Inferring attributes with picture metadata embeddings

Bizhan Alipour Pijani, Abdessamad Imine, Michaël Rusinowitch

Inferring Attributes with Picture Metadata Embeddings

Bizhan Alipour Pijani
Lorraine University, Cnrs,
Inria, Loria
54506 Vandœuvre-lès
Nancy, France
bizhan.alipourpijani@loria.fr

Abdessamad Imine
Lorraine University, Cnrs,
Inria, Loria
54506 Vandœuvre-lès
Nancy, France
abdessamad.imine@loria.fr

Michaël Rusinowitch
Lorraine University, Cnrs,
Inria, Loria
54506 Vandœuvre-lès
Nancy, France
michael.rusinowitch@loria.fr

ABSTRACT

Users in online social networks are vulnerable to attribute inference attacks due to some published data. Thus, the picture owner's gender has a strong influence on individuals' emotional reactions to the photo. In this work, we present a graph-embedding approach for gender inference attacks based on pictures meta-data such as (i) alt-texts generated by Facebook to describe the content of images, and (ii) Emojis/Emoticons posted by friends, friends of friends or regular users as a reaction to the picture. Specifically, we apply a semi-supervised technique, node2vec, for learning a mapping of pictures meta-data to a low-dimensional vector space. Next, we study in this vector space the gender closeness of users who published similar photos and/or received similar reactions. We leverage this image sharing and reaction mode of Facebook users to derive an efficient and accurate technique for user gender inference. Experimental results show that privacy attack often succeeds even when other information than pictures published by their owners is either hidden or unavailable.

CCS Concepts

•**Computer systems organization** → **Embedded systems;** *Redundancy;* Robotics; •**Networks** → Network reliability;

Keywords

Social network, privacy, inference attack, gender inference, picture, Emojis, graph embedding

1. INTRODUCTION

Attribute inference from social network profiles and behaviors is a powerful mean to breach user privacy for malicious purpose or targeted advertisements. It amounts to derive private attributes of a target user (such as gender, age, political view, or sexual orientation) from publicly available data.

Existing works have investigated two types of attribute inference attacks on Facebook: friend-based [17] and behavior-based [1] inference attacks. Friend-based attacks follow the intuition that *you are who you know*. They proceed in two steps: the attacker first collects the friend list of the target user, and then from the target user and his/her friend's available data infers target hidden attributes. Behavior-based attacks follow the intuition that *you are how you behave*. In these attacks, the attacker monitors user behavior such as liked pages and joined groups to infer his/her private attributes. Most existing inference techniques proceed by analyzing data directly generated by the target user, or data obtained by crawling the user vicinity network. Recently, there has been a dramatic increase in the scalability of network embedding methods due to the introduction of the neural language model, word2vec [27]. The word2vec-based embedding methods, such as Deepwalk [31], and node2vec [18], analogize nodes into words and capture network structure via random walks, which results in a large corpus to train the node representations. Graph-based classification methods have applied to attribute inference attacks as well. [19] leverage graph-based method by considering either one or both types of attribute inference attacks, mentioned above, to infer target user attributes. However, these attacks consider either social friendship links or user behaviors that are rather unavailable to an attacker in the real scenario.

Unlike previous studies, we show how to detect Facebook user's gender through his/her shared images. With the huge amount of available information on Facebook, identifying user's gender from their online activities and shared data is an essential mechanism for targeted advertising or privacy breaking [7]. Gender is a valuable information source in developing more accurate classifiers for inferring other private attributes such as age [30]. In [15], the authors investigated *479k* Facebook users to determine the level of privacy awareness. They showed that about one-half of their collected Facebook users hide their gender. Note that users prefer to hide their gender for two reasons. First, they want camouflage against sexual harassment and stalking. The Facebook search bar lets users track down pictures of their female friends, but not the male ones [22]. Second, they want to reduce discrimination. Gender is the direct beneficial information that helps the private sector to present personalized services. Facebook faced criticism for enabling biased discrimination and misinformation. The American Civil Liberties Union (ACLU) [1] accused Facebook of en-

[1] https://www.aclu.org/blog/womens-rights/womens-rights-workplace/facebook-settles-civil-rights-cases-making-sweeping

abling employers to use targeting technology that excludes a woman from receiving job ads for some positions. Additionally, [8] studied how different kinds of self-presentation information on Facebook is interpreted by employers and the subsequent attraction in hiring decisions.

While many Facebook users hide their sensitive attributes (e.g., gender, age, political view), pictures are still available to the public. A social media sharing analysis conducted by *The New York Times* revealed that *68%* of their respondents share images to give people a better sense of *who they are* and *what they care about* [37]. Users in social media share pictures to receive feedback on their activities, especially from friends, and acquaintances, and provide a great sense of connectedness. However, they lose privacy control on their posted pictures due to extra information (i.e., *meta-data*) added by third-party during the publication process. Let us review this added *meta-data* that we consider in our attack.

(i) Generated alt-text. Facebook has designed and deployed automatic alt-text, a system to identify faces, objects, and themes from photos by applying computer vision technology. This system is proposed to help blind people to feel more connected and involved in Facebook. The alt-text generates a summary of the existing content for the image automatically. The technology can reliably recognize a list of *97* concepts (tags), including people (e.g., people count, smiling, child, baby), objects (e.g., car, building, tree, cloud, food), settings (e.g. inside restaurant, outdoor, nature) and themes (e.g., close-up, selfie, drawing) [39].

(ii) Emojis. Users in social media post Emojis to express their feelings directly. Since the *2010*s, Emojis emerged into communication to the point where Oxford Dictionaries[2] announced 😂, commonly known as *FACE WITH TEARS OF JOY*, as the word of the year.

