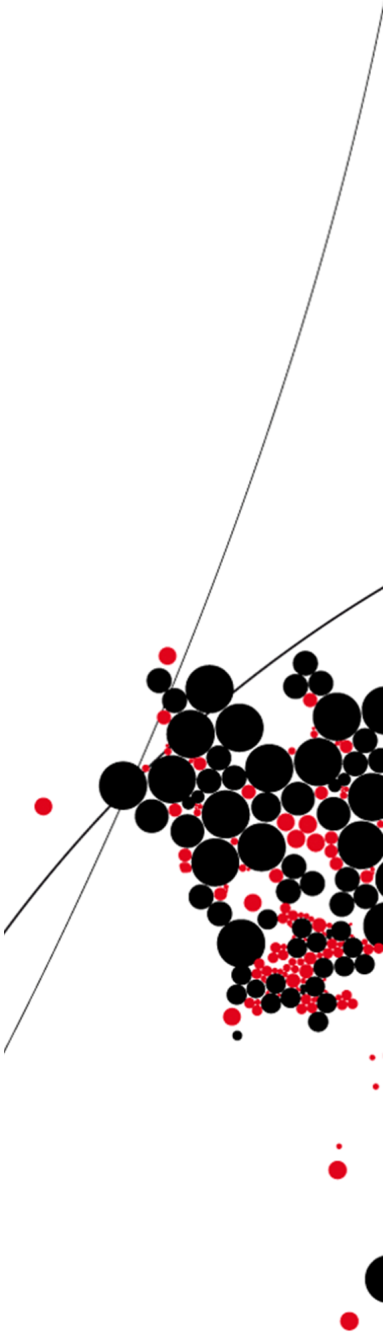
UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering, Mathematics & Computer Science

Using Stylometry to Track Cybercriminals in Darknet Forums

Anirudh Ekambaranathan

M.Sc. Thesis
July 2018



Graduation committee:
Dr. A. Peter
Dr. M. H. Everts
External supervisor:
Dr. S. Meiklejohn

Telecommunication Engineering Group
Faculty of Electrical Engineering,
Mathematics and Computer Science
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

Using Stylometry to Track Cybercriminals in Darknet Forums

Anirudh Ekambaranathan, Andreas Peter and Sarah Meiklejohn

***Abstract*—Darknet markets are becoming increasingly popular, making it important for law enforcement agencies to be aware of state of the art techniques on tracking and analysing key participants. In this work, we present an unsupervised method for linking user pseudonyms based on stylometry. We show on a Twitter dataset of 1,000 users that our method is 98.7% accurate. We also construct a dataset containing the user migration after a darknet market closure. Subsequently, we use this dataset to show that our linking technique can be used to track displacement of users, even when ground truth data is not readily available. The results show that using bi-grams as input features, linkability can be achieved on a large scale. Even though effective linkability requires a minimum of 25 posts per user, we can still link a majority of active members in darknet market migrations. We also test five countermeasures to evade our linking technique and show that none of the measures would uphold if law enforcement agencies decided to perform dedicated linkage attacks.**

I. Introduction

Darknet markets are much like traditional markets and provide a platform for users to exchange goods. However, unlike traditional markets, they reside on anonymous darknets, making them only accessible through special software, such as TOR. Darknets are attractive to those who wish to remain anonymous, such as whistleblowers, political dissidents, and people dealing in illegal goods. Though darknet markets can be used to buy and sell legal goods, it is estimated that more than half of the content on the darknet is illegal [34], with the majority of the offerings being drug related [11]. Over the past few years, these markets have rapidly gained popularity and have thereby caught the attention of law enforcement agencies.

Most commonly, the efforts at tackling online illicit trade are aimed at disrupting marketplaces (*crackdowns*) [5]. Since the rise of darknet markets, two majors police crackdowns have taken place. The last major crackdown, 'Operation Onymous', happened in November 2014 and led to the arrest of 17 people and the closing of multiple high profile marketplaces [12]. There is, however, little evidence that such crackdowns are an effective method for decreasing drug sales [12]. They have a time-limited impact, after which market participants use displacement techniques to continue their activities on different markets [13], [17], [36].

Instead, Decary et al. [12] suggest to target key players of the Dark Web community, as most of the sales are caused by a small portion of the vendors [30]. To this end, it may be useful to analyse the contents of the accompanying darknet forums to identify where key players *migrate* to after crackdowns. The aim is then to *match* or *link* aliases from different forums belonging to the same person. One way to do this, is by clustering accounts with similar usernames [23]. This heuristic has formerly been applied to measure displacements and identify multiple aliases across different marketplaces [12], [30]. However, measuring user migrations is often not so trivial, since individuals may operate under different pseudonyms and can change their usernames between markets. This challenge is therefore often tackled by analysing other structural clues left behind by users, such as message contents, nationality, and timestamps.

Stylometry is often used for author linkability, either by comparing sets of documents [4], [32], or by directly comparing distances between authors based on 'writeprints' (a stylometric fingerprint) [1], [3], [25]. However, existing methods often make use of synthetic datasets, where authors artificially split users into multiple identities. The problem is that there are no existing datasets specifically designed to experiment with alias matching. This can lead to a reduced accuracy when applied to actual separate accounts. Furthermore, linkability studies often operate under the assumption that it is known whether users have multiple accounts, which is not always the case for real world applications. For instance, in market migrations, ground truth data is unavailable, as it is not known beforehand whether users migrated or not. This makes linking challenging, since it is not possible to make use of supervised methods.

In this study we look at methods to overcome these limitations. We propose a technique for author linking based on stylometric features, which can match aliases even when ground truth data is unavailable. We analysed 13 darknet forums and more than 10 events, such as market closures and exit scams, to identify and create a dataset to effectively test our algorithm on real data. This was no trivial task, as migration patterns are often unruly and do not provide a basis to make ground truth measurements. Our main contributions are threefold:

Firstly, we present an unsupervised distance-based method for alias matching and show that it works on a controlled Twitter dataset. It is 98.7% accurate when applied on a set of 1,000 accounts. Furthermore, it is 94.1% accurate when applied in a setting where ground truth data would normally not be available.

Secondly, we analyse the migration of users after the exit scam of the darknet market *Evolution*. By applying the user-

A. Ekambaranathan was a student at the University of Twente, 7500 AE Enschede, the Netherlands (e-mail: a.g.ekambaranathan@student.utwente.nl)

A. Peter is with the department of Services, Cyber Security Safety Group, University of Twente, 7500 AE Enschede, the Netherlands (e-mail: a.peter@utwente.nl)

S. Meiklejohn is with the department of Computer Science, University College London, London WC1E 6BT, UK (e-mail: see https://smeiklej.com/)

name similarity heuristic, we can track 41.4% of the migrants. We use this to show that our alias matching technique works on darknet forum users with an accuracy of more than 90.0%. Subsequently, we apply it on the remaining users whose usernames have changed and show that active members can be linked with high confidence.

Lastly, we test our method against five different countermeasures and provide suggestions as to how linkability attacks can be evaded.

