Mohamad Imad Mahaini and Shujun Li, "Cyber Security Researchers on Online Social Networks: From the Lens of the UK's ACEs-CSR on Twitter," in *Security and Privacy in Social Networks and Big Data: 9th International Symposium, SocialSec 2023, Canterbury, UK, August 14–16, 2023, Proceedings*, *Lecture Notes in Computer Science* (LNCS), volume 14097, pp. 129–148, Springer, 2023, https://doi.org/10.1007/978-981-99-5177-2_8

Cyber Security Researchers on Online Social Networks: From the Lens of the UK's ACEs-CSR on Twitter

Mohamad Imad Mahaini and Shujun Li

Institute of Cyber Security for Society (iCSS) & School of Computing, University of Kent, Canterbury, UK
{mim, S.J.Li}@kent.ac.uk

Abstract. Much work in the literature has studied different types of cyber security related users and communities on OSNs, such as activists, hacktivists, hackers, cyber criminals. A few studies also covered no-expert users who discussed cyber security related topics, however, to the best of our knowledge, none has studied activities of cyber security researchers on OSNs. This paper fills this gap using a data-driven analysis of the presence of the UK's Academic Centres of Excellence in Cyber Security Research (ACEs-CSR) on Twitter. We created machine learning classifiers to identify cyber security and research related accounts. Then, starting from 19 seed accounts of the ACEs-CSR, a social network graph of 1,817 research-related accounts that were followers or friends of at least one ACE-CSR was constructed. We conducted a comprehensive analysis of the data we collected: a social structural analysis of the social graph; a topic modelling analysis to identify the main topics discussed publicly by researchers in ACEs-CSR network, and a sentiment analysis of how researchers perceived the ACE-CSR programme and the ACEs-CSR. Our study revealed several findings: 1) graph-based analysis and community detection algorithms are useful in detecting sub-communities of researchers to help understand how they are formed and what they represent; 2) topic modelling can identify topics discussed by cyber security researchers (e.g., cyber security incidents, vulnerabilities, threats, privacy, data protection laws, cryptography, research, education, cyber conflict, and politics); and 3) sentiment analysis showed a generally positive sentiment about the ACE-CSR programme and ACEs-CSR. Our work showed the feasibility and usefulness of large-scale automated analyses of cyber security researchers on Twitter.

Keywords: Cyber Security · Machine Learning · Online Social Network · Community Detection · Natural Language Processing · Topic Modelling · Sentiment Analysis · Twitter

1 Introduction

According to a recent report [36], the active online social network (OSN) users reached 4.76 billion in January 2023, more than half of the world population.

With the popularity of OSNs among people, identifying and finding users who form different online communities has become an interesting research topic for many because studying such communities can reveal useful insights about aspects such as their memberships, people's opinions, intentions and motivations of online users' activities. Such needs have led to a wide range of social network analysis (SNA) applications for different purposes, such as maximising the diffusing of new ideas or technologies, improving recommendations, and increasing the accuracy of expert finding tasks [19].

The application of SNA is also frequently applied to study cyber security related users on OSNs, e.g., cyber criminals [2,34,15], hacktivists [14,13], activists [24], and non-experts [25,30]. However, to the best of our knowledge, no past studies have investigated cyber security researchers on OSNs using a computational data-driven approach, even though many cyber security researchers and organisations are active on OSNs, and their activities can potentially have a significant influence on other users, e.g., how non-experts learn about cyber security. This paper tries to fill this research gap. Studying cyber security researchers and organisations' activities on OSNs could help us to learn more about many aspects, such as their memberships and social structures, their connections with other users, characteristics of their members, topics they often discuss, and their perception and opinions on different cyber security related matters. A better understanding of those aspects can help us better understand how they play a role in the wider online cyber security community.

As a case study, we chose to study the research network around the 19 Academic Centres of Excellence in Cyber Security Research (ACEs-CSR) in the UK on Twitter. ACEs-CSR are UK universities jointly recognised by the National Cyber Security Centre (NCSC, part of GCHQ) and the Engineering and Physical Sciences Research Council (EPSRC, part of UKRI – UK Research and Innovation) [20]. See [20] for a list of all ACEs-CSR. These universities are a good representative subset of cyber security researchers in the UK, allowing us to test how computational data-driven analysis can be done and to have a view of the important part of the UK cyber security research community on Twitter.

The main contributions of this paper can be summarised as follows:

1. We tested the performance of the machine learning (ML) classifiers reported in [18] for detecting cyber security related accounts in a real-world setting.
2. We developed a new ML classifier to detect cyber security research related accounts with good performance.
3. Using graph-based analysis and community detection algorithms, our study showed that such methods can produce useful insights about cyber security researcher communities on Twitter.
4. Using topic modelling, we identified a wide range of topics discussed by cyber security researchers on Twitter, including some less related to cyber security.
5. By applying sentiment analysis, we observed a generally positive sentiment on the ACE-CSR programme and the ACEs-CSR.

The rest of the paper is organised as follows. Some related work is reviewed in Section 2. We explain our research questions (RQs) and the methodology we used in Section 3. Section 4 describes the data collection process used in our research. The RQ-specific details of the methodology and the corresponding results are given in Sections 5–8. Further discussions and limitations can be found in Section 9. The last section concludes the paper with future work.

