A Light in the Dark Web: Linking Dark Web Aliases to Real Internet Identities

Ehsan Arabnezhad, Massimo La Morgia, Alessandro Mei, Eugenio Nerio Nemmi, Julinda Stefa
Department of Computer Science, University of Sapienza, Rome

Email: arabnezhadlotfabad.1734292@studenti.uniroma1.it, {lamorgia, mei, nemmi, stefa}@di.uniroma1.it

***Abstract*—Most users have several Internet names. On Facebook or LinkedIn, for example, people usually appear with the real one. On other standard websites, like forums, people often use aliases to protect their real identities with respect to the other users, with no real privacy against the web site and the authorities. Aliases in the Dark Web are different: users expect strong identity protection.**

In this paper, we show that using both "open" aliases (aliases used in the standard Web) and Dark Web aliases can be dangerous per se. Indeed, we develop tools to link Dark Web to open aliases. For the first time, we perform a massive scale experiment on real scenarios. First between two Dark Web forums, then between the Dark Web forums and the standard forums. Due to a large number of possible pairs, we first reduce the search space cutting down the number of potential matches to a small set of candidates, and then on the selection of the correct alias among these candidates. We show that our methodology has excellent precision, from 87% to 94%, and recall around 80%.

I. Introduction

People typically use plenty of social websites on the Internet. In some of them, they often disclose the real identity—users like that friends and colleagues can find them easily. On these sites, people share photos and experiences, and privacy with friends is not a real concern. The same users, when on forums, discussion boards, or live news aggregators, share less information and appreciate the weak anonymity they can get by using an alias. They do not use anonymity technology like Tor, they know they are not anonymous at all to the website and the authorities, and that's fine. It is different when using an alias in the Dark Web. Privacy is often the primary concern and being anonymous a necessity. One of the contributions of this work is to show that having multiple aliases on both the standard Internet and the Dark Web—a widespread circumstance in real life—is dangerous.

Plenty of people use the Dark Web. To give a better picture, consider that the amount of bitcoins traded in the Dark Web was estimated at 600 million USD in 2017 [1]. Sometimes the Dark Web is used to perform activities that are illegal, or questionable. Sometimes to discuss freely in forums about politics and culture in countries where there is no free speech. Many of the users of the Dark Web have several aliases in the Dark Web itself, *dark aliases*, and several aliases in the standard Web without the use of secure anonymity systems like Tor, *open aliases*.

First, we show how to break pseudo-anonymity between dark aliases. We can link dark aliases by using two methodologies: Daily activity profiles and stylometry. Daily activity profiles fingerprint users in terms of their posting activity during the day. Stylometry fingerprints users according to their very own style of writing—words, punctuation, and writing idiosyncrasies. When pseudo-anonymity between dark aliases is broken, you can know that two different aliases are just the same person, but you don't know who since both aliases live in the Dark Web.

Second, we show how to break anonymity. Note that this is more challenging: People might behave differently and use different writing styles when in the standard Web. We describe ways to mitigate this problem and show excellent performance—precision from 87% to 94% and recall around 80%. Once a dark alias is linked to an open alias, there is no more anonymity. Open aliases can be tracked by the authorities and easily linked to real identities.

The biggest problem we had to face in this work is how to assess our methodology. Indeed, there is no easy-to-get ground truth. We have overcome this challenge by human inspection. We have linked aliases of the same user in all those cases, checked manually, in which the user himself in one message declares that he is the owner of the two aliases. We have discarded the message from the dataset and used the aliases as ground truth to validate our methodologies and check performance.

Our main contributions are:

- **Methodology:** Authorship Attribution is a well-known field of study, but it is still an open problem when we scale from a few hundred candidate authors to thousands of them. We propose a methodology and a combination of stylometric and temporal features that achieve excellent results when compared with previous attempts of large scale authorship attribution.
- **Dark Web scenario:** To the best of our knowledge, this work represents a first attempt to link Dark Web aliases to standard Web ones. This task requires to handle several technical challenges such as finding websites on the standard web that are a meeting point of Dark Web users too, gathering the data, and overcome the problem of the scarcity and quality of the exploitable text. Indeed, in other works [2], [3] on large scale authorship attribution the authors retrieve the text of the users from blog posts—a unique, or few, single context long message. In our case,

we have forum posts—multiple, disjoint, short messages with slang and acronyms abuse.
- **De-anonymization:** We found 47 matching aliases between the standard Web and the Dark Web, and for at least 20 of them, we were able to confirm they belong to the same person. Moreover, we notice that Dark Web users behave care-less about their privacy in the standard Web, here they reveal habits, interests, and personal information that can be exploited to outline a fine grain personal profile of the user. We present a focus on John Doe, a Dark Web user for whom we were able to discover his age, the city where he lives, and even the model of his smartphone. This analysis shows the effectiveness of linking the Dark Web aliases to the open aliases.

In addition, we believe that our work contributes to the understanding of the Dark Web community. In the case of illicit activities, our methodology can support the authorities to drastically reduce the set of users under investigation and to collect some initial information on possible criminals. Finally, this work wants to make aware Tor users about the privacy risks of posting on the standard Internet.

II. Background

A. Tor and the Dark Web

Tor [4] is the most popular tool to get anonymity on the Internet. Users of Tor can surf the Web with excellent anonymity guarantees. Also, can websites hide behind the Tor system: in this case, the site is a so-called *hidden service*. The Dark Web is the Web of hidden services.

In the Dark Web, neither the user nor the server knows the IP address of the other since the connection happens in a rendezvous point that links the two entities using Tor. The Dark Web hosts several forums on the most diverse topics. Some of them are illegal, like drug markets, hacking activities, or child pornography. Others are just places where people meet to talk about sensitive topics; for example, in countries where free speech is limited. In all cases, the users of the Dark Web use aliases, and the protection of the real identity behind these aliases is critical.

B. Stylometry

Stylometry is a technique for analyzing texts and collecting evidence of authenticity, authorship, and much more. It was very used in the 19th century to determine the author of unknown or collaborative playwrights and in forensics analysis. The development of computers with their power to analyze massive amounts of data enhanced the possibility of these techniques. With the birth of the Web and the rise of illegal activities on the Web, being able to identify users starting from their writing style and habits has become a vital issue. A lot of research has been done to apply stylometry to social networks and forums, but successful results are still far.

Two main fields of research focus on this task: Authorship Attribution and Authorship verification. Authorship Attribution (AA) is the task of identifying the correct author of an unknown text for which we have a set of possible candidates. Authorship Verification (AV), instead, is the task of finding if the author is one of the candidates and, if it is, determine who among them. Although the two problems seem similar, a profound difference marks them as a more accessible and a harder challenge. In AA, we have a closed set, meaning that we are sure that one of the candidates is the author. In AV, we don't know that. The author of the text could be none of the candidates.

III. Datasets

First, we give an overview of the forums under investigation, the data we collected, and the procedure we used to polish the datasets.