(iii) Emoticons. An Emoticon[3] is a representation of human facial expression using only keyboard characters such as letters, numbers, and punctuation marks. They express emotions differently through facial gestures inside text-based communication.

1.1 Motivation

To increase awareness of Facebook users about threats on their privacy, we show that from very limited information, even when the user hides his/her comments, we can infer the user's gender. Previous gender inference attacks on Facebook have two main limitations. First, users friend-based and behavior-based data are extensively considered in the attack process, degrading prediction accuracy in the case of unavailability. Second, these attacks are limited to text-based knowledge. For example, a person's gender identity is reconstructed from linguistic features associated with male or female writing style on social media, decreasing the prediction accuracy when texts are multilingual or unavailable. In this work, we relax these two concrete limitations as follows:
1) We exploit non-user generated data (i) alt-text which is computed by Facebook, and (ii) Emojis/Emoticons added by other Facebook users while commenting on the picture.
2) We rely on Emojis/Emoticons as they are not limited to a specific language.

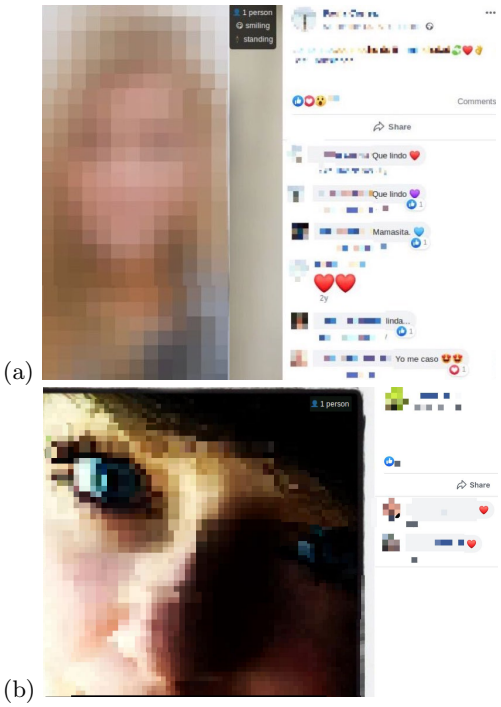


Figure 1: **Target user received comments: (a) Emojis and non-English words (b) only Emojis.**

The advantage of Emojis and Emoticons are twofold: it is a universal language, and it is a non-verbal communication way. On the other hand, alt-text is a widely available description of the picture that saves one from image processing tasks.
Female and male posted pictures can receive non-English comments as in Figure 1(a), or Emoji only comments as in Figure 1(b) and analyzing this non-user generated data is still sufficient to launch gender inference attack. Note, we have selected these female-owned pictures randomly from Facebook. For the sake of privacy, we blur the pictures and commenter's name. Consider the example in Figure 1(a), where the target user can be inferred as male since the picture received the Spanish word *lindo*. However, a word like *linda* may lead the inference process to an incorrect result. A possible remedy is to consider other available meta-data such as alt-text and Emojis into an end-to-end learning system. Figure 1(b) shows an example when the Emojis and generated alt-text are the only available data to the attacker. Emojis/Emoticons are language-independent and make the inference attack possible even when the received comments only consist of Emojis/Emoticons.
Our work shows that gender inference attack is possible even when, as in previous examples, essential information from the target user and his/her vicinity network is not available.

[2] `https://languages.oup.com/press/news/2019/7/5/WOTY`
[3] `https://en.wikipedia.org/wiki/Emoticon`

1.2 Problem Description

Using limited amounts of available data, we propose to investigate gender inference attack by leveraging non-user generated data. Therefore our attack is an indirect attack that targets Facebook users even when they are cautious about their privacy, and hides direct generated data such as friend list, liked pages, groups, writing style (e.g., comments), and profile attributes.

More concretely, we are interested in answering the following questions: Is there a significant difference in the usage of Emoji/Emoticon for commenting female and male-owned pictures? Do female and male receive different Emojis/Emoticons for pictures with the same theme and settings (similar alt-text tags)? Do female and male share picture with similar themes, setting, and objects? In addition to these questions, we also introduce the following assumptions:

1. We do not have access to the gender of commenters who wrote the comments underneath pictures.
2. We do not know whether the commenters are target user friends, friends of friends, or ordinary persons.
3. We do not perform any computer image processing on the target pictures.
4. We do not consider user profile name as an input to our attack process. Indeed, the Facebook user may use the shortened name as a chosen name. It is due to privacy concerns, and it is a popular tactic to be identifiable only to friends, but not so easily to a stranger.

The success of our gender inference attacks relies on finding accurate correlation of picture owner gender with alt-text, and Emojis/Emoticons used by commenters when reacting to these pictures. Supervised feature-based detection approaches are tedious and time-consuming to deal with this problem [32]. Indeed, they are based on manually assembling features and require an annotated dataset to be analyzed. Note that this labelling process might lack a deeper understanding of the interactions between users on Facebook.

1.3 Contributions

To tackle the above problem, we apply a semi-supervised approach called node2vec [18]. More specifically, we build a graph where nodes correspond to users and pictures (represented as sets of alt-texts and Emojis/Emoticons nodes) while edges denote picture ownerships. Next, we use a graph embedding model that automatically learn features by following an objective function to represent nodes in a low-dimensional vector space. The embedded vectors can be considered as the possible features for attribute inference attacks. In this work, we investigate vector representations capability in predicting Facebook users' gender from pictures metadata. For that, the generated vector representations are fed into the Logistic Regression classification algorithm. As a result, users who published similar pictures, or received similar reactions tend to be close to each other in the vector space.