2 Related Work

With the enormous content created by OSN users daily, researchers have access to a massive and wide range of individuals [1]. Different types of users can be found on OSNs, such as individuals, businesses, organisations and communities, hacktivists, and cyber criminals [24]. To the best of our knowledge, there has been no previous work on studying cyber security researchers using a data-driven approach based on OSN data. A lot of work has been done on studying cyber criminal groups on OSNs. For example, Aslan et al. [2] studied a list of 100 defacers on Twitter by analysing their activities, social structure, clusters, and public discussions on Twitter. While in [15], a clustering technique based on topic modelling was applied to study the comments of 30,469 users from three carding forums. In another study about cyber criminals [34], Tavabi et al. built and analysed a large corpus of messages across 80 deep and dark web forums to identify the discussion topics and to examine their patterns.

Moreover, several other researchers studied activist and hacktivist groups on OSNs. For instance, Jones et al. [13] analysed the presence of the Anonymous group on Twitter. They built an ML classifier and identified over 20k accounts from the Anonymous group. Then, the key players were identified using SNA and centrality measures. By applying topic modelling, the main topics were found and used to study similarities between the key accounts. Another interesting example is [24], where Nouh & Nurse studied a Facebook Activist group of 274 users with 670 posts. They created several graphs representing the users' friendships and interactions through the replies on the collected posts. Using SNA and different centrality measures, they analysed these graphs and identified the influential users. Also, sub-communities were found and studied. After that, they used sentiment analysis to study how user sentiment affected the group. Finally, they investigated trust relations using link analysis techniques.

A few studies related to analysing non-experts users on OSNs were found. In [25], Pattnaik et al. conducted a large-scale analysis on cyber security and privacy discussions of non-experts on Twitter. The researchers developed two ML classifiers, one for detecting non-expert users and the other for detecting tweets related to cyber security and privacy. Also, they used topic modelling to find the top topics discussed by non-experts. Using sentiment analysis, they discovered a general negative sentiment from non-experts when talking about such topics. Another interesting study was conducted by Saura et al. [30], where they studied cyber security related issues discussed by home users on Twitter using a large dataset of 938k tweets. They used sentiment analysis, topic modelling, and mutual information to find these security issues and studied their effects on user privacy.

Another topic related to our research in this paper is the use of ML classifiers to detect cyber security related accounts and discussions on OSNs. Aslan et al. [3] built a classifier using a small dataset of 424 manually labelled Twitter accounts to detect cyber security related accounts on Twitter and achieved good results using Random Forest and SVM classifiers. Also, in [18], we created a bigger dataset of almost 2k Twitter accounts and built a baseline classifier for cyber security related accounts and several sub-classifiers to detect other sub-groups (academics, hackers, and individuals), all with good results using several ML models.

3 Research Questions and Methodology

We found a gap in the literature about studying cyber security researchers on OSN. Thus, we wanted to explore this area, focusing on the UK ACEs-CSR network on Twitter as a case study. The main research objective is to study the cyber security researchers in the ACEs-CSR network and to see what insights can be obtained from their social structure and sub-communities on Twitter. Also, using quantitative methods (e.g., topic modelling and sentiment analysis), we analysed topics they discussed on Twitter. Thus, our research questions (RQs) for our study are:

- **RQ1**: How to identify cyber security research related accounts on Twitter?
- **RQ2**: What is the social structure of a typical cyber security research community on Twitter, such as the one formed by ACEs-CSR and their followers?
- **RQ3**: What topics do cyber security research related users in the ACEs-CSR network discuss online on Twitter?
- **RQ4**: What is the general sentiment of cyber security research related users when talking about the ACE-CSR program and the ACEs-CSR on Twitter?

RQs 1-3 depend on RQ1. To address RQ1, we used ML classifiers. Developing and evaluating such classifiers required us to collect Twitter data starting from a number of seed accounts of the ACEs-CSR (see Section 4 for more details). We studied RQ1 by i) applying two ML classifiers from the literature to detect cyber security related accounts and individual ones, and ii) building a new classifier to detect cyber security research related accounts on Twitter. For RQ2, we constructed the social graph from the connections of friends and followers of the cyber security research related accounts connected to the ACEs-CSR accounts. Then, we studied the graph's social structure and analysed different sub-communities using community detection algorithms. For RQ3, topic modelling analysis was applied using the latent Dirichlet allocation (LDA) algorithm to analyse the timelines of cyber security research related accounts to identify the main topics discussed in the ACEs-CSR network on Twitter. Finally, for RQ4, we used sentiment analysis to analyse all the tweets that mentioned any ACE-CSR account or talked about the ACE-CSR program. Then, we calculated the overall sentiment scores in each detected community from RQ2.

4 Data Collection

To study our RQs, we needed to select the right seed accounts and then crawl their friends and followers to get the needed accounts and connections between them to construct the social graph of the cyber security research related accounts in the ACEs-CSR Twitter network. The data collection for this study was carried out in June 2022. We created a list of 19 Twitter accounts, each corresponding to an ACE-CSR. First, we looked at each ACE-CSR's website and manually searched into Twitter to confirm their official Twitter account. In some cases, when no official account was identified, we chose the ACE-CSR lead's account as the seed account of the corresponding ACE-CSR. However, there was a single case when we found neither an ACE-CSR's official account nor its lead's account. In this case, we chose the account of the most well-known cyber security researcher in that ACE-CSR. Since our RQs are unrelated to the individuals themselves, but about the ACEs-CSR network as a whole, and to eliminate the risk of re-identification of individual researchers, the dataset was anonymised. To this end, this paper does not mention any personal detail related to any account, and our results do not refer to specific individuals or ACEs-CSR. This preserves individual researchers' privacy and avoids comparing individuals and ACE-CSR against each other. Note that such a treatment does not affect the reproducibility of the work presented in this paper.