The rest of the paper is structured as follows. In section II, we describe previous work done in the field of linkability and authorship analysis. In section III, we detail our linkability algorithm and explain the settings in which it will be tested. In section IV, we state and discuss the results. Lastly, in section V, we give the conclusion and make suggestions for future studies.

II. Related work

A. Alias matching and linkability

Alias matching can broadly be split into four different categories [16]: (1) string-based matching makes use of the alias names, (2) stylometric matching is based on the writing styles of authors, (3) time profile-based matching is based on the publication times of the posts, and (4) social network-based matching makes use of relationships between users. In this work we apply string-based (1) and stylometric matching (2).

a) String based matching: The similarity between the string-based aliases of users can be a useful feature for linking. Various edit distances between strings have been proposed, such as the Levenshtein [37] distance and the Jaro-Wrinkler distance [35].

Zafarani et al. [38] link usernames across different communities by extending base names with prefixes and suffixes. The method was tested by searching for candidate names through the Google search engine and was 66% accurate. Similarly, Perito et al. [23], analyse linkability and uniqueness of usernames across multiple domains and show that a large number of user profiles can be linked. They make use of a Markov chain model, which is trained on approximately 10 million usernames extracted from Google and eBay. Their model looks for similar substrings between usernames and determines how unique they are. If these substrings appear unique, it is more likely that the usernames belong to the same person.

b) Stylometric matching: Stylometry is the statistical analysis of writing styles [39] and is often used for authorship attribution and linking. Over the years a lot of methods have been proposed. Almishari et al. [4] show that a Naive Bayes classifier can be 95% accurate for linkability on a Twitter dataset of more than 7,000 users. However, they make the assumption that it is known whether users have multiple accounts.

Abbasi et al. [1] developed a method called *Writeprints*, which is used to construct a stylometric fingerprint of a user. The 'writeprints' are based on features which are important to one author and which are less important to other authors. Then, the 'writeprints' of different users are compared to determine whether they belong to the same person. Their approach is 91.3% accurate, for 100 authors, on a dataset of eBay comments, but is only 52.7% accurate on Java forum data. Furthermore, the performance of their method decreases as the number of users increases.

B. Darknet forums

In recent years, darknet markets have extensively been studied. However, authorship analyses and alias matching in darknet forums is limited. Spitter et al. [31], perform alias matching on the Black Market Reloaded forum, by artificially splitting the posts of 177 users. They reach a precision and recall of approximately 0.45 and 0.55 respectively.

Afroz et al. [2] develop a probabilistic algorithm called *Doppelgänger Finder*, which was tested on a dataset of two darknet forums: *L33tCrew* and *Carders*. Between the two forums they found 28 pairs of users, which their algorithm matched with a precision and recall of 0.85 and 0.82 respectively. A drawback of this method is that it requires a lot of manual handling of the input data. For example, they make use of parts-of-speech tagging, which is language dependent. This requires *a priori* knowledge of the language of the text and requires installing new taggers when texts in different languages are involved. The advantage of our method is that the feature extraction process is automated and that it is independent of the language.

Soska et al. [30] look at the use of multiple aliases between Silk Road 1.0, Black Market Reloaded and Sheep. They make use of the username similarity heuristic and also match aliases based on public PGP keys. From an initial list of 29,258 unique aliases, they reduced it to 9,386 vendors. We use a similar heuristic to construct a base dataset to measure the effectiveness of our algorithm. This way, instead of using an artificial dataset, as in [31], we can work with real data.

C. Author obfuscation

Author obfuscation is a generalised term for techniques relating to the obfuscation of writing styles, with the aim of evading author identification. These techniques are either performed manually, are computer assisted, or are entirely automated. Our aim in studying countermeasures is to understand how our technique can be bypassed and which measures criminals could be using to evade linking in practice. Below we briefly discuss manual and automated techniques.

Manual techniques were extensively studied by Brennan and Greenstadt [8], who asked 12 people to obfuscate their writing style and to imitate the writing styles of other authors. They show that it is possible to alter your own writing style, to the degree that automated authorship attribution performs no better than random. Brennan et al. [7] replicated this study with 45 writers, making use of additional crowdsourcing via Amazon's Mechanical Turk.

Rao and Rohatgi [26] propose an automated machine translation technique, wherein text is first translated to an intermediate language and then translated back to the original language. The theory here is that a round-trip of machine translations distorts the text enough to obfuscate the writing

style. Caliskan and Greenstadt [9] analyse this on authorship attribution and show that translated texts contain enough features to effectively attribute them to their original authors. But, translating through more intermediary languages does reduce the accuracy. We apply this technique by also randomising the languages in an effort to understand its effect on linkability.

Koshmood and Levinson [18], [20] suggest an approach where the writing style of a document is iteratively altered to match the writing style of a target document. Koshmood [19] also suggests altering sentence level structures, such as changing the tense, replacing certain words with their synonyms, and changing diction. However, the disadvantage of this is that the meaning of the message is often altered in the process, as it is difficult for computers to interpret the meaning of sentences.

III. Settings and methodology

In this section we explain:

1) the general task of alias classification.
2) our approach to solving this task.
3) how we test it in a controlled environment of Twitter data.
4) how it can be used to measure migrations between darknet forums.
5) and the countermeasures we use to test how our linking technique can be evaded.

A. Alias matching and pairing

In this work we make the distinction between two tasks: *alias matching* and *alias pairing*. Though in previous literature this distinction is not formally defined, it will help avoid confusion.

1) Alias matching: Given the set, $A = \{a_1, a_2, ..., a_n\}$, of n feature vectors and belonging to a pool of n authors, the task of *alias matching* is to cluster all vectors, $a_i, ..., a_j$, belonging to the same person. A feature vector is an n-dimensional vector of numeric values representing an alias. This is an unsupervised learning problem, as it is unknown beforehand whether authors have multiple aliases.

2) Alias pairing: Given two sets of feature vectors, $A = \{a_1, a_2, ..., a_n\}$, and $B = \{b_1, b_2, ..., b_n\}$, from a pool of $2n$ authors, the task of *alias pairing* is to match every feature vector from A to a feature vector from B.

This task is more common when working with artificial datasets, where the posts of users are split into two subsets. The assumption is then made that it is known beforehand whether users operate under multiple aliases. This task is useful for benchmark measurements.

B. Approach

We propose an unsupervised method to create 'writeprints', which we call *author embeddings*, that extends on the concept of *word embeddings*.

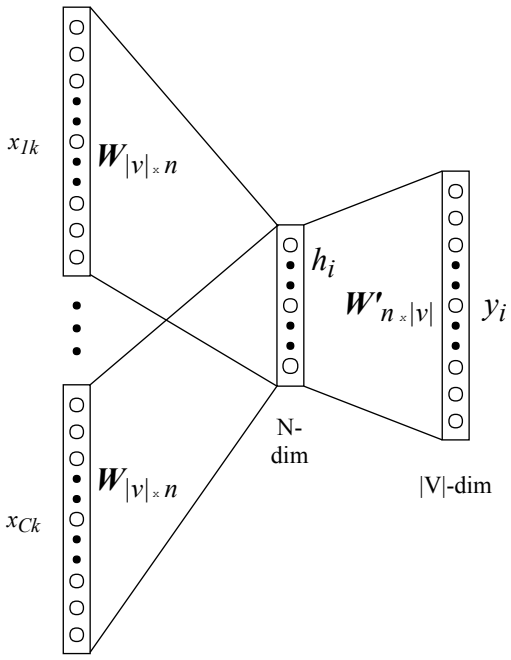


Figure 1. Neural network architecture of a Continuous-Bag-of-Words model.