In the following, the essence of our contributions and improvements over the previous works are:

1. Rather than considering the friend-based and behavioral-based data, which might be costly and unavailable in the real scenario, we provide a new approach for gender inference attack by considering picture meta-data.
2. We use Emojis/Emoticons as a universal and powerful language to infer the owner's picture gender. This inference has an advantage over any text-based inference attack as it is independent of any language restrictions.
3. We build a graph representing picture ownerships and we apply embeddings of this graph into low-dimensional vector space.
4. We conduct experiments to evaluate gender inference attacks by using AUC as a performance metric.

To ease reading, we introduce some specific definitions used throughout the paper. A *picture owner*, shortened as an *owner*, is the one who published pictures on Facebook. *Commenters* are other Facebook users who respond to owner pictures. A *response* is the commenters' Emojis/Emoticons preferences while commenting for owner pictures.

Outline. The paper is organized as follows: we review related work in Section *2*. In section *3*, we describe the gender inference attack. Section *4* presents our attack evaluation, and we conclude the paper in Section *5*.

2. RELATED WORK

In this section, we review recent works that are related to our research. For that, we consider three aspects: gender inference attack on social media, Emojis analysis, and graph-based approaches.

2.1 Gender Inference Attack on Social Media

Profiling users based on their activities has obtained great attention in the past decade. Especially, user profiling based on gender is important for recommendation systems. Recently, researchers have investigated on popular social media platforms to distinguish male and female based on content sharing [13] and behavior [26]. Prior works claimed that gender prediction is possible from the writing style of the target user [16], word usage [36] and phrase choice [33]. Gender inference attack by evaluating the target user name performed by [21] across major social networks. However, [35] proved that the performance of this type of attack is biased towards countries of origin. The authors of [12] propose user gender identification through user shared images in Fotolog and Flickr, two image-oriented social networks. They perform image processing task on each crawled image (in the offline mode), which is not feasible in an online attack.

To sum up, the above works depend on the availability of user-generated data, which is costly in a real scenario. In contrast, we perform gender inference attacks by relying only on small information that is not under the direct possession of the user. We do not explore the user network, which has two advantages: (i) makes the attack robust even when the entire personal data and his/her vicinity network is unavailable, and (ii) makes the attack suitable for online mode.

Additionally, our attack is not limited to textual language as we use Emoji/Emoticon, a universal language. We have shown the benefit of non-user generated data analysis to infer the picture owner gender by relying on the textual part of the comments, regardless of the Emoji/Emoticon usage [4]. However, this work is complementary to our previous work.

2.2 Emoji Usage Analysis

Several works have analyzed Emoji usage in recent years. Researchers have studied the individual intercept on messages containing Emojis [9]. They have performed experiments on how people use Emojis, an emerging universal language for stating emotions in different countries [24] and culture [6]. Emoji is a rich resource for sentiment analysis and emotion measurement. For example, [2] performs the first quantitative study to correlate Emoji usage to its semantic. Additionally, [5] analyzed messages of *Wechat*, and *IM APP* users in China, to learn the diversity of usage preferences of Emoji in frequency, type, and sentiment. The diversity and global usage of Emojis lead researchers to perform analysis of Emoji usage according to gender [10]. This study collected the data through the *Kika Keyboard*, and they rely on the usage preference of the user himself. This method may be affected in two way: (i) if the user interacts more with opposite-gender friends, his/her Emoji usage may have affected by them [29], and (ii) if the user is careful in choosing the Emojis. Our work is different in two senses. First, we skip the user Emoji usage and rely on other Facebook users Emotional response to solve the above limitations. Second, we engage the content of the picture as a powerful impact on individuals' emotional responses. Emoji can be interpreted differently according to the platform, which might influence communication [28]. Besides, some researchers have investigated the power of Emoji in the cross-lingual sentiment classification task [11] and have performed large scale empirical study on how developers used Emoji on GitHub [25].

To conclude, these approaches depend on the target user Emoji's usage. It might be straightforward to guess the Emoji publisher's gender. In contrast, we study gender inference attacks on Facebook by considering Emojis/Emoticon's preferences of other Facebook users (e.g., friends) while commenting on target-owned pictures. Although this approach is more complicated, it has two advantages over previous works: (i) target user personality does not affect the performance, (ii) the attack is still possible even when the target user is careful enough to manipulate Emoji/Emoticon neutrally.

2.3 Graph Based Method

Similar to the word and document embedding algorithms, graph embeddings are vector representations that locate users who have similar behaviors and preferences close to each other. Recent methods use neural networks to represent nodes in low-dimensional vectors. [38] introduces graph-based convolutional neural networks to infer social media users' attributes. Their method leverage visible user-profiles and social links to predict target users missing attributes. [3] describes a method for predicting the occupational class and the income of Twitter users given information extracted from their extended networks by using the graph embedding method. The authors of [34] introduce privacy inferring attacks by expressing the attacker's prior knowledge using knowledge graphs. All the above works use published attributes to generate the graph or consider the background knowledge of the attacker about the user. The main drawbacks of the above works are the difficulty of accessing such background and complete knowledge about the users. Our study investigates the effectiveness of applying these unsupervised methods used for learning node embeddings, to non-user generated data. The advantage of our method lies in the fact that the attacker does not need to have prior knowledge, which makes the attack suitable in an online mode.

3. GENDER INFERENCE ATTACK

Our gender inference attack aims at predicting a Facebook user gender from some limited amount of available information about his/her pictures. It consists of two steps. In the first step, we generate a connected graph from the picture metadata. In the second step, we apply a node2vec approach to the generated graph to obtain a user vector representation. The attack relies on constructing informative features for each user, and comparing two users' vectors, with the assumption that vector representations of male users are closer to each other than female users'. The key steps of our attack are detailed in the rest of this section.