For each seed account (Level 1, denoted by Lv1), we fetched its friends and followers using the Twitter API at Level 1 (i.e., Lv2). Then, we did the same for the accounts in Lv2, which led to nodes at Level 3 (i.e., Lv3). We fetched only the first 5,000 accounts (determined by the Twitter API) of friends and followers for each Lv2 and Lv3 account, as some accounts had a very large number of followers or friends. After that, we used Lv1, Lv2, and their connections. The retrieval of Lv3, which contained almost 16 million nodes, was necessary to capture all the connections between Lv2 accounts. Finally, we got 42,028 accounts in total for further analysis (19 in Lv1 and 42,009 in Lv2). Lastly, using the Twitter API, we obtained the timelines of these accounts (up to 3,250 tweets per account due to a limit of the API).

5 ML Classifiers

Studying the ACEs-CSR network on Twitter required identifying accounts that are both cyber security and research related. Thus, two classifiers were needed. Additionally, we needed a classifier to detect whether a Twitter account belongs to an individual or non-individual (e.g., group, organisation, government, NGO, news channel). Thus, a third classification task was also needed.

5.1 Cyber Security Related and Individual Classifiers

Regarding the cyber security related and individual classifiers, we used two classifiers we developed in 2021, reported in [18]. Before using these two classifiers, we re-trained and re-evaluated their performances (see Appendix A for more details). We extracted the required feature sets for our data collection as described in [18]. After that, the selected trained classifiers were used to predict the class of each account in the data collection according to each classification task. The prediction statistics are listed in Table 2. The Individual classifier was applied following the Cyber Security (Baseline) classifier to detect cyber security related individuals. Also, we applied the Individual classifier after the Research classifier – described in the next subsection – to detect whether a research related account is for an individual (e.g., researcher) or a non-individual.

5.2 Research Related Classifier

To identify cyber security research related accounts, we needed a new classifier for research related accounts. We considered a data sample as a positive case if it is involved with any research work or activity related to research. This is judged based on the account's description and timeline. This makes any cyber security researcher a positive case even if they does not work in academia or is not associated with any research organisation. This is the significant difference between our **Research** classifier and the **Academia** classifier reported in [18].

Feature Extraction: Besides the features we extracted for the Baseline and Individual classifiers, we introduced new features for this new Research classifier named the Research (**R**) group, which contains the following features. A) **Connections with seeds**, which is a metric of two values. The first is the number of seed accounts that follow this account, while the second is the number of seed accounts that this account follows. B) **Researcher Keywords**, using a compiled list of 27 keywords that can be found in the Twitter "Display Name" and "Description" fields and can refer to an account that is related to research, e.g., "Professor", "Academic", "Lecturer", "Reader", "Scientist", "Research", "Researcher", "Researching", "Research Assistant", "Research Associate", "Research Fellow", "Faculty", "University", and "PHD". These features form a 54-D vector, and each value reflects whether one of the 27 keywords appears at least once in the "Display Name" or "Description" field. C) **Verified**, which is a binary value corresponding to the **Verified** profile attribute in the Twitter account, as indicated by the blue check mark. D) **Website category**, which is derived from the "Website" field of the account's profile. Sometimes a link for a page can tell a lot about the Twitter account owner. We processed the URL found in this field and identified the host of each URL, and then used some regular expressions with manually created lists of hosts, main domains, and top-level domains to assign the parsed URL to one of the following three categories. 1) "Research": this category represents a website more likely related to research, such as a university or a research institute. Some entries used in this category's domain list are ".edu", ".ac.", ".academy", "orcid.org", and "scholar.google%". We noticed that universities do not have a unified domain in some countries. Thus, we used an additional list of university hosts [11] to capture as many cases as possible. 2) "Mixed": here, the website is not specifically related to research, but it might be. Some examples of the hosts and domains in this category are "linkedin",

Table 1: Experimental results of all the machine learning classifiers

Task	Featurs	#F	#S	Decision Tree			Random Forest			Extra Trees			Logistic Reg.			XGBoost			SVM (Linear)			SVM (RBF)		
				F1	**Prec**	**Rec**	**F1**	**Prec**	**Rec**	**F1**	**Prec**	**Rec**	**F1**	**Prec**	**Rec**	**F1**	**Prec**	**Rec**	**F1**	**Prec**	**Rec**	**F1**	**Prec**	**Rec**
Baseline	**PBCL**	149	1974	0.88	0.88	0.89	0.91	0.90	0.95	0.91	0.91	0.94	0.90	0.91	0.91	0.91	0.90	0.94	0.91	0.91	0.92	0.90	0.91	0.91
Individual	**PBCL**	149	957	0.84	0.84	0.84	0.89	0.91	0.87	0.88	0.93	0.84	0.89	0.90	0.88	0.91	0.92	0.90	0.89	0.91	0.87	0.87	0.91	0.83
Academia	**K:UCIDF**	200	245	0.81	0.68	1.00	0.90	0.82	1.00	0.92	0.85	1.00	0.00	0.00	0.00	0.82	0.69	1.00	0.00	0.00	0.00	0.43	0.71	0.58
Research	**R**	46	1003	0.78	0.94	0.67	0.81	0.94	0.72	0.81	0.94	0.71	0.82	0.97	0.72	0.81	0.94	0.72	0.82	0.97	0.72	0.83	0.96	0.73

"medium", "github", ".info", ".net" and ".com". 3) "Other": any other websites that are less likely related to research and do not fall under the previous two categories.

Classifier Training Dataset & ML Models: The training sub-dataset for this classifier was created as follows. After using the Baseline classifier to predict the labels of the 42k accounts, we kept only the accounts that were predicted as cyber security related accounts. Then, we randomly selected around 1,200 samples from the new group to label them manually. The selection and labelling process was repeated until we got a balanced dataset of 1k data samples. The same seven ML models were used for training and testing, including ET and XGBoost (see Section A). Moreover, we experimented with different feature sets to compare their performance scores and report which ones were the best for this new classifier.