A. Reddit

Reddit [5] is an American social news aggregation and discussion website in position 6th in the US and 20th worldwide in the most visited websites ranking made by Alexa [6]. It has more than 330 million users, called redditors. Reddit consists of "subreddits", sub-forums that people can create to discuss specific topics. As for today, there are more than 138,000 active subreddits. One of the reasons for its success is that every redditor can open, organize, and manage a new subreddit without explicit approval from Reddit.

This policy has led to the rise of a lot of controversial subreddits. Among them, you can find DarkNetMarkets. In DarkNetMarkets, you can talk about drugs, deals, vendors, shipping methods, and experiences as they do on the Dark Web. In its own way, this is useful to reduce the risk of fraud, a big problem in the community. This subreddit was very active, with more than 180,000 redditors. Then, after a policy change in March 21 2018 [7], it was shut down together with other similar subreddits. However, a few days later, the old subreddits were replaced with new ones with, fundamentally, the same activities.

The idea is that some of the people of the Dark Web can be users of these subreddits too. The challenge is to find who. We start getting the topics—from the most upvoted to the least. Usually, the more upvoted topics are also those with more comments. For each of the first 1000 topics, we collected all the users that wrote a comment. Afterward, for each user, we collected the last 1000 messages across all the subreddits. Using this procedure, we were able to collect 16,567 users and their messages. The messages are from 33,000 different subreddits. To reduce the number of topics, we discarded all the subreddits with less than 10 messages from our users. This way, the number of subreddits dropped to 656. Then, we defined 12 topics and manually labelled each subreddit with one of these topics. The result is shown in Table I.

Since we are talking about Dark Web users, it is no surprise that the most popular topic is drugs. More surprising is that some of the users are addicted video-gamers and that others are interested in cultural, political, and financial topics. We can deduce that these users are relatively young, with a notable level of education.

TABLE I
REDDIT DATASET COMPOSITION BY TOPIC.

Topic	subreddit(#)	subscriptions(%)	messages(%)	popular subreddit	messages(#)
Culture	18	4.7%	2%	r/science	17,442
Cryptocurrencies	39	3.2%	6%	r/bitcoin	96,407
Drugs	117	15.6%	33.7%	r/DarkNetMarkets	670,483
Entertainment	166	39.1%	22.4%	r/pics	75,454
Financial	15	1.6%	0.9%	r/personalfinance	11,590
Lifestyle/Sports	72	9.9%	9.5%	r/LifeProTips	12,109
News	18	4.8%	4.5%	r/worldnews	89,189
Places	43	1.4%	3%	r/canada	11,291
Politics	24	4%	5.9%	r/politics	119,238
R18+	12	1.6%	4.5%	r/sex	10,676
Psychological help	11	1.7%	0.5%	r/GetMotivated	3,733
Tech/Tor	52	5.4%	3.6%	r/technology	26,919
Videogame	61	7.0%	7.3%	r/gaming	41,183

B. Dark Web forums

To build our dataset, we looked into the Dark Web for hidden services with the highest number of users. We selected The Majestic Garden and the Dream Market forum—drug markets with almost all the sections accessible by new users. Of course, these sites do not have open APIs; we had to scrape the content of the forums to collect the data set.

1) The Dream Market: The Dream Market is one of the most popular marketplaces in the Dark Web since 2013. On the Dream Market, vendors can sell any drugs, counterfeit stuff, and stolen data in exchange for bitcoins, bitcoin cash, or monero. The marketplace is pretty well organized: They provide cryptocurrency wallets to the customers and escrow service. From March 27, 2019, the home page of the marketplace informs the customers that the service is shutting down, and moving to another hidden service, currently offline.

The Dream Market consists of a hidden service dedicated to the market place and a hidden service that serves as the official discussion forum [8]. The forum had 4 main sections: Products and Vendor Reviews, Marketplace discussions, Advertising and Promotions, and Scams. In the last one, the users could report cheats or potential scams to the administrators. Although on the Dream Market customers could buy several kinds of stuff, in the forum, most of the messages were about drugs. From this hidden service, we collected 6,348 users and their messages.

2) The Majestic Garden: The Majestic Garden [9] arose from the ashes of Silk Road. After the seizure of Silk Road, a group of former users, passionate about psychedelic drugs, founded the forum called The Majestic Garden [10], [11].

The forum is born as a place where people can share experiences related to the assumption of psychedelic drugs. Different from the Dream Market, The Majestic Garden forum does not have an official marketplace. Here, the qualified vendors directly sell goods to the members without any intermediation. Thus, the forum neither manages the money of the users directly, nor it gains from the transactions. It is just a meeting point. Since there is no marketplace, each vendor has its thread. Usually, the first post of the thread is the vendor showcase, while subsequent messages are the reviews of the users about the quality of the goods and the vendor. Most discussions are related to the assumption of drugs, but there are also sections dedicated to the literature of psychedelic and spiritual experiences, and how-tos on drug cooking. Because of the nature of this forum, the messages are longer than average and more digressive. We globally collected from The Majestic Garden 4,709 different aliases.

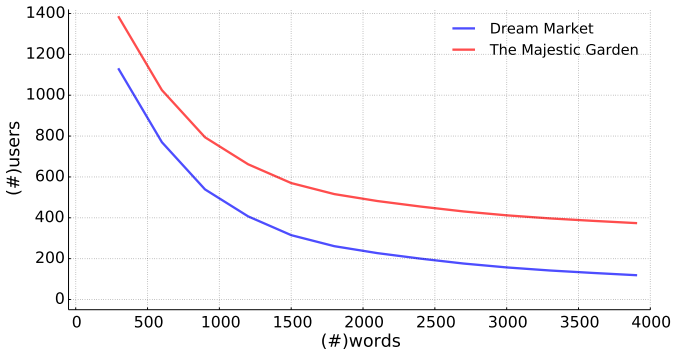


Fig. 1. Cumulative distribution of the number of words per user on Dark Web forums.

C. Polishing the data

Users on the web use slang and common saying more than people in other contexts. This results in dirty data that need particular attention to be polished and prepared for our purposes.

We perform the following steps:

1) We delete all the accounts whose nickname end or start with ”bot” because we found that most of them are not human beings. We notice that while manually inspecting the data. Especially on Reddit, Bots tend to have the ”bot” substring in the nickname to better recognize them.
2) We remove all the duplicates of the messages. This is especially important in the dark web, where users that are vendors often post the same message after a short

period of time or on Reddit where users post the same message in different subreddits (i.e. crossposting)
3) We normalize urls, keeping only the hostname (i.e. *www.reddit.com* is transformed into *reddit*.
4) We remove emojis from the messages.
5) We get rid of messages shorter than 10 words, they are usually not meaningful. We notice that most of them are just an expression of agreement or disagreement concerning other users statements.
6) We discard messages with less than 0.5 as the ratio between the number of different words and the total number of words. A lot of spam messages are made of a single sentence written multiple times. We empirically set this threshold after test on different spam and normal messages.
7) We select only the messages written in English to build the text, while we keep track of all the information associated like timestamps and subreddits.
8) We remove quotes from messages, and keep only the part written by the writer. On Reddit, quote have a pattern that can be easily detected through a regular expression. This is because we do not want to evaluate texts written by a different user with respect to the owner of the account.
9) When a user edits a posted message, the software of the platform add a string 'Edit by username'. Since there is the username in the string, this part of message can alter our results, so we remove it.
10) We substitute each mail address with the tag '_mail_'.
11) We delete PGP Keys from messages. In particular we notice that in the dark web forums the PGP key is preceded by a pattern of text that introduce the PGP key.
12) We get rid of words that are longer than 34 characters. We can safely assume that they are not meaningful words, such as jokes, ascii art, or PGP keys that do not follow the aforementioned pattern.

