rifà-Pous, and Joan Manuel Marquès Puig. Privacy analysis on microblogging online social networks: A survey. *ACM Computing Surveys (CSUR)*, 52(3):1–36, 2019.

[25] Lucia Passaro, Alessandro Bondielli, Alessandro Lenci, Francesco Marcelloni, et al. Unipi-nle at checkthat! 2020: approaching fact checking from a sentence similarity perspective through the lens of transformers. In *CEUR WORKSHOP PROCEEDINGS*, volume 2696. CEUR, 2020.

[26] Christoph Renner. Privacy in online social networks. *Swiss Federal Institute of Tech., Zurich*, pages 11–13, 2010.

[27] Jack Shepherd. 33 Essential Facebook Statistics You Need To Know In 2023, 6 2023.

[28] Agrima Srivastava and G Geethakumari. Measuring privacy leaks in online social networks. In *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 2095–2100. IEEE, 2013.

[29] Frederic D Stutzman, Ralph Gross, and Alessandro Acquisti. Silent listeners: The evolution of privacy and disclosure on facebook. *Journal of privacy and confidentiality*, 4(2):2, 2013.

[30] Mingxing Tan and Quoc Le. Efficientnetv2: Smaller models and faster training. In *International conference on machine learning*, pages 10096–10106. PMLR, 2021.

[31] Ash Turner. How Many Users Does Twitter Have? (Jul 2023), 6 2023.

[32] Yung Shin Van Der Sype and Walid Maalej. On lawful disclosure of personal user data: What should app developers do? In *2014 IEEE 7th International Workshop on Requirements Engineering and Law (RELAW)*, pages 25–34. IEEE, 2014.

[33] Chenguang Wang, Zhu Tianqing, Ping Xiong, Wei Ren, and Kim-Kwang Raymond Choo. A privacy preservation method for multiple-source unstructured data in online social networks. *Computers & Security*, 113:102574, 2022.

[34] Pamela Wisniewski, AKM Islam, Heather Richter Lipford, and David C Wilson. Framing and measuring multi-dimensional interpersonal privacy preferences of social networking site users. *Communications of the Association for information systems*, 38(1):10, 2016.

[35] Shijie Wu and Mark Dredze. Beto, bentz, becas: The surprising cross-lingual effectiveness of bert. *arXiv preprint arXiv:1904.09077*, 2019.