Experimental Results: Using the ML Python library Scikit-Learn [26] and the above models, we experimented with the following feature set combinations: R, PR, BR, CR, PBCR, and PBCLR. All models were trained and tested with 5-fold stratified cross-validation. The testing results are shown in Table 1, where we keep only the best-performing feature sets. A colour scale from red to green was used for the F1-scores. The highest F1-score is 83% using the R, BR, CR, PBCR, PBCLR feature sets, and the SVM-R (SVM with RBF kernel), ET, and RF models. Although we wanted to select the best classifier based on the F1-score, we had to consider the **Precision** as well since it corresponds to the accuracy of the positive class (i.e., the research related account). By choosing Precision over Recall, we decided to prioritise false positives (FPs) over false negatives (FNs) since our OSN analysis required working with positive samples and inspecting their profiles, timelines and connections. Moreover, since we were studying the communities resulting from positive samples, we needed the predicted positive samples to be more accurate and the FPs to be as minimum as possible. The

Table 2: The prediction results of the used machine learning classifiers

Task	Features	Model	#(Samples)	Prediction Samples	Positive	Negative
Baseline	PBCL	RF	42,028	42,028	9,377	32,651
Individual	PBCL	RF	42,028	9,377	4,795	4,582
Research	R	SVM-R	42,028	9,377	1,684	7,693

highest Precision score is 97%. Finally, the best-performing models are SVM (RBF and Linear kernel) and LR (Logistic Regression).

Applying the Research Classifier: For the prediction of the research related accounts in our data, we selected the trained Research classifier built using the R feature set and the SVM-R model (F1-score = 83%, Precision = 96%). Since the Research classifier is also a cascaded classifier following the Baseline classifier, we only considered positive samples (9,377) predicted by the Baseline classifier as the input for this classifier. The prediction statistics are listed in Table 2. Finally, we got 1,684 positive samples and 7,693 negative samples.

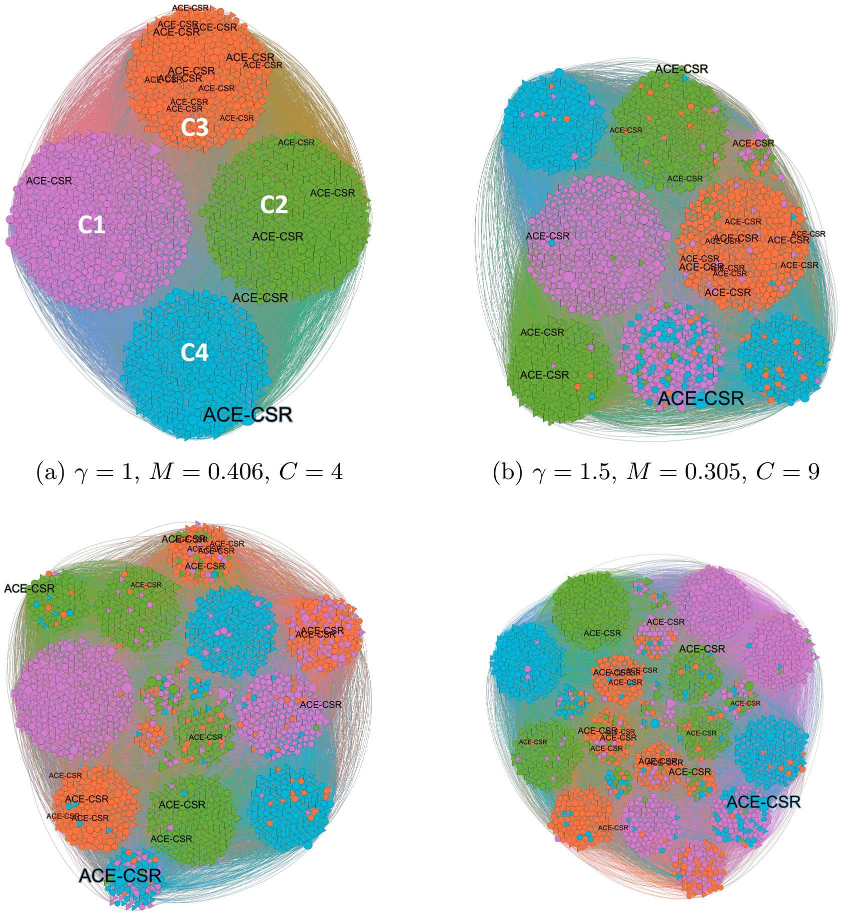
6 Social Structural Analysis

6.1 Social Graph Construction

To construct the social graph of the ACEs-CSR network, we had to identify the nodes and their edges. For nodes, we used the ML classifiers explained in Section 5 to find cyber security and research related accounts. As a result, we got 1,684 nodes, and after manual verification, some false positives were captured. Thus, the selected nodes were 1,817. For edges, we filtered the connections extracted in Section 4, where we kept only those where both ends are in selected nodes. As a result, we built a directed graph with 1,817 nodes and 64,826 edges. The constructed OSN graph was visualised using Gephi [4]. Figure 1 shows four example visualisations of the ACEs-CSR graph with different numbers of communities under different parameters. The nodes' sizes are scaled using their in-degree centrality. We can notice a few ACE-CSR nodes that are remarkably bigger than the other ACE-CSR nodes.