To extract the language of the messages in the above process, we use the Python library langdetect [12]. It is a Python porting of the Google Java library language detection [13]. The library generates language profiles from Wikipedia and achieves a precision of over 99% for 55 languages. In Fig. 1 is shown the cumulative distribution of words per users on both Dark Web forums.

IV. Methodology

The two primary goals of this work are to break pseudo-anonymity in the Dark Web: To link two Dark Web aliases across two different Dark Web forums. Also, to break anonymity: To link the Dark Web aliases to open aliases of the standard Internet. To reach our goal, we exploit writing style and behavioral patterns. These problems are challenging, especially in the standard Internet, where the number of potential candidates is massive. So, we need to reduce the search space to lower the computational cost without losing the correct associations. In the following sections, we describe our approach.

A. Text pre-processing and text features

In the data collection phase, we take care of selecting meaningful messages leaving the text format as it is. Now, we need to polish and normalize the text to make it fit for the processing phase. The first step of our pre-processing phase is to tokenize the text. Tokenization is the process of breaking up a stream of text into linguistic units such as words, punctuation, or other meaningful elements. This step is pretty standard in text analysis, and it is essential in our case. Web writers usually abuse emojis, do not pay attention to the text formatting rules and grammar conventions.

After the tokenization step, we transform each token to its base form. This operation is called lemmatization and it reduces an inflected word to its lemmas (e.g., *am, are, is* $\rightarrow$ *be*). Thanks to this standardization, we can analyze words with different inflections as a single item. Finally, we extract the following features from the lemmatized text:

- **Word n-grams:** A sequence of N contiguous words in a text. For our detection, we use n-grams of length 1, 2 and 3 (Word1-3 in Table II)
- **Character n-grams:** A sub-sequence of N contiguous characters of a larger sequence. We use character n-grams of length from 1 up to 5 (Char1-5 in Table II).
- **Frequency based:** Frequency of punctuation, numbers and special characters (Freq of in Table II).

After the extraction, we order the n-grams by their frequency across the dataset, we select the top N features, and we compute their weight with the Tf-Idf. The Tf-Idf, Term Frequency-Inverse Document Frequency, is a measure associated with each term. Tf-Idf grows with the number of times the term appears in a document and lowers with the number of documents in the dataset that contain the term. This measure gives more importance to features that are frequently used by only one user and less importance to popular features such as stop-words. Table II reports on the number of features.

B. The daily activity profile

Dark Web users of forums do their best to not leave traces. What we usually know about them is just what they write and the timestamps of the posts recorded by the hidden service. Starting from the timestamps, we can build a profile based on the frequency of the posts of the user during the hours of the day. We use the same approach described in [14]. Similarly, we set the minimum number of timestamps necessary to build the profile to 30 without considering the weekend and the holidays, since in these days users typically change their habits. So, for each user u, we compute the profile of activity during the day. More formally, let bit $a_u(d,h)$ indicate whether user u has posted in the h^{th} hour of day d. The profile P_u is then defined as follows:

$$P_u[h] = \frac{\sum_d a_u(d,h)}{\sum_{d,h'} a_u(d,h')}, h \in \{0,\ldots,23\} \qquad (1)$$

TABLE II
FEATURES USED FOR SPACE REDUCTION AND FINAL CLASSIFICATION.

Type	Space Reduction	Final
Word n-grams 1-3	60,000	50,000
Char n-grams 1-5	30,000	15,000
Freq. of punctuation ('.', ',', ':')	11	11
Freq. of digit ('0', '1', '2')	10	10
Freq. of special chars ('@', '#')	21	21
Daily activity profile	24	24

Furthermore, since each forum reports a time aligned on a different time-zone, we align the timestamps by adjusting all the profiles to UTC.

As a first measure to evaluate if two profiles, computed on different forums, belong to the same user, we use the similarity between the daily activity of the two aliases. In our Reddit dataset, we have more than 10,000 possible candidates for each Dark Web user. With this large number of candidates, as also stated in [2], it is neither practical to learn a single classifier for 10,000 classes, nor to learn 10,000 one-versus-all binary classifiers. Moreover, previous works [15], [16] show that similarity or distance methods perform better than machine learning approaches when there are so many candidates.

In our approach, we measure the similarity between aliases pairs with the cosine similarity. The cosine similarity (2) is a measure of similarity between two non-zero vectors. X and Y are vectors of attributes and X_i and Y_i are the components of X and Y. In our case, X and Y are the term frequency vector and the daily activity profile of the two users. Since all the values of X and Y are in the positive space, the cosine value range from 0 to 1.

$$\cos(\mathbf{X}, \mathbf{Y}) = \frac{\mathbf{X}\mathbf{Y}}{\|\mathbf{X}\|\|\mathbf{Y}\|} = \frac{\sum_{i=1}^{n} \mathbf{X}_i \mathbf{Y}_i}{\sqrt{\sum_{i=1}^{n} (\mathbf{X}_i)^2}\sqrt{\sum_{i=1}^{n} (\mathbf{Y}_i)^2}} \tag{2}$$

The higher the cosine similarity, the higher the probability that two aliases belong to the same user.

C. Search space reduction through k-attribution

K-attribution is a relaxed version of the Authorship Attribution problem. While in the Authorship attribution problem, given a set S of documents and a document d, we look for the author in S with the maximum likelihood to be the author of d, in the k-attribution problem we look for a set of dimension k, with $k < S$, of possible authors that wrote d with maximal likelihood.

Our goal is, given an alias a, to first solve the k-attribution problem to reduce the search space from the number of aliases in the set S to a set of k candidates. To achieve this, we keep in mind two key factors. First, we want k to be small but still large enough to capture the real alias. Second, many users of the Dark Web forums have a limited amount of text that we can exploit, and we want to evaluate as many users as we can. Thus, we chose to fix the value of k at 10 and evaluate the number of words needed to balance these two key factors.

TABLE III
k-ATTRIBUTION ACCURACY AT DIFFERENT NUMBER OF WORDS USED.