3.1 Graph Generation

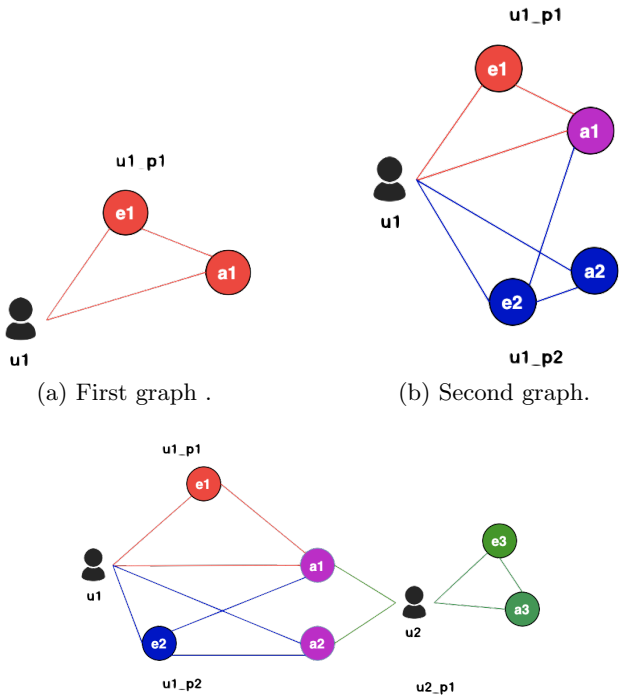
For the convenience of narration, we introduce following notations. Let $U = \{u_1, u_2, u_3, ..., u_m\}$ be the set of users, and their set of pictures defined by p where $P_i = \{p_i^1, p_i^2, .., p_i^n\}$ is the set of pictures for u_i. A picture $p_i \in P_i$, published by user u_i, is defined by $p_i = \langle a_i, e_i \rangle$ where a_i is an alt-text generated by Facebook and e_i is the set of Emojis/Emoticons posted by commenters for that picture.

In the following, we give an example of the steps to generate the graph from the available metadata. As illustrated in Figure 2, we are given a user u_1 owning two pictures $P_1 = \{p_1^1, p_1^2\}$. In the first step, we generate the graph by adding all the nodes in p_1 (tags and Emojis/Emoticons) together and to u_1 (see Figure2(a)). In the second step, for p_2, we connect all nodes appearing in p_2 (tags and Emojis/Emoticons) to u_1 and p_1 nodes if they appear together (see Figure2(b)). For another user u_2, we connect nodes of the current graph (tags, and Emojis/Emoticons) to u_2 , if u_2 pictures contain those nodes (see Figure2(c)). User posted pictures can receive different types of metadata.

To cover all possible scenarios based on the pictures metadata availability, we generate a graph for each one of the following cases: (1) Facebook generated alt-text; (2) commenters' comments; (3) both alt-text and comments. We represent each scenario graph in the next subsection and represent the accuracy result of each graph separately in Subsection 4.4.

3.2 Graph Embedding

To capture the vector embedding, we use the node2vec



(a) First graph . (b) Second graph.

(c) Third graph.

Figure 2: **Steps for generating a graph from picture metadata.**

method. Node2vec is a representational learning framework of graphs, which can generate continuous vector representations for the nodes based on$p$ the network structure. The advantage of using node2vec over feature-based approach is threefold: (i) it is semi-supervised methods that make the online inference attack possible, (ii) it automatically detects the feature which improves the feature-based lacks, and (iii) it is robust to unseen words as it is not based on texts. The objective of node2vec is to get a low-dimensional embedding that preserves the graph characteristics and represents owners that share pictures with the same style (themes, setting, background) and receive a similar reaction, as close vectors. To achieve this objective, we have to tune some parameters. To generate a representative embedding, node2vec harmonizes between immediate/distance neighbors of/from the source node by combining Breath-first Sampling (*BFS*) and Depth-first Sampling (*DFS*) strategy. To that end, a biased random walk is introduced with parameters (i) p to control the possibility of revisiting a target node during the random walk, and (ii)q to control how far the walks can move from a target node. In general, we examine four node2vec parameters which are: p,q, the number of walks to be generated per node in the graph, and the dimension, d, which is the output of the underlying word2vec model. We discuss parameters setting in Subsection 4.3.

We consider pictures as the minimum information concealed by users. We also define a set of minimum knowledge for attackers as follow:
1) Pictures metadata are available to the attacker (some or all).
2) The attacker has partial access to users' gender in the graph. The reason for that is, after getting the vector of each user and clustering base on the vectors, the attacker has to know which cluster belongs to which gender (male or female). This partial knowledge will allow thereafter the attacker to locate users with unknown gender in one of both clusters.

We generate three different graphs and vectorize users in each scenario (defined in Subsection 3.1). To test the effectiveness of the trained vector based on the available metadata, we first reduce the vector dimension to *2* by using Principal Component Analysis (*PCA*) method [20]. *PCA* is a linear technique to reduce high dimensional data to lower dimensions while helps to find the most effective transformation of existing attributes. Our user representation is shown in Figure 3, in which pink dots represent female users, while blue dots show the male users. Here we set the dimension, p, and$q$ parameters as *300*, *1*, *4*, respectively, considering the performance of the model. The comparisons of tuned parameters (p,q, dimension and number of walks) are shown in Figures 5 and 6.