1) Word and document embeddings: Word embeddings were first introduced in the early 2000s [6] and are functions mapping words to high dimensional vectors. More recently, the `Word2Vec` software has gained popularity and provides state of the art word embeddings [22]. It is a suite of two algorithms: Continuous-Bag-of-Words (CBOW) and Skip-Gram. Intuitively, both algorithms determine the vector values of words based on its context. Here we briefly explain the Continuous-Bag-of-Words model.

Continuous-Bag-of-Words. Given a *focus* word w and its *context* words c, from a document D, the aim of CBOW is to maximize the conditional probability of w:

$$\arg\max_{\theta} \prod_{(w,c) \in D} p(w|c;\theta) \tag{1}$$

where θ is the parameter which will be optimized. This can be parametrized with a neural network, by modelling the conditional probability $p(w|c;\theta)$ using soft-max:

$$p(w|c;\theta) = \frac{e^{v_w \cdot v_c}}{\sum_{w' \in V} e^{v_{w'} \cdot v_c}} \tag{2}$$

where V is the vocabulary, and v_w and v_c are vector representations for w and c respectively. Figure 1 represents this graphically.

The weights between the input layer and the hidden layer is a $|V| \times N$ matrix $\mathbf{W}$ (and form the word embeddings), where N is the dimensionality of the embeddings.

Taking the log likelihood of (1) leads to the following equation:

$$\arg\max \sum_{(w,c)\in D} \log p(w|c;\theta) = \sum_{(w,c)\in D} (\log e^{v_w \cdot v_c} - \log \sum_{w'} e^{v_{w'} \cdot v_c}) \tag{3}$$

For a more elaborate discussion of word embeddings using `Word2Vec` the reader is referred to [29], [14].

Doc2Vec. Similar to `Word2Vec`, it is also possible to create embeddings of entire documents [21]. `Doc2Vec` extends Word2Vec by adding an extra feature input vector into the neural network: the document ID. When the word vectors are trained, the document vector is trained as well.

The objective of the `Doc2Vec` algorithm is to maximize the probability of a word, given the context and the document ID:

$$\arg\max_{\theta} \prod_{(w,c,d_{id})\in D} p(w|c,d_{id};\theta) \tag{4}$$

The assumption is that similar documents have similar document vectors.

2) Our approach: We propose to extend this model to create *author embeddings*. To do so, we first expand the corpus with character and word level n-grams, which capture lexical preferences of authors. By using multiple values of n, it is possible to create embeddings which contain a full range of the stylistic features of an author. The advantage of n-grams is that they do not require any *a priori* knowledge of the grammar of the language, and they have shown to be effective for authorship analysis [15], [33].

Then, similar to `Doc2Vec`, the corpus embeddings are created with the author ID as an extra input feature, where the corpus now consists of all tokens after expansion:

$$\arg\max_{\theta} \prod_{(t,c,a_{id})\in D} p(t|c,a_{id};\theta) \tag{5}$$

The hypothesis is that authors with similar writing styles, have similar vectors. We use a basic approach to measuring this, namely by computing the cosine distance between the feature vectors. The cosine distance between two vectors, u and v is defined as:

$$1 - \frac{u \cdot v}{||u||_2||v||_2} \tag{6}$$

C. Datasets

We will apply our algorithm in different settings. Firstly, we test the effectiveness of *author embeddings* using a Twitter dataset. Secondly, we adapt our techniques to darknet forums. Lastly, we test how well our algorithm fares against various countermeasures. For this, we make use of a dataset constructed by Brennan et al. [7] and also manipulate the Twitter and darknet data.

Table I
TWITTER DATASET STATISTICS

Parameter	Value
Total # of Tweets	6,822,774
Total # of Tweeters	67,719
Max. # of Tweets per Tweeter	2,200
Min. # of Tweets per Tweeter	1
% Tweeters with $\geq$ 1000 Tweets	0.42%
% Tweeters with $\geq$ 500 Tweets	7.31%
% Tweeters with $\geq$ 200 Tweets	7.93%
% Tweeters with $\geq$ 50 Tweets	22.23%
% Tweeters with $\geq$ 10 Tweets	95.6%
% Tweeters with 1 Tweet	0.53%

Table II
DARKNET FORUMS DATASET STATISTICS

Parameter	Nucleus	Evolution
Total # of Posts	98,879	493,688
Total # of Authors	7,688	21,991
Max. # of Posts per Author	2,516	3,608
Min. # of Posts per Author	1	1
% Authors with >= 1000 Posts	0.04%	0.16%
% Authors with >= 500 Posts	0.22%	0.56%
% Authors with >= 200 Posts	1.04%	2.06%
% Authors with >= 50 Posts	5.03%	8.88%
% Authors with >= 10 Posts	18.86%	29.12%
% Authors with 1 Post	47.83%	27.73%

1) Twitter dataset: We make use of the CIKM dataset collected by Cheng et al. [10] and spans a period of 6 months from September 2009 to January 2010. We merged the training and test set, and after preprocessing (explained in section III-D), we were left with 67,719 Tweeters and 6,822,744 Tweets. The majority of the users posted between 10 and 50 Tweets. The details of the dataset are described in Table I.

2) Darknet forums: We scraped 13 darknet forums between February 2012 and October 2017. For this experiment we are focusing on two markets, Nucleus and Evolution. Evolution was active from January 2014 to March 2015 and Nucleus from September 2014 to August 2015. Table II details the statistics of these forums.

3) Extended Brennan-Greenstadt Corpus: The Extended Brennan-Greenstadt Corpus [1] consists of writing samples of 45 authors. Every author submitted three samples: (1) a baseline sample of at least 500 words which is 'scholarly' in nature, (2) a sample in which the author tries to obfuscate his or her writing style, (3) and a sample in which the author tries to imitate the writing style of another author. The corpus is publicly accessible and free to download.