# of words	$K=1$ (text)	$K=1$ (all)	$K=10$ (text)	$K=10$ (all)
400	16.4%	20%	29.6%	35.5%
600	32.5%	37.8%	51.7%	58.2%
800	49.7%	55.8%	70%	75.2%
1000	64.6%	69.6%	79.7%	84.4%
1100	68.3%	73.2%	83.7%	87.6%
1200	73.7%	76%	87.2%	89.2%
1300	78.6%	82.3%	89.1%	92%
1400	81.3%	84.4%	89.7%	93.4%
1500	84.8%	87.7%	93.4%	95.5%
1600	85.3%	87.9%	94.7%	96.5%
1700	87%	90%	95.7%	97%

1) Value of k and number of words per user: The number of words per document, hence per user, is crucial for Authorship Verification [17], [18]. In our case, a further complication is that we are upper bounded by the Dark Web datasets, where users usually write fewer messages than on Reddit. This is due to the nature of these forums. While Dark Web forums are single-topic, Reddit, with more than 100,000 communities, covers a wide range of topics. We do the following experiment to understand how the length of the messages, expressed as the number of words, affects the identification accuracy in k-attribution.

For this experiment, we select 11,679 users from Reddit, that we call dataset A, consisting of the users who have enough messages to build the daily activity profile. Then, from dataset A we built the alter-ego (procedure explained in Section IV-D) for 1,000 users randomly selected. We call this last set of alter-ego dataset B. We perform the cosine similarity between A and B twice, once using only text features and a second time using both text and daily activity features. We repeat this experiment several times, increasing each time the number of words. At the end of each experiment, for each user in B, we rank the users in A by cosine similarity score. Table III shows the accuracy for $k = 1$ and $k = 10$ with only text features (text) or both features (all), for different text size. In the light of the above results, we select 1,500 words as the size of the text for each alias and $k = 10$, because of the high accuracy, 96%, that this configuration achieves. Fig 4(a) shows the performance of the accuracy for the configuration chosen at the different values of k.

D. Refining the datasets and alter-ego generation

Once we defined the configuration in terms of the number of words, we re-sample our polished datasets to get good datasets for our experiments. So, for all the three forums we discard the users that have less than 30 timestamps, those needed to build the daily activity profile, and less than 1,500 words. Since we do not have a ground-truth to test our results, we need to generate datasets made by alter-ego aliases. So, we

TABLE IV
DATASETS FINAL COMPOSITION.

Name	(#)Aliases
Reddit	11,679
AE_Reddit	10,133
TMG	422
AE_TMG	196
DM	178
AE_DM	66

select users with more than 3,000 words and more than 60 usable timestamps. To generate two users starting from one, we randomly divide the messages of each user in two sets—one for the original user, the other for the so-called alter-ego. As for the timestamps, we evenly divide them between the original user and the alter-ego in a randomized way. So, the intersection of messages between the original user and the alter-ego is the empty set, and they can be seen as two aliases of the same person.

Running our experiments we notice that some users and their alter-egos achieve an extremely high cosine score. In a manual investigation of these users we have seen that most of them are bots, others are users that write multiple times the same messages changing just some words. Thus, to get reliable datasets we get rid of these kinds of aliases. Finally, for each user, we sort the messages by length and select the messages from the longest to the shortest until we reach the limit of 1,500 words. At the end of this procedure, we built 6 datasets, 2 for each forum. One contains the original users—Reddit, TMG, DM, while in the other one the alter-egos—AE_Reddit, AE_TMG, AE_DM. Of course, for each pair of datasets, the one with the alter-egos has fewer users, since it was not possible to build an alter-ego for every one of the original users. Table IV shows the composition of the above-mentioned datasets.

E. Finding pairs through threshold

In section IV-C, we reduced the number of the most likely candidates to 10 for each alias, now we look to determine the actual author among these candidates if there is one. The idea to achieve this goal is to find a threshold above which we can find only the correct pairs with good precision.

To find a suitable threshold, we take into account the Reddit and the AE_Reddit datasets. From AE_Reddit we randomly select 1,000 users and split them into two sets of 500 users each. We call these two sets W_1 and W_2. To find the threshold, we define the dataset W_1 as the set of the unknown aliases, and the users in Reddit as the set of the known aliases. Then, we compute the 10-attribution procedure, and we get for each user in W_1 10 possible matches from Reddit. Finally, we re-compute the text features and apply a new step of cosine similarity, for each user and their 10 candidates. Note that in this second step the new features are not the same as the previous one since the set of elements is different. At this point, we use the cosine similarity scores as possible thresholds, and we calculate the precision-recall curve.

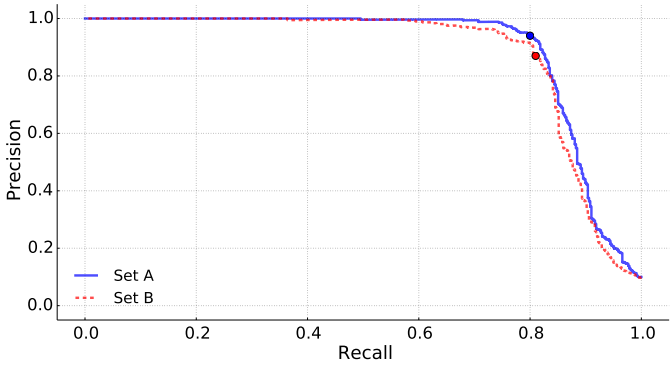


Fig. 2. Precision-recall curves for set A and B.

Analyzing the curve, we find that a good trade-off between precision and recall is where the cosine similarity score is equal to 0.4190. For this value, we get a precision of 94% and a recall of 80%. Now, we need to prove if the same threshold also holds for the other dataset. So, we repeat the procedure using the dataset W_2 as the set of the unknown authors. At the end of the process, we apply the same threshold discovered before to the new curve and this time we get a precision of 87% and a recall of 82%. Fig. 2 shows the precision-recall curves of both the sets. The blue line represents the set W_1, the red line the set W_2, while the dot marks the point where the threshold hits the curves. As we can see the two curves behave very similarly, moreover using the same threshold, the precision and the recall values are close for both of the sets.

F. Baseline comparison

To evaluate our results, we compare the proposed methodology with two baselines. For the first one, we use the characters free space 4-grams and the cosine similarity as a similarity function, since it is the standard baseline in literature for our task [2], [19], [20], [21]. We will refer to this method as *Standard Baseline*. The second one is the implementation of the methodology proposed by Koppel et al. [2] to solve the authorship attribution problem in the presence of thousands of candidates. In particular, starting from the original features set, they randomly select a 40% subset of features. Then, they calculate the cosine similarity between users and assign a score of 1 at the most similar user and 0 to the others. They repeat this procedure 100 times. In the end, they sum up the score of each candidate, normalize the score, and use it as a possible threshold to build the precision-recall curve. We refer to this method as *Koppel Baseline*.

For this comparison, we select 1000 alter-egos from the AE_Reddit dataset, and we look for a match into the Reddit dataset.