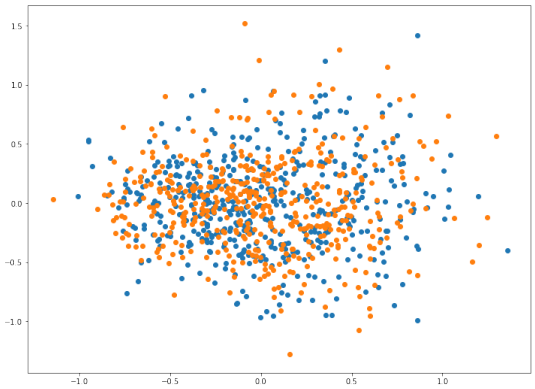
In Figure 3, we extract picture metadata separately and run them through the node2vec algorithm to get users' vectors of predefined dimensions, and then use cosine similarity score [27] to compute similarities among users for clustering. Figure 3(a) is the result of generating vector representation for female and male users by using only Facebook generated alt-texts.

In Figure 3(b), we use commenters' Emojis/Emoticons preferences in reacting to males and females-owned pictures to generate the vector representation. In Figure 3 (c), we give user representation based on both alt-text and reaction mode. To interpret the graphs, as we feed a rich dataset to the model, we obtain closeness between genders. For example, in Figure 3 (a) users are distributed all over the space, while in Figure 3 (b) users are getting far away from each other but they are not still really close to each other in the clusters. Unlike the other two graphs, in Figure 3 (c) female and male users are placed closer to each other in the vector space, which implies the power of emojis and alt-text combination in gender inference attack. Although Figure 3 (c) has overlap, the users (female, or male) are close to each other and represent as a stream in the space. Therefore, we conclude that the similarity of vectors among users in one gender is much larger than users across genders.

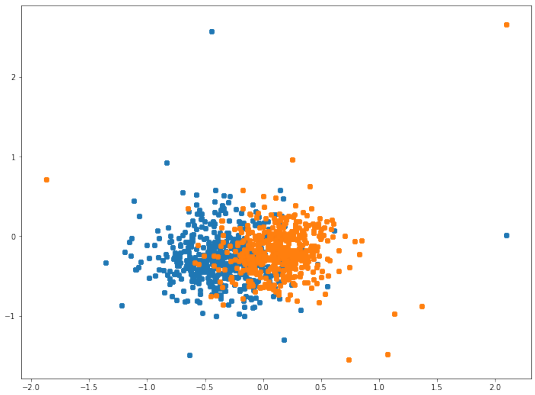
To conclude, we note that our user embedding is useful to cluster female and male users even with a limited amount of available metadata. We also notice the female and male clusters get separable as valuable metadata are accessible, which implies that the best scenario for the attacker is the last one when he has access to all metadata.

4. ATTACK EVALUATION

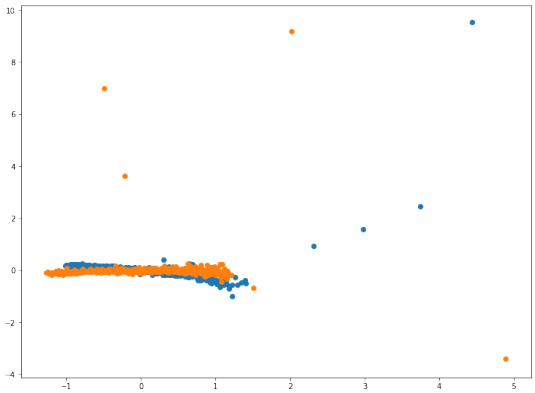
We evaluate our proposed gender inference attack in this section. We first describe the dataset, evaluation metric, and parameter setting. Then, we experimentally study the sensitivity of the parameters involved in our inference attack. Finally, we assess the performance of our attack when the



(a) Generated graph from pictures alt-text.

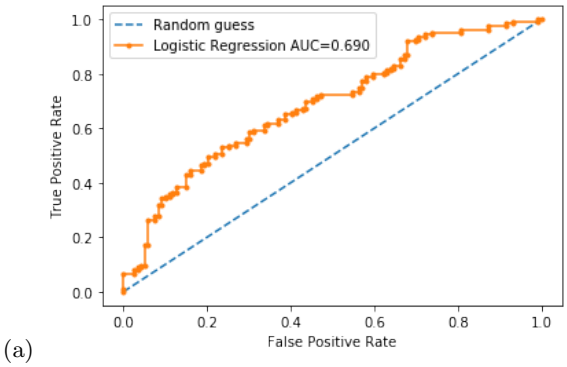


(b) Generated graph from received emojis.

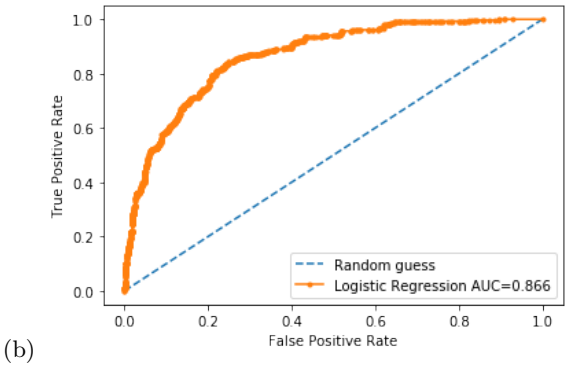


(c) Generated graph from both pictures alt-text and received emojis.

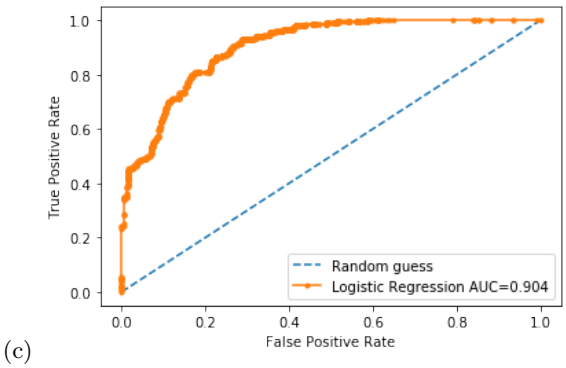
Figure 3: **User vector representation. Pink dots shows the female-owned pictures and blue dots illustrate male-owned pictures.**



(a)



(b)



(c)

Figure 4: **Result: (a) AUC alt-text, (b) AUC emojis, (c) AUC alt-text and emojis.**

attacker has partial or full access to picture metadata (alt-text, and comments).