D. Experimental setup

1) Benchmark measurements: For the benchmark measurements, we make use of the Twitter dataset. We take the following steps.

a) Creating the dataset: We create an artificial dataset by splitting n users into two separate users, u_{ai} and u_{bi}, where $1 \leq i \leq n$. We have chosen $n = 1000$, so that we can test our technique on a large scale and the users are randomly sampled. Each user u_{ai} and u_{bi} is given, without overlap, a subset of

[1] https://www.psal.cs.drexel.edu/index.php/Main_Page

all the posts of user u_i. The number of posts assigned to each user varies for each experiment between 5 and 100. The two sets, $A = \{u_{a1}, u_{a2}, ...u_{an}\}$ and $B = \{u_{b1}, u_{b2}, ...u_{bn}\}$, form the (artificial) ground truth data.

b) Data preprocessing: Similar to [28], we make sure to remove all retweets from the dataset. Since retweets capture stylistic features of other authors, it may taint the embeddings. We also change all URLs to 'URL' tags, as URLs are independent of the author's writing style.

c) Alias pairing: In the first experiment we make the assumption that it is known that every user has a second alias. We merge all accounts into a single pool and, subsequently, for every author, u_{ai} and u_{bi}, we compare their feature vectors to the feature vectors of all other authors. Then, authors with the shortest distances are paired.

d) Alias matching: Next, we remove the assumption that it is known whether users have multiple aliases. We do this by by removing half of the users from set B. This way, not all users from set A have a second account in set B. There are now two subtasks: (1) determining which users have a second alias and (2) identifying which alias this would be. This task is solved by making use of a threshold distance. Using the *alias pairing* dataset, an average distance between true aliases is computed and is used as a threshold for matching. If two accounts have a distance shorter than this threshold, the assumption is made that they can be matched. Of the list of candidate authors, it is matched with the one which has the shortest distance.

e) Varying the parameters: We measure the results for varying parameters. We vary the corpus sizes of the users, the input features, and the number of candidate authors. We also do one experiment with 5000 authors to test the technique on a very large scale.

2) Migration measurements: The second part of the experiment is centred around the user migration and activity on darknet forums.

a) Data preprocessing: After initially plotting the user activity, we noticed abnormal patterns which were caused by spammers. In an attempt to remove the spam, all duplicate posts are removed. This does not remove computer generated spam, where the contents of the messages differ for each post. Fortunately, these spam attacks happened on days which are not relevant to our analysis. Similar to the Twitter dataset, quotes within posts from different users are removed, as these would taint the embeddings of the original author. All URLs are replaced with an 'URL' tag and all PGP blocks are removed as well.

b) Measuring user activity: We start by plotting the user activity of all the forums. This is done by measuring (1) the number of posts made per day, (2) the number of users per day, and (3) the number of new users per day. This reveals when the activity goes up an down. By reading the contents of the posts around anomalous days, we can infer what causes the activities to fluctuate.

c) Measuring migrations: By looking at the number of new users per day for each forum, we can understand where large displacements are taking place. We noticed that new users typically appear on forums, when a market recently closed. For this reason, we decided to focus on market closures. More specifically, the closure of the marketplace Evolution resulted in a clear defined migration pattern in the Nucleus forums. Therefore we decided to continue with these two forums only.

d) Naive pairing: In an attempt to gather ground truth data, we link aliases between the forums based on the username. We only link aliases, if after lowercasing the usernames are the same. To avoid false positives, we do not make use of complex string-based matching techniques.

Furthermore, Decary et al. point out that the number of active dealers recover within a month of a market closure. Therefore, we only consider the influx of users in Nucleus for one month after the closure of Evolution. A user is considered to have migrated if he or she stopped being active on one forum and at least made 1 post in a new forum.

e) Alias pairing: We make the assumption that aliases paired based on the username heuristic are true aliases of the same user. This gives ground truth data to test the accuracy of *author embeddings* on real data instead of simulated data. Here, the users from the forum Evolution form the set $A = \{u_{a1}, u_{a2}, ...u_{an}\}$ and the users from Nucleus form the set $B = \{u_{b1}, u_{b2}, ...u_{bn}\}$. The task is now to pair every user from A set to a user in B. In this setting, we make the assumption that there is a one-to-one mapping and that users have not created multiple accounts on Nucleus.

f) Alias matching: The remaining users, which could not be paired using the username similarity heuristic, are now paired using a distance heuristic. It is not known which users have multiple aliases and, therefore, we measure the average distance between aliases of the same username. Subsequently, we pair the closest user between the forums, if their distance is shorter than this average.

E. Countermeasures

We analyse two manual and three automated techniques for evading linkability attacks. All measures are tested using the Twitter dataset. The best working solution is tested on the darknet forums.

1) Manual obfuscation: We test the effectiveness of manual obfuscation by trying to solve the problem of alias pairing on the Extended Brennan-Greenstadt Corpus. We create two sets of users, $A = \{u_{a1}, u_{a2}, ...u_{an}\}$ and $B = \{u_{b1}, u_{b2}, ...u_{bn}\}$, where set A contains the feature vectors of the users before obfuscation and set B contains the feature vectors of the users after obfuscation. The task is then to pair every user from set A with the correct user in set B.

2) Using synonyms: The second approach is less intensive as the first one. We attempt to evade linkability by replacing words with their synonyms. We test this on one Tweeter, by comparing whether the distance between the altered and unaltered *author embeddings* increases sufficiently to trick our algorithm.

3) Text distortion: Our results show that character level input features are effective for linkability. Therefore, we aim to evade linking, by distorting the text on character level. We apply a naive approach to this, by randomly adding or

removing a character from a given word. For each word, there is a probability x, that the word will be distorted. Distortion happens by either adding or removing a single character at random.

Then, as before, two sets of users are created. One set where the feature vectors are constructed from the original text and one set where the feature vectors are constructed from the distorted text. We measure the accuracy for varying values of x.

4) Text jumbling: Next, we look at the effect of *text jumbling* on linkability. Text jumbling was first studied by Rawlinson [27] and he showed that scrambling the middle letters of a word had little effect on the ability of a reader to still read the word. In the early 2000s, this concept gained quick popularity and has extensively been studied. There are some caveats to this concept. For example, transposing letters which are more distant from each other makes it harder to read. Also, if the sound is relatively similar, is it easier to read. Here is an example of a jumbled text:

> 'Aoccdrnig to a rscheearch at Cmabrigde Uinervtisy, it deosnt mttaer in waht oredr the ltteers in a wrod are.'

The original text is: 'According to a researcher at Cambridge University, it doesnt matter in what order the letters in a word are'. By randomising the order of the letters, we are tainting the embeddings, which should result in a decreased accuracy for any form of authorship analyses.

5) Machine translation: Lastly we analyse the effectiveness of machine translation on linkability. The idea is to translate a post to an intermediate language and subsequently translate it back to the original language [26], [9]. The reasoning is that stylistic features of the original text will be lost in the translation process.

Not all intermediate languages give a sensible result. For example, using Japanese as an intermediate language, often changes the meaning of the text. We have found that Russian works well enough to retain the original meaning, while still changing stylistic features of the text.

Using Google's translation engine, we perform a total of two tests. In the first test, we translate all the posts to Russian and then back to English. In the second test, for each post, we randomly select one of six intermediate languages: Russian, German, Dutch, French, Spanish, and Italian. This way, we diversify the stylistic features over each user.

IV. Results

A. Twitter alias pairing

In the first setting we make the assumption that it is known whether users have multiple accounts. As explained in section III-A, we are trying to pair every user in set A to a user in set B.

1) Varying corpus sizes: We start by looking at the impact on linkability for varying the number of Tweets per user. Tabel III lists the results for different corpus sizes. The embeddings are created with character 1-4 grams and word 1-2 grams, and the dimensionality is 700.