6.2 Communities Detection & Analysis

To study the big ACE-CSR graph, we had to break it down into sub-graphs, where each graph represents a community or a group of Twitter accounts that have something in common. A community in a network is defined by [22] as a subset of nodes that are densely connected with each other but at the same time have a few connections to other network nodes. Since the graph nodes had no ground truth labels of any characteristic, using supervised classifiers was impossible. This is normal in such cases as we do not know the number of communities and whether they are roughly equal in size when we want to break



(a) $\gamma = 1$, $M = 0.406$, $C = 4$

(b) $\gamma = 1.5$, $M = 0.305$, $C = 9$

(c) $\gamma = 2$, $M = 0.244$, $C = 18$

(d) $\gamma = 2.5$, $M = 0.206$, $C = 28$

Fig. 1: Four different visualisations of the ACEs-CSR network with different clustering parameters (C: the number of communities, M: modularity)

a network into communities [22]. As a result, we used unsupervised clustering techniques to divide the graph nodes into clusters (i.e., communities).

We tested several community detection algorithms that are widely adopted in the literature. First, we tried DBSCAN [31], but it did not work with our dataset as the clustering results were not as good as the other methods. Then, we tried the Girvan-Newman algorithm [10]. Despite the long processing time, the results were also not good as it clustered all nodes in one cluster. After that, we examined modularity-optimisation-based algorithms as modularity is a well-known method for community detection [22]. We started to get good results using the Louvain algorithm [7]. However, due to some limitations in this algorithm (e.g., yielding arbitrarily poorly connected communities), we used the Leiden

Table 3: Statistics of discovered communities ($\gamma = 1$)

Community	Colour	Members	Size	Individual Accounts	Non-individual Accounts
C1	Purple	595	32.75%	72.61%	27.39%
C2	Green	465	25.59%	79.14%	20.86%
C3	Orange	382	21.02%	51.83%	48.17%
C4	Blue	375	20.64%	70.13%	29.87%

algorithm [35] instead. These two algorithms use a resolution parameter [16], which controls the size of the detected communities [21].

Increasing the resolution parameter γ in the Leiden algorithm results in more communities while reducing it does the opposite [35]. To illustrate this, we presented four instances of applying the Leiden algorithm in Figure 1, using the following γ values: 1.0, 1.5, 2.0, and 2.5. The node size and the label are proportionate with its in-degree centrality score. Using the predicted labels from the Individual classifier in Section 5.1, the node shape can be either a triangle (individual node) or a circle (non-individual node). Also, we grouped the nodes that belong to the same cluster together using the Circle Pack [8] layout with "hierarchy" set to "cluster" attribute in Gephi. To emphasise the size and members of the clusters, we used a distinctive colour for each cluster. Then, we preserved these colours in the next applications of the Leiden algorithm to understand how these communities split and create new sub-communities when the modularity decreases due to the increase in resolution. Selecting the right resolution depends on how many communities we want to work with. Analysing hundreds of communities manually would be impossible, and analysing 2 or 3 communities would be less indicative. As for the analysis of the detected communities, we could not list all the trials we had with each reasonable resolution and its corresponding communities. Instead, we listed below a few examples of the insights we learned about the ACEs-CSR network and sub-communities we discovered shown in Figure 1.

A) Initially, we expected each ACE-CSR Twitter account to have a strong community around its node in the graph, but this was not the case for a few of them unless the modularity was significantly reduced. However, that would not reflect a strong and densely connected community. One of the explanations for this is that the seed accounts for some ACE-CSR are not well connected to other cyber security researchers. B) Some ACE-CSR nodes always appear in the same cluster regardless of the chosen resolution. After manual inspection of several cases, one explanation for this might be that these ACEs-CSR are close to each other geographically. We also had some personal observations about this, where we noticed that researchers across these ACEs-CSR have worked together. In two particular cases, some researchers moved from one ACE-CSR to another. C) Using different values for resolution and checking the resulted communities each time, we observed some clusters that do not have any ACE-CSR nodes (see Figure 1b). We inspected these communities and checked their members'

Twitter profiles. We noticed they are also densely connected and represent a mix of national, European, and international research institutions. For simplicity and explainability purposes, we carried out some additional analysis focusing only on the communities corresponding to $\gamma = 1$ (see Figure 1a and Table 3).

Clusters Analysis – Individual Members: Knowing the percentage of individuals in the ACEs-CSR network is interesting as it might give insights into how many cyber security individual researchers these ACE-CSR accounts attracted on Twitter and how many other non-individuals e.g., research centres, universities, and companies are connected to these ACE-CSR accounts. The overall individual and non-individual percentages in the graph were 69.40%, and 30.60%, respectively. Using the four communities in Figure 1a as an example, we calculated the individual percentage of each community and the results are shown in Table 3. The individual percentage reached 79.14% for Community C2, which is higher than other communities. Upon inspecting C2, we found that individuals in this community are often well-known researchers and figures in the cyber security research domain.

Clusters Analysis – Location: The account's "Location" field is optional on Twitter, so not all account holders provide such information. The percentage of the accounts with the information provided in the whole data we collected is 61.41%, while it is 77.55% for the ACEs-CSR network. This higher percentage indicates that cyber security research related accounts had a tendency to use this field more often. We analysed the ACE-CSR communities based on their members' declared locations, hoping to gain more insights into how these communities were formed in the first place or what they represent. The "location" field is a free-formatted text where users can write anything they like. We observed names of places (e.g., towns, cities, countries, or even non-existing places), names of affiliations, GPS coordinates, postcodes, country codes (alphabetic such as "GB" and numeric such as "+44"), and Unicode symbols of national flags. Considering the different ways to indicate location information, we had to use a set of methods to extract such information. For some "location" fields representing the location information as GPS coordinates, country codes and national flag symbols, we could extract such information using bespoke Python scripts. For other "location" fields that could not be processed using the previous method, we preprocessed them by removing any email address(es), URL(s), Twitter handle(s), special ASCII character(s), IP address(es)[1] and isolated number(s), and then fed them to the Location Tagger Python library [33] to extract possible location information. The extracted location information was automatically checked against cities' names downloaded from the GeoNames website [9] to resolve the ambiguity that is usually raised when detecting location information from free-formatted texts. For about 10% of "location" fields, the above automated methods could not produce any location information, so we manually inspected them to detect and recover such information. Based on all the extracted location information, we calculated geographical statistics about the

[1] IP addresses can sometimes carry location-related information. We considered such information less reliable and too complicated to process, so decided to exclude it.