In figure 3, we show the precision-recall curve of the experiments. The green line shows the curve for the Standard Baseline. It is the method that has the worst results among all the others, with an AUC value of 0.1. The orange line

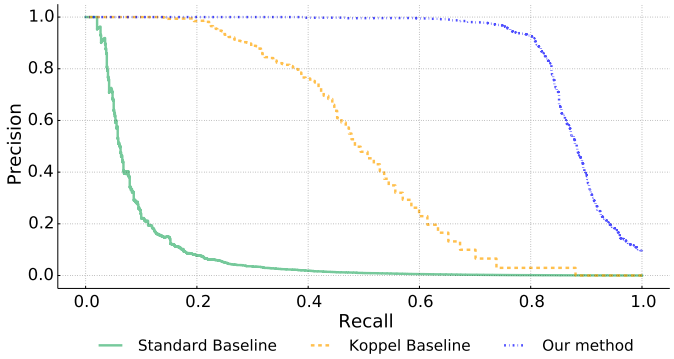


Fig. 3. Precision-recall curves for the Baseline (green), Koppel Baseline (orange), Our method (blue)

represents the Koppel Baseline, with an AUC of 0.49. As we can see, it has comparable performances with the ones reported in the original work [2]. Finally, the blue line represents the precision-recall curve of our method. As we can see, our method outperforms both the baselines with an AUC of 0.88.

Finally, for the sake of curiosity, we also analyze the execution time of the three methodologies. We run the experiments on a Mac OS X 10.14.6, 2.7 GHz Intel Core i5 processor, and 16 GB 1867 MHz DDR3 memory. The Standard Baseline took less than three minutes (155 seconds) but with the worst curve. Our algorithm finished the processing and outputted the result after only 1541 seconds, making it the runner up. The Koppel baseline, instead, is significantly slower with an execution time of more than 40 minutes (2501 seconds).

G. Our methodology on the Dark Web

We found out that our procedure and the threshold perform very well on the Reddit datasets. Since the Dark Web datasets have fewer users, to get a more robust test we merge the TMG with DM dataset, and the AE_TMG with the AE_DM. We refer to them as DarkWeb and AE_DarkWeb, respectively. We define the aliases in AE_DarkWeb as the set of unknown authors, while the aliases in DarkWeb as the set of known authors. So, we perform 10-attribution, and at the end of the process we obtain an accuracy of 98.4%, which is slightly more than the accuracy achieved on Reddit. This means that the search space reduction can be successfully applied also on the Dark Web forums.

Now, we need to verify if the threshold we set works on both forums. So, let AE_TMG be the set of aliases for whom we want to find the aliases in the set TMG. We first apply the reduction, then for each user in AE_TMG we compute the similarity score with their best 10 candidates, and apply the threshold. This way, we get a precision of 90% and a recall of 84%. We repeat the same experiment with the datasets DM and AE_DM and we get a precision of 98% and a recall of 78%. In Table V we report for all the datasets the thresholds associated with 80% recall, and the precision and recall values for the chosen threshold.

TABLE V
PRECISION-RECALL WITH DIFFERENT THRESHOLDS.

Forum	threshold	Precision	Recall
Reddit_A	0.4190	94%	80%
Reddit_B	0.4210	91%	80%
DM	0.4096	96%	80%
TMG	0.4222	94%	80%
Reddit_A	0.4190	94%	80%
Reddit_B	0.4190	87%	82%
DM	0.4190	98%	78%
TMG	0.4190	90%	84%

TABLE VI
AUC VALUES.

Forum	AUC with reduction	AUC without reduction
Reddit	0.89	0.79
TMG	0.93	0.91
DM	0.94	0.91

As we can see, the results on the Dark Web forums are slightly higher than those of Reddit. We believe that it depends on two factors. The first is that Reddit has many more users than the Dark Web. The second is that while both The Majestic Garden and the Dream Market are only focused on drugs, hence all the messages belong to the same domain, the Reddit dataset is a 'quasi-single domain' since the majority of messages are about drugs, but not all of them. Experiments show that our methodology performs well in all the three forums and that the same threshold can be applied to all of them.

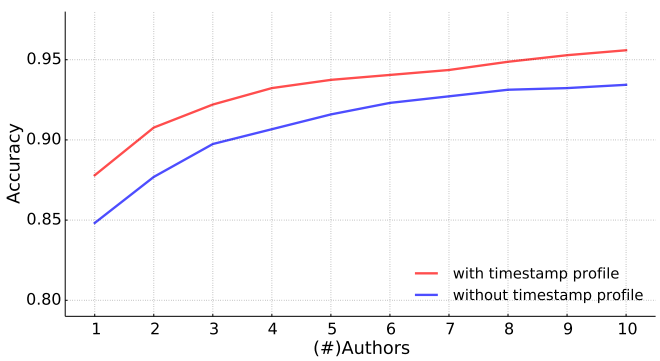
H. Performance improvement

In this section we focus on the effectiveness of the daily activity feature and the search space reduction.

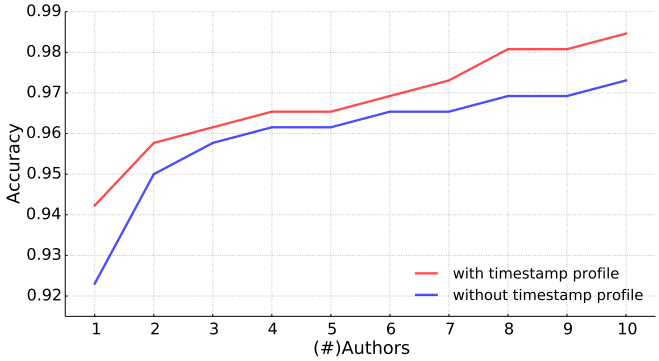
Fig. 4 shows k-attribution on the Reddit dataset (Fig. 4(a)) and on the Dark Web forums (Fig. 4(b)). The blue lines on the figures show the accuracy achieved varying k if we use only the text features, while the red lines when we also add the daily activity profile. As we can see in both the cases, using the daily activity improves the accuracy performance. This boost allows us to use less text in our procedure, so we can evaluate more users. Regarding the space reduction, to estimate how this step improves the performance, we compute the precision-recall curve on the same datasets with and without the reduction, using AUC as evaluation metric. AUC is the area under the curve of the precision-recall curve, the value of AUC can vary from 0 up to 1—the higher is the value, the better. Tab. VI shows the AUC values for our forums before and after the reduction step. As we can see, the AUC values after the reduction are always higher. In Fig. 5 we show the precision recall curves, before and after space reduction.

I. The final algorithm

Our algorithm performs two steps of cosine similarity using the features in Table II. At first, we extract the text features

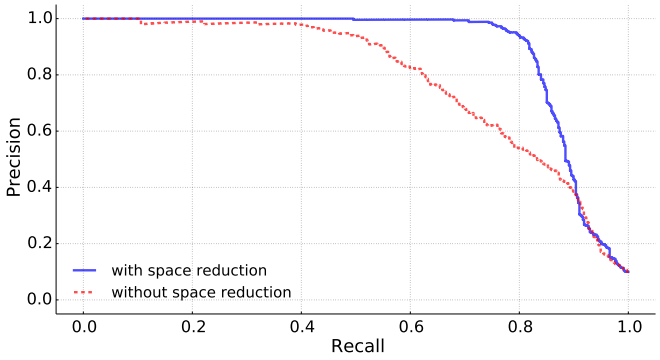


(a) k-attribution accuracy on Reddit.