Table III
Twitter accuracy for varying corpus sizes

$A \backslash B$	**5**	**10**	**20**	**50**	**100**
100	60.6	75.8	88.9	97.1	98.7
50	56.9	70.3	82.6	94.4	
20	40.1	50.6	61.6		
10	29.5	34.3			
5	17.9				

Table IV
Twitter accuracy for different values of N

A	**n=5**	**n=4**	**n=3**	**n=2**	**n=1**
100	99.4	99.4	99.3	99.3	98.7
50	99.4	99.2	98.8	98.4	97.1
20	84.4	95.0	94.4	92.8	88.9
10	87.8	86.4	83.5	81.0	75.8
5	72.1	70.7	68.6	65.5	60.6

As expected, the embeddings are most accurate for larger corpus sizes. Impressively, it is 98.7% accurate when every user has 100 Tweets. Even with simply 50 Tweets per user, it is possible to link aliases with 94.4% accuracy. The performance decreases as the number of posts becomes sparse, with linkability being just 17.9% accurate if each account only has 5 Tweets.

2) Varying candidate authors: In real world settings, it is often enough to narrow down secondary accounts to a list of n potential candidates. For this reason, we measure the accuracy for different values of n and varying corpus sizes. The results are listed in table IV. The corpus sizes of all users in set B is 100 Tweets. The corpus sizes for users from set A are listed in the leftmost column. As expected, the accuracy increases for higher values of n. Even when the number of Tweets for set B is as low as 5, it is 72.1% accurate for $n = 5$, indicating that a relatively few numbers of posts is required for a second account to be linked to an original account.

3) Varying input features: Table V lists the accuracy for different input features and different values of n. Using all features gives the best performance, however character 2-grams and word 1-grams yield near perfect results as well. Linking with character 2-grams is 97.7% accurate for $n = 1$ and is 93.5% accurate for word 1-grams ($n = 1$). When using all features, for $n = 1$, it is 98.7% accurate.

4) Discussion: Linkability works best if users have at least 50 Tweets. As the number of Tweets goes down, the performance is strongly affected. This can be explained by the fact that a smaller corpus size often does not capture the full range of stylometric diversity of individual authors.

Table V
Accuracy for different input features

Feature	**n=5**	**n=4**	**n=3**	**n=2**	**n=1**
Char. 1-gram	82.7	81.0	79.1	73.6	65.1
Char. 2-gram	99.2	99.0	98.8	98.5	97.7
Char. 3-gram	87.0	85.9	83.8	80.9	74.7
Char. 4-gram	94.0	93.1	92.2	90.8	87.5
Word 1-gram	99.0	98.8	98.3	96.1	93.5
Word 2-gram	62.6	60.1	56.3	51.1	42.6
All features	99.4	99.4	99.3	99.3	98.7

In this case, it helps to expand the list of candidate authors. Even when one of the two accounts only has a few posts, a second account can be narrowed down to a list of 5 potential authors in more than 70.0% of the cases. The accuracy for corpus sizes of 100 and 50 Tweets are the same for higher values of n, suggesting that the advantage of the algorithm for corpus sizes above 50 becomes smaller.

Lastly, the fact that character 2-grams yield such good results has some important consequences. Namely, it is approximately 10 times faster to compute embeddings based on character 2-grams than when all features are used. This is because the neural network for computing the embeddings first constructs a Bag-of-Words as a means to parametrize the input features. Suppose that the input corpus is constructed of the letters of the English alphabet. The Bag-of-Words is then at most $26 \times 26 = 676$ tokens. With character 3-grams, this size already increases to a possible $26 \times 26 \times 26 = 17576$ tokens, considerably slowing down the training time. Character 2-grams can thus be used to compute embeddings on a large number of authors. For instance, we constructed a Twitter dataset of 5000 users and ran the same experiment using embeddings based on just character 2-grams. We made the corpus size 100 Tweets and it is 94.4% accurate. The fact that linking works on a large scale, suggests that the performance may be independent of the number of users, but rather is dependent on the stylometric diversity of the users. A last important consequence is that linkability with character 2-grams possibly becomes topic independent, hinting that it could work cross-domain.

B. Twitter alias matching

In the second experiment, it is unknown which and whether users have multiple accounts. To solve this linking task, there are several strategies we experiment with, but all of them are based off metrics computed in the previous experiment.

Initially we compute the average, minimum and maximum distance between two aliases of the same author. This is listed in the second column of table VI and the accuracy for linkability for each of these metrics is listed in the third column. The embeddings are computed using character 1-4-grams and word 1-2-grams, i.e. all input features.

a) Random: The simplest strategy is to make the assumption that every user has a second account. This strategy is 49.7% accurate, but note that in our experimental setting, where it is known that half of the users have a second account, we can at most be 50.0% accurate.

b) Maximum: The maximum distance between two accounts from the same user was 0.5691. We now make the assumption that for any account, the nearest second account belongs to the same user, *if* the distance between the two accounts is less than this value. This gives an accuracy which is close to the random strategy, because this strategy still makes the assumption that most of the users have a second account.

c) Minimum: The minimum distance between two accounts from the same user is 0.0427. We now make the assumption that for any account, the nearest second account belongs to the same user, *if* the distance between the two accounts is less than this value. This strategy, however, assumes that most of the users do not have a second account, making it exactly 50.0% accurate.

d) Average: Lastly, the most intuitive metric is to use the average distance between accounts of the same user. This value is 0.4353 and is 94.1% accurate, which is close to the accuracy in the former experiment (98.7%).

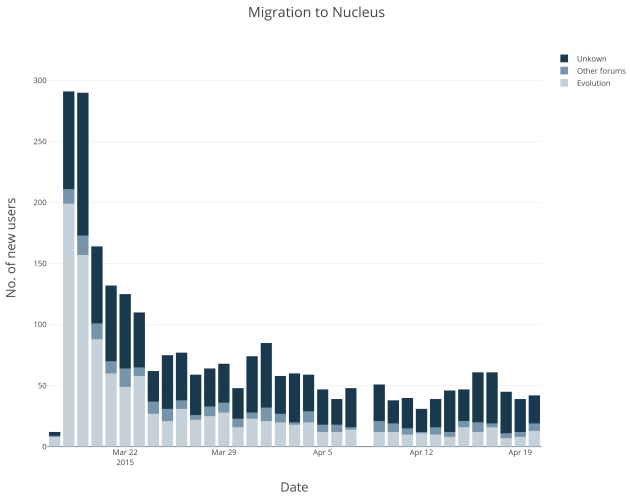


Figure 2. Nucleus migration graph. The majority of the users can be linked to Evolution and few can be linked to other forums. Approximately half of the users are unaccounted for.