4.1 Dataset

To collect the ground truth, we utilize a python crawler to collect all the required information from the HTML file of each picture. The data collection was conducted from April to June *2019*. We have collected *141,812* pictures and their *446,655* messages. Our statistics showed that *1291* different types of Emoji appear in our dataset. We randomly select Facebook users to avoid a region or country usage bias. We use female and male labels. To get a representative yet usable dataset, we perform some pre-processing on the collected data. First, since the generated alt-text contains conjunction, redundant tags, and appended text, we reformulate the generated alt-text to create a proper graph. Second, we replace the Emoticons with their corresponding

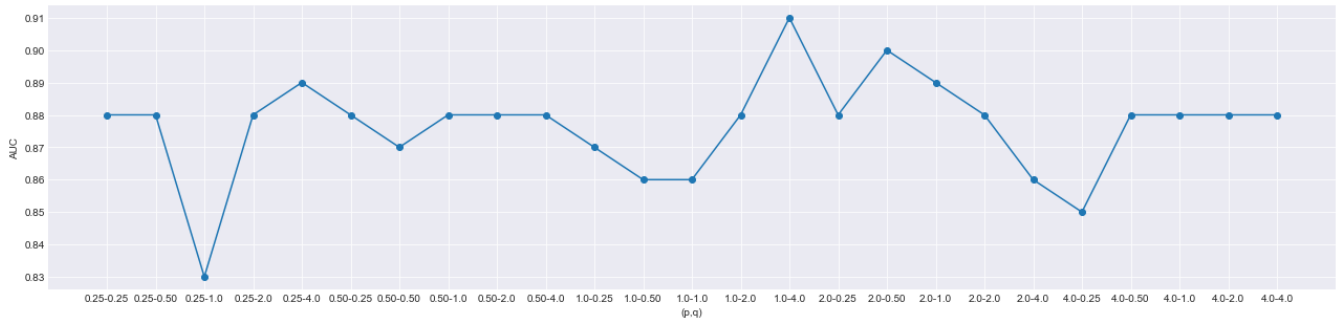


Figure 5: **The effect of different (p, q) on the performance of LR.**

Emojis. To tackle the problem of an imbalanced dataset, we adopt the down-sampling strategy used in [18]; that is, we randomly sample the same number from both classes.

4.2 Metric

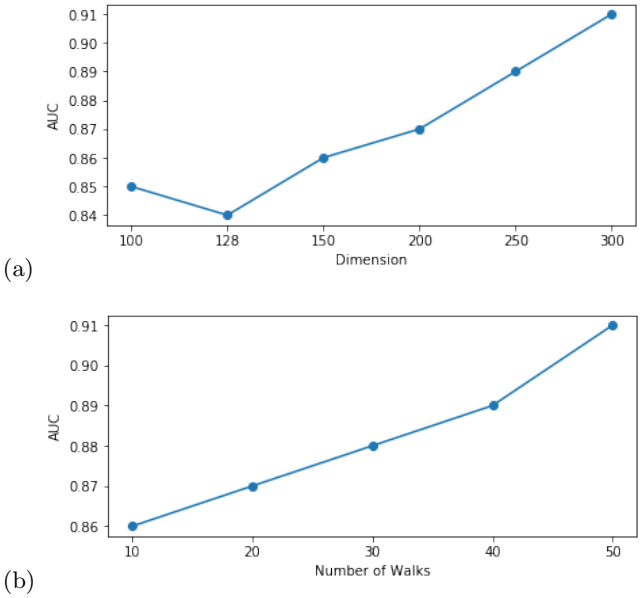
To evaluate the performance of each scenario, we utilize AUC (area under the ROC curve) as the evaluation metric for two reasons. First, it is not sensitive to the label distribution [23]. Second, It is easy to interpret the attack's performance as the AUC range is $[0.5, 1]$ where 0.5 is equal to random prediction, and 1 is an accurate prediction.

4.3 Parameters

We examine how Node2vec parameters affect its performance on our dataset. We learn the best values of p and$q$ parameters using 10% labeled data with a grid search over $p, q \in \{0.25, 0.50, 1, 2, 4\}$. Except for these two parameters, all the other remaining parameters are set to their default values. The result of all the possible combinations of p and$q$ parameters is given in Figure 5. We examine how the dimensions d, and the number of walks num_walk affect the performance. For these parameters testing, except p and$q$ which we find in Figure 5, the remaining ones are set to their default settings. Figure 6(a) shows the effect of different dimensions on AUC performance. We observe the performance increases as the value of d gets larger. Similarly, we perceive an increase in the number of walks improves performance, since we have a larger sample to learn the representations. The number of walk values connects to the amount of data being fed into the skip-gram model. Finally, we select $d = 300$, $num\_walk = 50$, p= 1 and q=4 in our inference attack which gives the best result. Note, all the parameters are fixed in the last scenario, where we received the best result.

4.4 Results

To address the problem of fairly estimating the performance of the classifier, and make sure the classifiers can generalize to unseen data, we divide the dataset into three parts, train (70%), validation (10%), and test (20%). We train the model on the training dataset and search for the sufficient parameters to our attack in the validation dataset. Finally, we use the test dataset to represent the result of the model. We consider the gender inference attack as a binary classification problem, and Logistic Regression (LR) is used as the classifier to evaluate the user embedding. LR is adapted to infer between female and male users and it takes a linear equation as input and uses a logistic function and log odds to perform a binary classification task. $L2$ regularization [14] is preferred for improving the performance of the model.