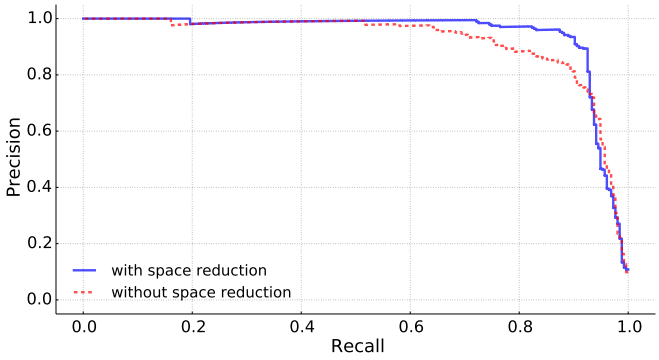


(b) k-attribution accuracy on Dark Web forums

Fig. 4. Impact of the daily activity feature.



(a) PR curves with and without space reduction on Reddit.



(b) PR curves with and without space reduction on Dark Web forums.

Fig. 5. Precision and Recall with and without search space reduction

from the documents associated with the set of known users Z, we rank the n-grams by frequency, and then we select the top N according to Table II. We extract the same selected features from the unknown document D. Finally, we weight the features with the Tf-Idf. Along each text feature vector, we concatenate the daily activity profile of the user. We compute the cosine similarity between each known profile against the unknown one and sort the results in descending order keeping the first k. We then perform feature extraction, and we recompute the Tf-Idf on the documents of these k users. This changes the sequences of words and chars selected by frequency and consequently the Tf-Idf weighting, providing a new feature vector for each user. Of course, this procedure changes the feature vector of the unknown alias too. We finally determine the cosine similarity between the k profiles and the unknown profile, and we output the pair if the similarity score is higher than the threshold t.

J. Process the data in batch

Dealing with a large amount of aliases and features, can lead to the possibility to incur in hardware constraint— lack of available RAM. To overcome this limitation, we build the following iterative procedure.

Let B, indicate the maximum number of candidates aliases that the hardware can handle. We divide the total number of aliases in batches of B. Then, we apply on each batch the 10-attribution step. Now, if the sum of the resulting candidate is less than B, we apply the final step of our methodology, else we divide the candidate aliases again in batches, and we repeat the 10-attribution step. To validate this procedure, we apply it on the same dataset used for the comparison of the baseline in section IV-F, with B equals to 100. At the end of the computation, we get for the same threshold (0.4190) a precision of 91% and a recall of 81%.

V. Results

After validating our methodology, we used it to break the pseudo-anonymity between the two Dark Web forums, and to de-anonymize Dark Web users.

A. Evaluation methodology

Since we do not have a ground truth, we manually evaluate our results. For each pair above the threshold, we analyze the full data retrieved for both the aliases looking for evidence. Then, we classify the pair in the following manner:

- **True**: If we find clear evidence that the two aliases belong to the same user. For example, the user declares her username on the other forums or leak some data that is unique to the user such as the same mail address.

- **Probably True**: If we find evidence that the two users could be the same person, but we can not be sure.
- **Unclear**: If the messages do not leak any information or evidence that we can exploit.
- **False**: If the two matching users disclose information about themselves that are contradictory.

B. The Majestic Garden vs Dream Market

In this experiment, we want to find users that actively participate in The Majestic Garden and Dream Market forums, to break the pseudo-anonymity across the two forums. For these experiments, we used the TMG and DM datasets, that are made respectively of 422 and 178 users. The algorithm outputs 11 possible matches. Out of these 11 pairs, 7 are **True**, since in their raw text we find references to their alias in the other forum, 1 is **Unclear** since there is no reference but we find common opinions expressed in the same way, and the other 3 are **False**.

C. Reddit vs Dark Web

As our last experiment, we want to find matching aliases between the Dark Web and the standard Internet. So, we look for both TMG and DM users on Reddit. At the end of the experiment, we get 47 possible matches. After manual investigation, we classify 20 of these matches as **True** pairs. Some of them are vendors in one of the two forums. These users refer to themselves in the Dark Web. In fact, since they are vendors, they use their name as a brand. One of them posted a link on Reddit, and some hours after she posted the same link in the Dark Web stating that she posts the same link on Reddit to look at users' reaction. Exploiting the link again, we catch another match. In this case, she advertised a web site platform through a referral link and she uses the same link both on Reddit and in the Dark Web, moreover, in the URL there was her nickname. Another user convinced us to be the same person because both on the standard Internet and in the Dark Web she states to live in Miami, she declares to assume the same kinds of drugs, but mostly she complained about the same vendor that sold her poor quality "white molly", describing the drugs in the same way and just at one day apart from one platform and the other. The last one has the name of a philosopher in her nickname in the Dark Web, she frequently suggests to take yoga lessons or cooking classes, and she shows knowledge of the blockchain technology. The Reddit alter-ego speaks about the philosopher, and she is really addicted to cooking, yoga, and the bitcoin subreddit.

Probably True: We labeled 2 users as Probably True. For these users, we find some common patterns, such as they declare to live in the same country and to buy the same drugs from the same vendor, or they show to have the same hobbies.

Unclear: These users do not leak any kind of information about themselves, or they leak just on one of the two platforms. Other kinds of users that we insert in this list are the users that share only the kind of drugs they use. Reading the forum we noticed that per se it is not discriminative information. In this set, there are 20 users.

False: We labeled as false 5 pairs. In particular, one match declares to be 20 years old on the Dark Web and to be 34 on Reddit. Another user states to be Christian and her counterpart to be Atheist. One is a Trump supporter and she is a subscriber of subreddit pro-Trump, the alter-ego instead argued on the Dark Web against Trump. One pair live in Poland, and the other pair are American citizens. The last one declares to assume MDMA regularly, the alter-ego states that she never tried this kind of drugs.

D. Exploiting the Reddit posts

An in-depth analysis of active users on Reddit can lead to build a very detailed profile. We did this experiment on one of the True pair we found, using the Reddit web interface that shows the comments of a user[1]. We gathered the following information about our John Doe. John is a 27 years old man from Edmonton, Alberta, Canada. He lives with his parents and his brother. He had a job but lost it because of drug abuse. He was in a relationship for more than 2 years. He likes to play video games, in particular online ones, and has an account on Fallout, League of Legends, COD4, and Counter Strike. His phone in 2017 was a Samsung Galaxy S4. He travels frequently to the USA, New York in particular. We also gathered other information that we believe is too sensible to be disclosed such as medical conditions, social events he joined in Edmonton, and the bars where he usually goes. And, of course, now we know his alias in the Dark Web.