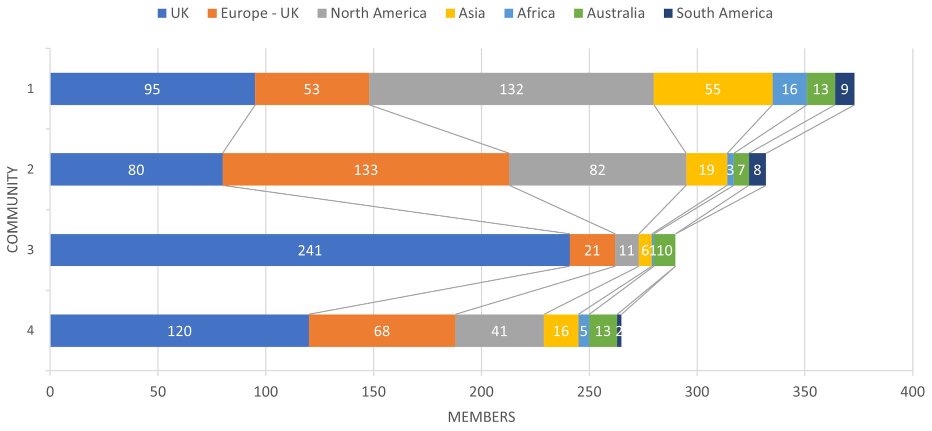


Fig. 2: Continent-specific statistics of the four communities shown in Figure 1a

nodes in the ACEs-CSR network. Figure 2 shows continent-specific statistics of the four communities shown in Figure 1a. We split Europe into two sub-groups, UK and Europe excluding UK, in order to know which communities are more national (UK) or international (non-UK) from the perspective of ACEs-CSR.

The location-based analysis revealed interesting insights about the discovered communities. First, for the four communities in Figure 2, Community C3 seems a more UK-centric one, but the other three are highly international. Communities C1 and C2 are dominated by non-UK accounts – the most accounts were from North America for C1 and from the non-UK part of Europe for C2. Second, across all communities, there are much fewer accounts from Africa, Australia and South America, indicating more biased international connections with Europe, North America and Asia. Third, Community C1 seems to be the most international cluster, where almost an equal number of accounts were from Europe (excluding the UK) and from Asia. The percentage of Asian accounts in C1 is substantially higher than the other three communities, indicating it may be the one representing the UK-Asia links. Finally, when considering UK against non-UK accounts, Community C4 looks like a more balanced cluster with an approximately 1:1 ratio between national and international accounts.

7 Topic Modelling Analysis

We utilised topic modelling to automatically identify topics discussed by the cyber security research related accounts in the ACEs-CSR network. We used the LDA algorithm [6], one of the most widely used topic modelling algorithms in the literature [25,2]. LDA is an unsupervised method for clustering N documents into k categories, i.e., topics. LDA assigns a document to a topic in a probabilistic manner, where each document is assigned to each topic with a probability, and the sum of all these probabilities is 1.0 per document [15]. The LDA algorithm

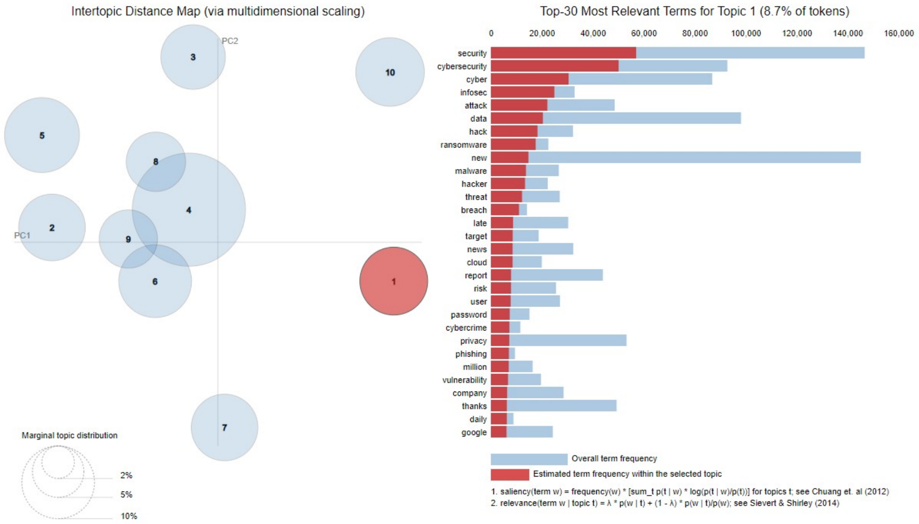


Fig. 3: Visualisation of the estimated topics by the LDA algorithm

works in iterations to do two estimations, the distribution of words (i.e., tokens) into topics, and the distribution of topics over documents [5]. Thus, it requires two essential parameters to work, which are k, the number of topics, and r, the maximum number of iterations.