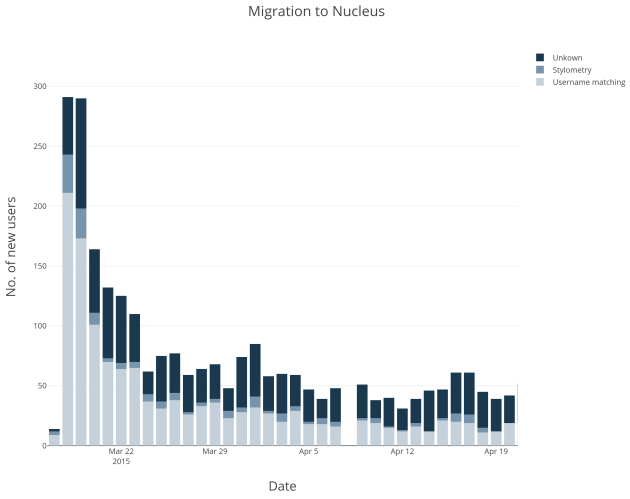


Figure 3. Nucleus migration graph after applying Stylometry.

Table VI
TWITTER ALIAS MATCHING

Metric	Value	Accuracy
Random	-	49.7
Ave. dist.	0.4353	94.1
Max. dist.	0.5691	50.3
Min. dist.	0.0427	50.0

1) Discussion: Using threshold values it is possible to accurately link users. The average distance between two true accounts is a rather simple metric. One could also experiment

Table VII
FORUM ALIAS PAIRING

Rec. size	Users	n=1	n=2	n=3	n=4	n=5
≥ 0	1496	19.2	20.9	22.3	22.9	23.2
≥ 10	426	74.9	79.6	81.9	84.3	85.9
≥ 25	214	92.1	93.9	95.8	96.3	96.3
≥ 50	121	96.7	97.5	97.5	97.5	97.5
≥ 100	52	1.0	1.0	1.0	1.0	1.0

with more sophisticated techniques for calculating threshold values, for example by first removing outliers.

The fact that this technique is accurate has the important consequence that this allows us to determine which users may be operating under multiple pseudonyms. This is also useful in other domains, for example to remove users from posting multiple reviews to promote a product.

C. Darknet forums: user activity

To measure the activity of the forums, the number of new users appearing per day is plotted for each of the thirteen forums. On some forums, there are clear defined patterns relating to activities taking place on the darknet, such as market closures and technical vulnerabilities. Of the different types of events, market closures cause the most fluctuations. After the closure of Evolution, there is a sudden burst of activity on Nucleus, as can be seen on in Figure 2. We can assume that this activity is caused by the migration of users from Evolution to Nucleus. This displacement therefore serves as a good case study.

1) Username matching: Initially, the cross activity and migration between the markets is measured using the username similarity heuristic. The colours on the migration graphs indicate from which market users are coming from. In Nucleus, 41.4%, a total of 1496 users, can be linked to *Evolution*. A few users can be linked to other markets, such as *Pandora*, *Agora* and *Black Market Reloaded*.

2) Alias pairing: Using the aliases found in the previous step, the performance of our technique is assessed for different corpus sizes. The results are listed in Table VII. The first column lists the corpus size, i.e. the minimum number of posts made by users. For example, the first row includes all users who have made ≥ 0 posts. This totals to 1496 users, for which linking is 19.2% accurate. As the corpus size increases, the accuracy also increases, which is the case with the Twitter dataset as well. The number of users with more than 25 posts is 214, for which linking is 92.1% accurate.

3) Alias matching: The results from alias pairing show that when users have more than 25 posts, they can be paired with more than 90.0% accuracy. We therefore only consider the case when users meet this threshold.

There are 2117 users in Nucleus who cannot be linked by username. Of these, 273 users have more than 25 posts. In evolution there are 1957 users who have more than 25 posts. The aim is to link accounts from Evolution to accounts from Nucleus. To this end, we make use of the metric which yielded the best performance in the alias matching problem of the Twitter dataset, namely the average distance between two true aliases. In the alias pairing of the darknet forums, this value was equal to 0.525. Using this metric, we successfully link all 273 accounts from Nucleus to an account in Evolution. This is shown in Figure 3.

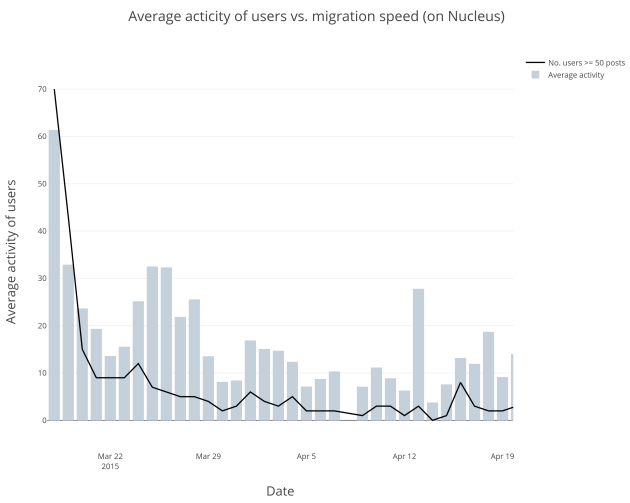


Figure 4. Average activity of the users compared to their migration speed. Users who displace fast appear to be more active on the forums.

4) Migration measurements: Even though market migrations are not the focus of this study, the displacement measurement of Evolution produces relevant findings.

In the first month after Evolution closed, a total of 2494 new users appeared on the Nucleus forums. More than half of these users appeared within the first week after the closure and 30% of the users appeared within the first 3 days. Figure 4 plots the average activity of users compared to their speed of displacement. The graph shows that more active members displaced within the first few days and the majority of the users who made more than 50 posts, migrated within the first three days. Users who posted on the forum for the first time after one month, generally tended to be less active on the forums.

Also, Decary et al. [12] showed that market activity resumes to normal after one month of a market closure. Though the majority of the users became active on Nucleus within the first few weeks, we measured a steady influx of users for at least four months after the closure of Evolution.

5) Discussion: The username similarity heuristic shows that a large number of users can be tracked after a market closure. A more thorough look at the contents of the forum posts reveals two reasons for this. Firstly, vendors rely on their reputation for sales. Using a new identity in a new market would mean that they would have to start their business from scratch. This makes it attractive for them to retain their username. Secondly, once a trade relationship has been established, buyers tend to stick their vendors. Even when vendors switch markets, buyers follow after them and even enquire on the forums whether their old vendors are still active. In some cases, buyers supposedly offer rewards for finding former vendors.

Username matching can thus be used to construct a dataset for further authorship analyses. Analysing the migration into

Table VIII
COUNTERMEASURES

Method	Dataset	No. users	Accuracy
None	Twitter	1000	92.2
None	Forums	214	92.1
Manual obfuscation	Brennan	45	0.16
Distortion ($x = .2$)	Twitter	1000	88.6
Distortion ($x = .5$)	Twitter	1000	86.8
Distortion ($x = .8$)	Twitter	1000	82.9
Distortion ($x = 1.0$)	Twitter	1000	81.4
Text jumbling	Twitter	1000	67.5
Text jumbling	Forums	214	39.4
Machine translation (Russian)	Twitter	1000	84.1
Machine translation (random)	Twitter	1000	81.7

Nucleus, reveals that author embeddings are an effective tool for linking users. Furthermore, even when ground truth data is unavailable, a large number of users can still be linked. The requirement is that users should have a minimum number of, approximately, 25 posts. This means that less active users are not easily tracked, but more active and key members would be easily linkable.