VI. Discussion

From Reddit to the full de-anonymization of the user Previously we declared the user matched on Reddit as fully de-anonymized. This is completely true if the user uses a classic HTTPS connection to surf Reddit. In fact, in this case, she unveils her IP address, and Reddit servers can log it. We believe this scenario is highly possible as analyzing the messages of users on Reddit we notice that some of them, when out of the Dark Web, as shown in V-D, let their guard down and forget about their privacy. They use the same alias to write on the Dark Web markets subreddits and on subreddit related to their personal life interests, writing in their native language and disclosing personal information. On the other hand, there are some more careful users, that use the Reddit alias to write only on Drugs related subreddits, and that do not share private information. Moreover, reading the forums, we found that some users have a dedicated computer only to surf the Dark Web. These users probably use Tor connection even to write on Reddit, meaning that we cannot de-anonymize them. Although, Reddit is a point of failure for them. Authorities can, in fact, focus on a few aliases, and with Reddit help, they could perform different kinds of attacks to de-anonymize a specific user.

About the sampling time range. As we have shown, exploiting the timestamps provides a strong boost on detection performance, but we believe that this could not always be true.

[1] https://www.reddit.com/user/{nickname}/comments/

In the long run, people can change their habits because of the birth of a son or simply because they change work, hobbies, and so on. It is important that the timestamps collected from the authors to compare belong to the same time range. Of course, the shorter is the time range and more timestamps we have, the more reliable will be the generated profile. The same holds for written messages, users change their writing style, maybe more gradually with respect to the habits. In our work almost all the posts in the datasets were written in the same year, 2017, and we discard the timestamp of the post written on weekends and holidays in order to be less prone to this kind of bias.

Avoiding the attack. Our attack relies on a combination of temporal and stylometric features, so we analyze them separately. The best way to protect themselves against daily activity profiles attack on different platforms is to post on a completely different time, for example on one forum in the morning and the other in the evening. Although this is possible, we believe that is almost impractical. Handling properly when and where to write, means that the user is not free to act naturally, and if she wants to be active in the community, she must re-shape her habits. Moreover, the users we investigate about, write on NSFW topics— Not Safe For Work, this means that usually they can read or write on these forums only when they are alone, or no one can see what they are doing. As for the text, a user can use adversarial stylometry tools in order to obfuscate her linguistic features. The use of these tools requires a constant effort on behalf of the user, since each time she wants to post a new message she needs to use the tools and check if the meaning of the original message is the same as the obfuscated message, otherwise she needs to re-adjust it. The best way to avoid the detection is to use disposable aliases, but in this case, the user becomes indistinguishable from the other community members, so she can neither build her personal reputation inside the community nor have any kinds of relationship with other members, hence the user is cut off from the community life.

VII. Ethical Considerations

In this work, we analyzed 11,058 anonymous users of two different forums in the Dark Web, 16,567 users on Reddit, for a total number of more than 7 million posts. The data collected from the Dark Web and Reddit were encrypted and stored for a limited amount of time in our servers. In addition, user nicknames were hashed to further protect user's privacy. Original data were not shared directly or placed on platforms from where it could be downloaded. Consequently, and accordingly to the policy of our IRB, we did not need any explicit authorization to perform our experiments.

VIII. Related works

Account matching. Nowadays, users have one or more accounts on multiple social media and sites. Linking them to the same user can be a precious information to obtain. Vosoughi et al. [22] perform an analysis of digital stylometry for linking profiles across two social networks: Facebook and Twitter. They collect a total of 5,612 users and 11,224 accounts. They built a profile for each user in one of the dataset and then, given n number of users from the other dataset, their model rank them from the most likely to be the correct match, to the last one using cosine similarity as measurement. Their best accuracy result is 31%. Johansson et al. [23] perform a similar analysis with similar features on an Irish Web forum dump. For each of the 1000 users they collected, they create 2 aliases splitting their text into two parts. They measure the similarity between aliases to match them correctly, reaching an accuracy up to 56%. The work of Spitters et al. [24] focuses on the Dark Net markets Black Market Reloaded. They collected 177 users and create the ground truth splitting each user into two aliases. Their best result is 91% precision and 25% recall.
Authorship Attribution and Verification. While aforementioned works link accounts exploiting any kind of information associated to the user, in the following we described some works that rely only on the use of stylometric features. In Narayanan et al. [3] they used a dataset of 100,000 blogs of different users. They reach their best result using Nearest-Neighbor/RLSC combination with an accuracy of 80% with a 50% of recall. Abbasi et al. [25] focused on unsupervised methods to perform user identification and similarity detection. Their best result is in the eBay comments dataset with 96% of accuracy on 25 users. Some works focused on Neural Network to succeed in this task. Shrestha et al. [26] were the first to use a Convolutional Neural Network. They reach an accuracy of 72% on a dataset of 50 users with 1000 tweets each. On Twitter's tweets, also focused Layton et al. [19]. Using SCAP method [27], they reached an accuracy of more than 70% with 50 users and 140 tweets each. Overdorf et al. [28] focused on cross-domain Authorship. Their results are in a range of accuracy between 20% and 80% for cross-domain and in domain respectively. Schwartz [21] et al. uses k-signature of an author, along with other features, to train an SVM classifiers that lead to an accuracy of 71%. Afroz et al. [29] perform this task starting from SQL dumps of three underground forums. Their best result is 84%. Recently, Brocardo et al. [30] explore the use of a Deep Belief Network to perform Authorship Verification. Their results, expressed as Equal Error Rate(EER), are EER of 8.2% on Twitter and of 5.4% on the forgery dataset. Ruder et al. [31] perform an extensive comparison between several CNN approaches and traditional ones. They outperform the state-of-the-art results in three out of six datasets they tested on. This task is also addressed in PAN, a series of scientific events and shared tasks on digital text forensics and stylometry [32]. The last task was in 2019 [33]. First [34] and second [35] ranked in the Cross-Domain Authorship Attribution task, reach an F1_score of 69% and 68% respectively. Finally, for a comprehensive view, Neal et al. [36] published a survey of stylometry techniques and applications that better described the state of the art so far.

IX. Conclusion

In this paper, we have shown that the privacy of Dark Web users is in danger when in the standard Internet. Daily activity profile and writing style are enough to link the dark alias to the real identity with excellent performance. Indeed, using our algorithm, we found 58 matches from the Dark Web to the standard Internet. For 27 of them, we were able to confirm that they belong to the same person. Lastly, we discovered that most of the de-anonymized users are careless when using the open alias, and that it is often easy to outline very detailed personal profiles.

Lastly, our findings bring up the need for more work on software that is able to anonymize writing patterns. Although a few applications do exist—like Anonymouth [37]—this is still an open problem.

Acknowledgment

This work was supported in part by the MIUR under grant "Dipartimenti di eccellenza 2018-2022" of the Department of Computer Science of Sapienza University.