We used the Scikit-Learn implementation of LDA to process the documents in our dataset, which are timelines of the cyber security research related accounts. Although there are 1,817 accounts, only 1,771 have public timelines. The timelines were preprocessed as follows:

- URLs, emails, Twitter handlers and the beginning word “RT” were removed.
- The text was tokenised using the Gensim library [27].
- Punctuation marks, isolated numbers, and very short tokens were removed.
- Stopwords removal using a list of Gensim and NLTK [23] stopwords.
- Lemmatisation was then applied using the TextBlob library [17].

After that, the tokens were passed to the LDA algorithm. We tried to find the optimum values for the LDA parameters automatically by training the LDA model using a series of values for each parameter. Each time, we used the coherence model from the Gensim library [28] to calculate the UCI coherence score of the created topics [29]. Ultimately, we chose the best value of each parameter that corresponds to the highest coherence score. For k, the tested values were from 2 to 20 with a step size of 1. The potential best values are 5 and 12. For r, the values were from 20 to 300 with a step size of 20. The potential best values are 200 and 220. While several past studies in the literature utilised coherence measures in similar experiments to find the best values for k [25,13], several other studies agreed that a manual inspection approach for the topics in each cycle is

Table 4: LDA topics with top 15 keywords, ranked in descending order by size

ID	Topic Name	Size (%)	Top Keywords
4	General Terms	24.2	like, people, think, time, good, work, know, need, look, year, thing, day, great, want, way
5	Cyber Security for Students	10.6	student, today, great, day, new, cyber, work, look, event, research, talk, join, team, uk, year
6	Data Protection Laws	10	data, privacy, law, new, right, digital, eu, ai, internet, tech, work, protection, facebook, online, gdpr
10	Vulnerabilities & Threats	8.9	new, security, malware, attack, tool, vulnerability, release, exploit, code, hack, blog, use, android, linux, update
1	Cyber Security Incidents	8.7	security, cybersecurity, cyber, infosec, attack, data, hack, ransomware, new, malware, hacker, threat, breach, late, target
2	Security Research & Education	8.4	research, new, work, security, social, read, join, look, digital, data, online, study, report, project, researcher
7	Cyber Conflict & Politics	8.4	cyber, state, russia, new, russian, china, war, ukraine, government, attack, world, country, intelligence, military, report
3	Cryptography & Privacy Research	7.9	paper, security, work, research, new, privacy, talk, crypto, open, program, phd, bitcoin, student, computer, blockchain
8	Cyber Security Events	6.6	cybersecurity, security, cyber, join, learn, new, register, ic, today, check, day, event, talk, team, course
9	ICT Industry	6.4	ai, iot, technology, data, learn, new, business, tech, future, digital, market, innovation, report, industry, world

better to find the best values of these parameters [15,2], which was confirmed in our case as well. Considering the coherence model, the manual inspection, and the visualisation-aid analysis (using the pyLDAvis Python library [32]), we set k to 10 and r to 200.

The results in Table 4 demonstrate the topics discussed by cyber security research related accounts in the ACEs-CSR network. Using the inter-topic distance map (shown in Figure 3), we can notice that the correlation between topics is minimum, which was caused mainly by topic T4, a topic with general keywords and non-related to the cyber security domain. This kind of topic is expected to be found in similar textual sources like tweets. The topic distribution is shown in Table 4. Apart from topic T4, all the other topics are relatively balanced in size, ranging from 6.4% to 10.6% with an average of 8.4%. We can spot several topical themes by looking at the generated topics: research, privacy, education, technical, and politics. Ignoring T4, the top discussed topic was T5 ("Cyber Security for Students", 10.6%), followed by T6 ("Data Protection Laws", 10%), T10 ("Cyber Security Vulnerabilities & Threats", 8.9%), T1 ("Cyber Security Incidents", 8.7%), and T2 ("Security Research & Education", 8.4%). Interestingly, politics-related and cyber conflict discussions in T7 also had a good share with 8.4%. Upon checking some tweets, we noticed sub-topics that many researchers discussed within politics, e.g., the Russia-Ukraine cyber conflict and the Trump elections. Finally, by checking the document-topic matrix, we found that the top two main topics across all documents are T5 and T3.

8 Sentiment Analysis

For RQ4, we utilised sentiment analysis to achieve a better understanding of how the cyber security research community perceive the ACE-CSR programme

Table 5: Sentiment analysis results for tweets related to ACEs-CSR

		Positive		Neutral		Negative	
Accounts Group	Tweets	Count	%	Count	%	Count	%
Non Research related	13,915	9,306	66.88	3,377	24.27	1,232	8.85
Research related C1	608	406	66.78	134	22.04	68	11.18
Research related C2	1,613	988	61.25	459	28.46	166	10.29
Research related C3	4,485	2,888	64.39	1,205	26.87	392	8.74
Research related C4	753	476	63.21	188	24.97	89	11.82
All accounts	21,374	14,064	65.8	5,363	25.09	1,947	9.11

and the ACEs-CSR. The ACE-CSR programme started almost a decade ago, and such an analysis can provide useful insights about what to do in the future with the ACE-CSR programme. To this end, we created a dataset of tweets by filtering the timelines of the 42,028 accounts in our dataset, searching for tweets related to the ACE-CSR program or any of the ACEs-CSR using a set of selected keywords. Moreover, we added tweets that mentioned any of the 19 seed accounts we used, as such mentions were considered direct or indirect interactions with an ACE-CSR. Finally, we excluded tweets created by the seed accounts as these accounts might be biased when they talked about the ACE-CSR program or themselves. In the end, a total of 21,374 tweets were obtained for the sentiment analysis. The tweets were preprocessed by removing Twitter handlers, URLs, email addresses, and the beginning word "RT" (for retweets).