(a)

(b)

Figure 6: **The effect of different parameters selection on performance of LR: (a) Dimension (b) Number of walks.**

Figure 4 shows the performance of LR classifiers in each scenario. A random guess classifier is one that cannot discriminate between the classes and would predict a random class in all cases. Figure 4(a), (b), and (c) shows the result of classifier after using only alt-text, only emojis, and both alt-text and emojis as the input data to generate the graph, respectively. According to the result, LR can infer the picture owner gender with AUC of 0.69, 0.86, and 0.90 based on the availability of the picture metadata, which represents a good prediction result.

To conclude, non-user generated data (Emoji preferences of other Facebook users and generated alt-text) are adequate and easily accessible data to the attacker that can be used to infer the picture owner's gender. The experimental re-

sult shows that the prediction accuracy of the last scenario is higher than the other two scenarios. Therefore, the performance of the classifier gets better as the input data get richer, which matches the representation of users in Figures 3 (c).

5. CONCLUSION

Identifying users' gender from their online activities and data sharing behavior is an important topic in the growing research field of social networks. This study has investigated *141,812* images and their *446,655* comments. We presented a new perspective of gender inference attack by embedding users based on their picture sharing style and comments that they received from users (non-user generated data). We studied feature learning by relying on a sophisticated method, called node2vec. This perspective gave us multiple advantages. First, The attack is possible even though all the data except pictures metadata are hidden. Second, it simplifies the feature selection process which is a tedious task. Third, it can be implemented in an unsupervised fashion, that alleviates the problem of collecting labeled dataset. Lastly, it can scale to large networks. We showed the possibility of gender inference attack even when all user attributes/activities are hidden, such as profile attributes, friend list, liked pages, and joined groups. The attacker can infer the picture owner's gender more accurately if he has access to complete picture metadata. Our experiment results showed that picture metadata are sufficient to infer the owner's gender. As future work, we plan to (i) to consider other attributes, and (ii) adjusting the same technique to propose counter-measures to picture owners.

6. ACKNOWLEDGMENTS

This research work is supported by the DIGITRUST `http://lue.univ-lorraine.fr/fr/article/digitrust`.

7. REFERENCES

[1] C. Abdelberi, G. Ács, and M. A. Kâafar. You are what you like! information leakage through users' interests. In *19th Annual Network and Distributed System Security Symposium, NDSS*, San Diego, California, USA, 2012. The Internet Society.

[2] W. Ai, X. Lu, X. Liu, N. Wang, G. Huang, and Q. Mei. Untangling emoji popularity through semantic embeddings. In *Proceedings of the Eleventh International Conference on Web and Social Media, ICWSM*, pages 2–11, Montréal, Québec, Canada, 2017. AAAI Press.

[3] N. Aletras and B. P. Chamberlain. Predicting twitter user socioeconomic attributes with network and language information. In *Proceedings of the 29th on Hypertext and Social Media*, pages 20–24. 2018.

[4] B. Alipour, A. Imine, and M. Rusinowitch. Gender inference for facebook picture owners. In *International Conference on Trust, Privacy and Security in Digital Business, TrustBus*, pages 145–160, Linz, Austria, 2019. Springer.

[5] J. An, T. Li, Y. Teng, and P. Zhang. Factors influencing emoji usage in smartphone mediated communications. In *Transforming Digital Worlds - 13th International Conference, iConference*, pages 423–428, Sheffield, UK, 2018. Springer.

[6] F. Barbieri, G. Kruszewski, F. Ronzano, and H. Saggion. How cosmopolitan are emojis?: Exploring emojis usage and meaning over different languages with distributional semantics. In *Proceedings of the Conference on Multimedia Conference, MM*, pages 531–535, Amsterdam, The Netherlands, 2016. ACM.

[7] T. Belinic. Personality profile of social media users how to get maximum from it. `https://medium.com/krakensystems-blog/personality-profile-of-social-media-users-how-to-get-maximum-from-it-5e8b803efb30`, Apr. 2009.

[8] R. L. Brouer, M. Stefanone, R. L. Badawy, and M. J. Egnoto. Gender (in)consistent communication via social media and hireability: An exploratory study. In *50th Hawaii International Conference on System Sciences, HICSS*, pages 1–10, Hawaii, USA, 2017. ScholarSpace / AISeL.

[9] S. E. Butterworth, T. A. Giuliano, J. White, L. Cantu, and K. C. Fraser. Sender gender influences emoji interpretation in text messages. *Frontiers in psychology*, 10:784, 2019.

[10] Z. Chen, X. Lu, W. Ai, H. Li, Q. Mei, and X. Liu. Through a gender lens: Learning usage patterns of emojis from large-scale android users. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW*, pages 763–772, Lyon, France, 2018. ACM.

[11] Z. Chen, S. Shen, Z. Hu, X. Lu, Q. Mei, and X. Liu. Emoji-powered representation learning for cross-lingual sentiment classification. In *The World Wide Web Conference, WWW*, pages 251–262, San Francisco, CA, USA, 2019. ACM.

[12] M. Cheung and J. She. An analytic system for user gender identification through user shared images. *TOMCCAP*, 13(3):30, 2017.

[13] M. D. Choudhury, S. S. Sharma, T. Logar, W. Eekhout, and R. C. Nielsen. Gender and cross-cultural differences in social media disclosures of mental illness. In *Proceedings of the Conference on Computer Supported Cooperative Work and Social Computing, CSCW*, pages 353–369, Portland, OR, USA, 2017. ACM.