References

[1] blog.chainalysis.com. [Online]. Available: https://blog.chainalysis.com/reports/decodingdarknetmarkets

[2] M. Koppel, J. Schler, and S. Argamon, "Authorship attribution in the wild," *Language Resources and Evaluation*, vol. 45, no. 1, pp. 83–94, 2011.

[3] A. Narayanan, H. Paskov, N. Z. Gong, J. Bethencourt, E. Stefanov, E. C. R. Shin, and D. Song, "On the feasibility of internet-scale author identification," in *2012 IEEE Symposium on Security and Privacy*. IEEE, 2012, pp. 300–314.

[4] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The second-generation onion router," Naval Research Lab Washington DC, Tech. Rep., 2004.

[5] Reddit. [Online]. Available: www.reddit.com

[6] Amazon. (2001) Alexa. [Online]. Available: www.alexa.com

[7] L. Franceschi-Bicchierai. (2018) Reddit bans subreddits dedicated to dark web drug markets and selling guns. [Online]. Available: https://www.vice.com/en_us/article/ne9v5k/reddit-bans-subreddits-dark-web-drug-markets-and-guns

[8] Dream market forum. [Online]. Available: http://lchudifyeqm4ldjj.onion

[9] The majestic garden. [Online]. Available: http://http://talismanrestz7mr.onion

[10] E. Ormsby. (2016) The dark web's largest acid dealer is running an art competition. [Online]. Available: https://www.vice.com/en_uk/article/ppv4dm/the-dark-webs-largest-acid-dealer-is-holding-an-art-competition

[11] ——. (2016) Dark web lsd stalwarts jesusofrave in their own words. [Online]. Available: https://allthingsvice.com/2016/11/09/dark-web-lsd-stalwarts-jesusofrave-in-their-own-words/

[12] langdetect. [Online]. Available: https://pypi.org/project/langdetect/

[13] Google java library. [Online]. Available: https://code.google.com/archive/p/language-detection/

[14] M. La Morgia, A. Mei, S. Raponi, and J. Stefa, "Time-zone geolocation of crowds in the dark web," in *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2018, pp. 445–455.

[15] M. Koppel, J. Schler, S. Argamon, and E. Messeri, "Authorship attribution with thousands of candidate authors," in *SIGIR*, vol. 6. Citeseer, 2006, pp. 659–660.

[16] K. Luyckx and W. Daelemans, "Authorship attribution and verification with many authors and limited data," in *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*. Association for Computational Linguistics, 2008, pp. 513–520.

[17] ——, "The effect of author set size and data size in authorship attribution," *Literary and linguistic Computing*, vol. 26, no. 1, pp. 35–55, 2011.

[18] M. Eder, "Does size matter? authorship attribution, small samples, big problem," *Digital Scholarship in the Humanities*, vol. 30, no. 2, pp. 167–182, 2013.

[19] R. Layton, P. Watters, and R. Dazeley, "Authorship attribution for twitter in 140 characters or less," in *2010 Second Cybercrime and Trustworthy Computing Workshop*. IEEE, 2010, pp. 1–8.

[20] C. Sanderson and S. Guenter, "Short text authorship attribution via sequence kernels, markov chains and author unmasking: An investigation," in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2006, pp. 482–491.

[21] R. Schwartz, O. Tsur, A. Rappoport, and M. Koppel, "Authorship attribution of micro-messages," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013, pp. 1880–1891.

[22] S. Vosoughi, H. Zhou, and D. Roy, "Digital stylometry: Linking profiles across social networks," in *International Conference on Social Informatics*. Springer, 2015, pp. 164–177.

[23] F. Johansson, L. Kaati, and A. Shrestha, "Detecting multiple aliases in social media," in *Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining*. ACM, 2013, pp. 1004–1011.

[24] M. Spitters, F. Klaver, G. Koot, and M. van Staalduinen, "Authorship analysis on dark marketplace forums," in *2015 European Intelligence and Security Informatics Conference*. IEEE, 2015, pp. 1–8.

[25] A. Abbasi and H. Chen, "Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace," *ACM Transactions on Information Systems (TOIS)*, vol. 26, no. 2, p. 7, 2008.

[26] P. Shrestha, S. Sierra, F. Gonzalez, M. Montes, P. Rosso, and T. Solorio, "Convolutional neural networks for authorship attribution of short texts," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, vol. 2, 2017, pp. 669–674.

[27] G. Frantzeskou, E. Stamatatos, S. Gritzalis, C. E. Chaski, and B. S. Howald, "Identifying authorship by byte-level n-grams: The source code author profile (scap) method," *International Journal of Digital Evidence*, vol. 6, no. 1, pp. 1–18, 2007.

[28] R. Overdorf and R. Greenstadt, "Blogs, twitter feeds, and reddit comments: Cross-domain authorship attribution," *Proceedings on Privacy Enhancing Technologies*, vol. 2016, no. 3, pp. 155–171, 2016.

[29] S. Afroz, A. C. Islam, A. Stolerman, R. Greenstadt, and D. McCoy, "Doppelgänger finder: Taking stylometry to the underground," in *2014 IEEE Symposium on Security and Privacy*. IEEE, 2014, pp. 212–226.

[30] M. L. Brocardo, I. Traore, I. Woungang, and M. S. Obaidat, "Authorship verification using deep belief network systems," *International Journal of Communication Systems*, vol. 30, no. 12, p. e3259, 2017.

[31] S. Ruder, P. Ghaffari, and J. G. Breslin, "Character-level and multi-channel convolutional neural networks for large-scale authorship attribution," *arXiv preprint arXiv:1609.06686*, 2016.

[32] Pan. [Online]. Available: https://pan.webis.de/

[33] M. Kestemont, E. Stamatatos, E. Manjavacas, W. Daelemans, M. Potthast, and B. Stein, "Overview of the Cross-domain Authorship Attribution Task at PAN 2019," in *CLEF 2019 Labs and Workshops, Notebook Papers*, L. Cappellato, N. Ferro, D. Losada, and H. Müller, Eds. CEUR-WS.org, Sep. 2019.

[34] L. Muttenthaler, G. Lucas, and J. Amann, "Authorship attribution in fan-fictional texts given variable length character and word n-grams," *Working Notes Papers of the CLEF*, 2019.

[35] A. Bacciu, M. La Morgia, A. Mei, E. N. Nemmi, V. Neri, and J. Stefa, "Cross-domain authorship attribution combining instance-based and profile-based features," *Working Notes Papers of the CLEF*, 2019.

[36] T. Neal, K. Sundararajan, A. Fatima, Y. Yan, Y. Xiang, and D. Woodard, "Surveying stylometry techniques and applications," *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, p. 86, 2018.

[37] A. W. McDonald, S. Afroz, A. Caliskan, A. Stolerman, and R. Greenstadt, "Use fewer instances of the letter "i": Toward writing style anonymization," in *International Symposium on Privacy Enhancing Technologies Symposium*. Springer, 2012, pp. 299–318.