Plotting the migration activity reveals that the majority of the displacement happens within the first week after a market closure. This reinforces the view that crackdowns are perhaps not an effective tool against illegal darknet markets. Active members reappear within the first few days after a market closure. We showed that these users can be tracked using stylometry, if username matching proves unsuccessful.

D. Countermeasures

Table VIII lists the results for the different countermeasures. The Twitter dataset was used to analyse the effectiveness of different methods, after which the best performing method was tested on the Darknet forums. The first row shows that linkability is 92.2% accurate for the Twitter dataset without using any countermeasures. For Twitter, a corpus size of 50 Tweets was used and for the Darknet forums a minimum corpus size 25 posts was used.

1) Manual obfuscation: The most effective method to evade linkability attacks is by manually obfuscating one's writing style. Out of the 45 users in the Extended Brennan-Greenstadt corpus, only 7 could effectively be linked.

2) Using synonyms: Changing the posts of a user by replacing words with synonyms was not very effective. In fact, the distance between the accounts without any manipulation is 0.416 and after manipulation 0.410. The distance has thus become shorter by a very small amount, indicating that they are stylometrically more similar after altering the posts.

3) Text distortion: The effect of text distortion was tested on the Twitter dataset for varying values of the distortion rate x. For example, if x is equal to 0.2, this means that 20% of the words were distorted by either removing or adding a random letter. Distorting all the words of a user, $x = 1.0$, reduces accuracy from 92.2% to 82.1%.

4) Text jumbling: Text jumbling reduces the accuracy of linking from 92.2% to 67.5%. Because of the effectiveness of this countermeasure, it is also tested on the darknet forums. Without any countermeasures, 214 users from Evolution could be linked to Nucleus with 92.1% accuracy. By applying text jumbling, this is reduced to 39.4%, suggesting that this technique is effective for evading linkability attacks.

5) Machine translation: Machine translation reduced the accuracy to 84.1%, on the Twitter dataset, when using Russian as the intermediate language. When randomly selecting one of six languages, linkability is 81.7% accurate. The accuracy is similar to distorting the text for $x = 1.0$.

6) Discussion: At first glance, the most effective countermeasure against linkability attacks seems to be to manually obfuscate one's writing style. However, the corpus size of the obfuscated text of the Extended Brennan-Greenstadt Corpus is approximately 2,000 characters for each author. On the other hand, the corpus size for Twitter users with 50 Tweets can be $50 \times 140 = 7,000$ characters. The low accuracy for linking can thus also be explained by the fact that there is not enough data. However, even if manual obfuscation is an effective method, it might still not be a feasible method. Assuming that users are active over multiple platforms, a person who wishes to remain anonymous would have to author each post on every platform in a different writing style. This can be burdensome and difficult for users whose native language is not English. Automated techniques are thus important to evade linkage attacks.

A simple automated technique is text distortion, but it seems to be the least effective method. Even when all words are distorted, linkability is still accurate above 80.0%. Furthermore, text distortion can drastically affect readability of the text. Here is an example of an undistorted Tweet:

> "It is werid to look outside and not see palm trees and mountains"

The same Tweet which is distorted with a rate of $x = 1.0$ looks as follows:

> "It s erid o look otside and nt s pal trees nd mountains.."

Though a reader could extract the meaning from it, it is still less readable. The fact that the original messages could also contain spelling mistakes, makes it more difficult.

Text jumbling is the most effective countermeasure. The same Tweet as above in jumbled form looks as follows:

> "It is wried to look otiudse and not see plam teers and moutanins"

A reader who is fluent in English could still determine the original message. Technically speaking, this is the best method for evading linkability attacks based on author embeddings. However, there are some practical drawbacks to using this technique. Firstly, darknet forums are a tool for participants to effectively communicate with each other. Vendors use it to promote their products and establish trade relationships. Jumbling all of the posts can have an adverse effect on their dealings, as users may find it bothersome or hard to read. Secondly, not all users are native English speakers. This means that they might have extreme difficulty in reading the messages. Thirdly, software exists which can unscramble words and therefore reverse the jumbling process. Therefore, text jumbling would not be useful if attackers are tracking

individual users and have the resources to manually analyse messages. However, text jumbling could work against large scale authorship analyses, where attackers do not have the resources to analyse users individually.

Even though text jumbling may have its practical limitations, it does reveal something about how linkage could be evaded. In essence, text jumbling randomises the letters of words which are larger than three characters. This reduces the number of same occurrences of words, making frequent words appear in a diverse number of contexts. This tricks the computer into believing that an author has a very large vocabulary. Imitating this, by simply replacing words with synonyms, is not effective. In our case, it even improved linkability, suggesting that it is important to also change the word order. We think, that to apply this concept in manual obfuscation, an author would have to (1) make use of a lot of synonyms and (2) change the word order of frequently used words. As mentioned before, this can be doable, but is an arduous process.

Automated approaches to obfuscation would therefore make countermeasures more accessible to the public, but are difficult to develop. The main reason behind this, is that a computer does not inherently understand the meaning of a text like humans do. It is therefore difficult to interchange words and use synonyms, while retaining the original meaning. Machine translation does not provide the solution for this and as Potthast et al. [24] point out: machine translation is a black box and therefore does not allow us to control the outcome. Furthermore, translation techniques are constantly changing. As they are improving, machine translations will provide less stylistic changes.

All in all, evading linkage is a challenging task. If law enforcement agencies decide to perform dedicated attacks, they will be able to link the majority of the users. Automated countermeasures are not yet advanced enough to remain pseudonymous.

V. Conclusion

In this work we presented a new technique for stylometric analysis and showed that it can be used to effectively link aliases belonging to the same user. Initially, we tested it in controlled environment by artificially creating a dataset using Twitter messages. Subsequently we constructed a real dataset from a darknet user migration and showed that aliases can effectively be linked on a large scale if a user has posted enough messages. The findings show that even when ground truth data is unavailable, users can still be linked with high confidence. Lastly, we looked at countermeasures to evade our linkability algorithm. Text jumbling is a technical solution against linkability, but has some practical limitations due to social and practical factors, such as that software can easily unscramble text. However, it does tell us that linkability can be evaded by profuse use of synonyms and swapping word order. However, manual obfuscation can be tedious for users and automated obfuscation methods are not yet advanced enough.

A. Future work

Author embeddings can be used for different authorship analyses, such as author verification and document attribution. Furthermore, they can also be used for general classification tasks, such as, gender, geo-location, and age classification.

Given the good results on the Twitter dataset, it is interesting to see how well linkability performs in cross-domain settings. Formerly, this is has been a difficult task because of the diversity of the topics. However, with character 2-grams showing promise, they could solve this problem.

Lastly, we created the embeddings with the bare minimum input features. It is possible to add extra features, such as nationality, time, gender, alphanumeric frequencies, and anything else that could be relevant. On some forums, such data is readily available and could improve linkability.