[14] B. Conroy and P. Sajda. Fast, exact model selection and permutation testing for l2-regularized logistic regression. In *Artificial Intelligence and Statistics*, pages 246–254, 2012.

[15] R. Farahbakhsh, X. Han, Á. Cuevas, and N. Crespi. Analysis of publicly disclosed information in facebook profiles. *CoRR*, abs/1705.00515, 2017.

[16] L. Flekova, J. Carpenter, S. Giorgi, L. H. Ungar, and D. Preotiuc-Pietro. Analyzing biases in human perception of user age and gender from text. In *Proceedings of the 54th Annual Meeting of the Association for Computational*, pages 843–854, Berlin, Germany, 2016. ACL.

[17] N. Z. Gong and B. Liu. You are who you know and how you behave: Attribute inference attacks via users'

social friends and behaviors. In *25th Security Symposium*, pages 979–995, Austin, TX, USA, 2016. USENIX.

[18] A. Grover and J. Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864, 2016.

[19] J. Jia, B. Wang, L. Zhang, and N. Z. Gong. Attriinfer: Inferring user attributes in online social networks using markov random fields. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1561–1569, 2017.

[20] I. T. Jolliffe. Principal component analysis. In *International Encyclopedia of Statistical Science*, pages 1094–1096. Springer, 2011.

[21] F. Karimi, C. Wagner, F. Lemmerich, M. Jadidi, and M. Strohmaier. Inferring gender from names on the web: A comparative evaluation of gender detection methods. In *Proceedings of the 25th International Conference on World Wide Web, WWW*, pages 53–54, Montreal, Canada, 2016. ACM.

[22] A. Lenton. Facebook wants you to search photos of your female friends at the beach, but not your male mates. `https://www.whimn.com.au/talk/people/facebook-wants-you-to-search-photos-of-your-female-friends-at-the-beach-but-not-your-male-mates/news-story/bbc21ee6883bd07bfbbbe76a0c8ca54c`, Feb. 2019.

[23] R. N. Lichtenwalter, J. T. Lussier, and N. V. Chawla. New perspectives and methods in link prediction. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 243–252, 2010.

[24] X. Lu, W. Ai, X. Liu, Q. Li, N. Wang, G. Huang, and Q. Mei. Learning from the ubiquitous language: an empirical analysis of emoji usage of smartphone users. In *Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp*, pages 770–780, Heidelberg, Germany, 2016. ACM.

[25] X. Lu, Y. Cao, Z. Chen, and X. Liu. A first look at emoji usage on github: An empirical study. *CoRR*, abs/1812.04863, 2018.

[26] P. S. Ludu. Inferring gender of a twitter user using celebrities it follows. *CoRR*, abs/1405.6667, 2014.

[27] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[28] H. J. Miller, D. Kluver, J. Thebault-Spieker, L. G. Terveen, and B. J. Hecht. Understanding emoji ambiguity in context: The role of text in emoji-related miscommunication. In *Proceedings of the Eleventh International Conference on Web and Social Media, ICWSM*, pages 152–161, Montréal, Québec, Canada, 2017. AAAI Press.

[29] D. Nguyen, D. Trieschnigg, A. S. Dogruöz, R. Gravel, M. Theune, T. Meder, and F. de Jong. Why gender and age prediction from tweets is hard: Lessons from a crowdsourcing experiment. In *25th International Conference on Computational Linguistics, Proceedings of the Conference, COLING*, pages 1950–1961, Dublin, Ireland, 2014. ACL.

[30] C. Peersman, W. Daelemans, and L. V. Vaerenbergh. Predicting age and gender in online social networks. In *Proceedings of the 3rd International Workshop on Search and Mining User-Generated Contents, SMUC*, pages 37–44, Glasgow, United Kingdom, 2011. ACM.

[31] B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710, 2014.

[32] B. A. Pijani, A. Imine, and M. Rusinowitch. You are what emojis say about your pictures: language-independent gender inference attack on facebook. In C. Hung, T. Cerný, D. Shin, and A. Bechini, editors, *SAC '20: The 35th ACM/SIGAPP Symposium on Applied Computing, online event, [Brno, Czech Republic], March 30 - April 3, 2020*, pages 1826–1834. ACM, 2020.

[33] D. Preotiuc-Pietro, W. Xu, and L. H. Ungar. Discovering user attribute stylistic differences via paraphrasing. In *Proceedings of the Thirtieth Conference on Artificial Intelligence,*, pages 3030–3037, Phoenix, Arizona, USA, 2016. AAAI Press.

[34] J. Qian, X.-Y. Li, C. Zhang, and L. Chen. De-anonymizing social networks and inferring private attributes using knowledge graphs. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9. IEEE, 2016.

[35] L. Santamaría and H. Mihaljevic. Comparison and benchmark of name-to-gender inference services. *PeerJ Computer Science*, 2018.

[36] M. Sap, G. J. Park, J. C. Eichstaedt, M. L. Kern, D. Stillwell, M. Kosinski, L. H. Ungar, and H. A. Schwartz. Developing age and gender predictive lexica over social media. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP*, pages 1146–1151, Doha, Qatar, 2014. ACL.

[37] Shutterstock. The psychology behind why we share on social media. `https://www.shutterstock.com/blog/the-psychology-behind-why-we-share-on-social-media`, Jan. 2015.

[38] Y. Tian, Y. Niu, J. Yan, and F. Tian. Inferring private attributes based on graph convolutional neural network in social networks. In *2019 International Conference on Networking and Network Applications (NaNA)*, pages 186–190. IEEE, 2019.

[39] S. Wu, J. Wieland, O. Farivar, and J. Schiller. Automatic alt-text: Computer-generated image descriptions for blind users on a social network service. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW*, pages 1180–1192, Portland, OR, USA, 2017. ACM.