References

[1] Ahmed Abbasi and Hsinchun Chen. Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. *ACM Transactions on Information Systems (TOIS)*, 26(2):7, 2008.

[2] Sadia Afroz, Aylin Caliskan Islam, Ariel Stolerman, Rachel Greenstadt, and Damon McCoy. Doppelgänger finder: Taking stylometry to the underground. In *Security and Privacy (SP), 2014 IEEE Symposium on*, pages 212–226. IEEE, 2014.

[3] Mishari Almishari, Paolo Gasti, Gene Tsudik, and Ekin Oguz. Privacy-preserving matching of community-contributed content. In *European Symposium on Research in Computer Security*, pages 443–462. Springer, 2013.

[4] Mishari Almishari, Dali Kaafar, Ekin Oguz, and Gene Tsudik. Stylometric linkability of tweets. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*, pages 205–208. ACM, 2014.

[5] Thomas Babor. *Drug policy and the public good*. Oxford University Press, 2010.

[6] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. A neural probabilistic language model. *Journal of machine learning research*, 3(Feb):1137–1155, 2003.

[7] Michael Brennan, Sadia Afroz, and Rachel Greenstadt. Adversarial stylometry: Circumventing authorship recognition to preserve privacy and anonymity. *ACM Transactions on Information and System Security (TISSEC)*, 15(3):12, 2012.

[8] Michael Robert Brennan and Rachel Greenstadt. Practical attacks against authorship recognition techniques. In *IAAI*, 2009.

[9] Aylin Caliskan and Rachel Greenstadt. Translate once, translate twice, translate thrice and attribute: Identifying authors and machine translation tools in translated text. In *Semantic Computing (ICSC), 2012 IEEE Sixth International Conference on*, pages 121–125. IEEE, 2012.

[10] Zhiyuan Cheng, James Caverlee, and Kyumin Lee. You are where you tweet: a content-based approach to geo-locating twitter users. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 759–768. ACM, 2010.

[11] Nicolas Christin. Traveling the silk road: A measurement analysis of a large anonymous online marketplace. In *Proceedings of the 22nd international conference on World Wide Web*, pages 213–224. ACM, 2013.

[12] David Décary-Hétu and Luca Giommoni. Do police crackdowns disrupt drug cryptomarkets? a longitudinal analysis of the effects of operation onymous. *Crime, Law and Social Change*, 67(1):55–75, 2017.

[13] Mark Edmunds, Michael Hough, and Norman Urquía. *Tackling local drug markets*, volume 80. Home Office Police Research Group London, 1996.

[14] Yoav Goldberg and Omer Levy. word2vec explained: Deriving mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*, 2014.

[15] David L Hoover. Frequent word sequences and statistical stylistics'. *Literary and Linguistic Computing*, 17(2):157–180, 2002.

[16] Fredrik Johansson, Lisa Kaati, and Amendra Shrestha. Detecting multiple aliases in social media. In *Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining*, pages 1004–1011. ACM, 2013.

[17] Thomas Kerr, Will Small, and Evan Wood. The public health and social impacts of drug market enforcement: A review of the evidence. *International journal of drug policy*, 16(4):210–220, 2005.

[18] Foaad Khosmood and Robert Levinson. Toward automated stylistic transformation of natural language text. *Digital Humanities, Washington, DC*, 2009.

[19] Foaad Khosmood and Robert Levinson. Automatic synonym and phrase replacement show promise for style transformation. In *2010 Ninth International Conference on Machine Learning and Applications*, pages 958–961. IEEE, 2010.

[20] Foaad Khosmood and Robert A Levinson. Automatic natural language style classification and transformation. In *BCS Corpus Profiling Workshop, London, UK*. sn, 2008.

[21] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In *International Conference on Machine Learning*, pages 1188–1196, 2014.

[22] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[23] Daniele Perito, Claude Castelluccia, Mohamed Ali Kaafar, and Pere Manils. How unique and traceable are usernames? In *International Symposium on Privacy Enhancing Technologies Symposium*, pages 1–17. Springer, 2011.

[24] Martin Potthast, Matthias Hagen, and Benno Stein. Author obfuscation: Attacking the state of the art in authorship verification. In *CLEF (Working Notes)*, pages 716–749, 2016.

[25] Tieyun Qian and Bing Liu. Identifying multiple userids of the same author. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1124–1135, 2013.

[26] Josyula R Rao, Pankaj Rohatgi, et al. Can pseudonymity really guarantee privacy? In *USENIX Security Symposium*, pages 85–96, 2000.

[27] G. E. Rawlinson. The significance of letter position in word recognition. 1976.

[28] Anderson Rocha, Walter J Scheirer, Christopher W Forstall, Thiago Cavalcante, Antonio Theophilo, Bingyu Shen, Ariadne RB Carvalho, and Efstathios Stamatatos. Authorship attribution for social media forensics. *IEEE Transactions on Information Forensics and Security*, 12(1):5–33, 2017.

[29] Xin Rong. word2vec parameter learning explained. *arXiv preprint arXiv:1411.2738*, 2014.

[30] Kyle Soska and Nicolas Christin. Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. In *USENIX Security Symposium*, pages 33–48, 2015.

[31] Martijn Spitters, Femke Klaver, Gijs Koot, and Mark van Staalduinen. Authorship analysis on dark marketplace forums. In *Intelligence and Security Informatics Conference (EISIC), 2015 European*, pages 1–8. IEEE, 2015.

[32] Martijn Spitters, Stefan Verbruggen, and Mark van Staalduinen. Towards a comprehensive insight into the thematic organization of the tor hidden services. In *Intelligence and Security Informatics Conference (JISIC), 2014 IEEE Joint*, pages 220–223. IEEE, 2014.

[33] Efstathios Stamatatos. On the robustness of authorship attribution based on character n-gram features. *JL & Pol'y*, 21:421, 2012.

[34] Gabriel Weimann. Terrorist migration to the dark web. *Perspectives on Terrorism*, 10(3), 2016.

[35] William E Winkler. String comparator metrics and enhanced decision rules in the fellegi-sunter model of record linkage. 1990.

[36] Evan Wood, Patricia M Spittal, Will Small, Thomas Kerr, Kathy Li, Robert S Hogg, Mark W Tyndall, Julio SG Montaner, and Martin T Schechter. Displacement of canada's largest public illicit drug market in response to a police crackdown. *Canadian Medical Association Journal*, 170(10):1551–1556, 2004.

[37] Li Yujian and Liu Bo. A normalized levenshtein distance metric. *IEEE transactions on pattern analysis and machine intelligence*, 29(6):1091–1095, 2007.

[38] Reza Zafarani and Huan Liu. Connecting corresponding identities across communities. *ICWSM*, 9:354–357, 2009.

[39] Rong Zheng, Jiexun Li, Hsinchun Chen, and Zan Huang. A framework for authorship identification of online messages: Writing-style features and classification techniques. *Journal of the Association for Information Science and Technology*, 57(3):378–393, 2006.