We examined the two most popular methods for sentiment analysis. The first one we tried is the sentiment analyser in TextBlob [17], a popular Python library for text processing and NLP tasks. TextBlob relies on a lexicon-based sentiment analyser with predefined rules to calculate a "polarity" score between -1 and 1. This score tells whether a text can be considered positive, neutral, or negative. The second method we tried is VADER, a lexicon-based sentiment analyser with a simple rule-based model for general sentiment analysis [12]. The VADER sentiment analyser returns four scores for each piece of input text: "neg", "neu", "pos", and "compound". Each score corresponds to a sentiment type except the last which is a normalised combined value of the first three scores. For the actual implementation of VADER, we used the one in the NLTK library [23]. After applying both sentiment analysers to our data and manually inspecting the results, we concluded that VADER is a better method. Some example tweets wrongly by the TextBlob sentiment analyser can be found in Appendix B.

The results of the VADER sentiment analyser are shown in Table 5. 65.8% of all tweets are classified as positive, 25.09% as neutral, and only 9.11% as negative. These results showed that the cyber security research community perceived the ACE-CSR program and the ACEs-CSR largely positively on Twitter. Following our community analysis discussed earlier, we were also interested in if the sentiment analysis results would vary from one community to another, and between cyber security research related accounts and others in the ACEs-CSR network. To this end, we divided the tweets we selected into sub-datasets, each corresponding to an intended sub-group of accounts.

The sentiment analysis results of each sub-group are largely aligned with the main results for all. However, a few observations were noted, e.g., the percentage of the positive sentiment in Community C2 (the more "European" community) dropped to 61.25% while the negative percentage increased to 10.29%. On the other hand, the more UK-centric Community C3 saw the lowest negative sentiment percentage (8.74%) across the four communities, while the positive sentiment percentage was 64.39%. Comparing the sentiment results of Communities C2 and C3, one may wonder if the accounts' characteristics – e.g., location – can affect the results. One explanation for this observation is that UK-based accounts may be more interested in the ACE-CSR program than those European accounts outside of the UK.

9 Limitations and Future Work

The work presented in this paper has some limitations, but also suggests some future research directions. Our choice of ACEs-CSR in the UK can be seen as a very ad hoc one, but the methods we used can be easily applied to study other OSNs of cyber security researchers, other researcher communities in different research areas and disciplines, or even non-researcher communities. The performance of our Research classifier has an F1-score of 83%, which can be further improved by considering more candidate features and building a bigger dataset so that other hybrid ML models can be used, such as deep learning based ones. Our work is based on a single OSN platform (Twitter), so another future research direction is to consider other data sources to enlarge the diversity and richness of the data, such as LinkedIn and the websites of universities and research organisations. Considering a wider range of data sources will allow covering a more representative subset of the targeted research community and their online activities. Furthermore, we can also consider using scientific data services such as Google Scholar, ResearchGate and DBLP to explore potential correlations between online activities and scientific ones of researchers, e.g., if and how an enhanced level of presence on OSNs can have a positive or negative impact on the dissemination and use of the research work of a researcher or a research organisation, how topics discussed by researchers on OSNs correlate with topics of their research publications and research projects, and how researchers with similar research interests are connected on OSNs and how such connections correlate to their actual scientific or professional collaboration.

10 Conclusion

This paper reports our study on the presence of cyber security experts on OSNs, focusing on the UK's ACEs-CSR network on Twitter as a case study. We used two existing ML classifiers in the literature and developed a new one to help identify cyber security research related accounts for constructing an ACEs-CSR network on Twitter. The results showed that all the classifiers worked well for the case study. Based on the constructed ACEs-CSR network, we conducted a social structure analysis of the ACEs-CSR graph, topic modelling analyses, and sentiment analyses. The social structure analysis revealed some useful insights about the network's structure and sub-communities, e.g., a location-based analysis led to the discovery of a four-community structure: International, European, UK-centric, and balanced. The topic modelling analysis revealed a wide range of topics cyber security researchers of the ACEs-CSR network discussed on Twitter, e.g., cyber security incidents, system vulnerabilities, cyber threats, industry, data protection laws, and even politics and cyber conflicts. The sentiment analysis results showed that the accounts in the ACEs-CSR network talked about the ACE-CSR program and the ACEs-CSR mostly positively. Overall, our study has demonstrated the feasibility and usefulness of a largely automated data-driven approach for analysing cyber security research networks on OSNs.

References

1. Andreotta, M., Nugroho, R., Hurlstone, M.J., Boschetti, F., Farrell, S., Walker, I., Paris, C.: Analyzing social media data: A mixed-methods framework combining computational and qualitative text analysis. Behavior Research Methods **51**, 1766–1781 (2019). https://doi.org/10.3758/s13428-019-01202-8
2. Aslan, C.B., Li, S., Celebi, F.V., Tian, H.: The world of defacers: Looking through the lens of their activities on Twitter. IEEE Access **8**, 204132–204143 (2020). https://doi.org/10.1109/ACCESS.2020.3037015
3. Aslan, c.B., Belen Sağlam, R., Li, S.: Automatic detection of cyber security related accounts on online social networks: Twitter as an example. In: Proceedings of the 9th International Conference on Social Media and Society. pp. 236–240. ACM (2018). https://doi.org/10.1145/3217804.3217919
4. Bastian, M., Heymann, S., Jacomy, M.: Gephi: An open source software for exploring and manipulating networks. Proceedings of the International AAAI Conference on Web and Social Media **3**(1), 361–362 (2009). https://doi.org/10.1609/icwsm.v3i1.13937
5. Blei, D.M.: Probabilistic topic models. Communications of the ACM **55**(4), 77–84 (2012). https://doi.org/10.1145/2133806.2133826
6. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. Journal of Machine Learning Research **3**, 993–1022 (2003), https://www.jmlr.org/papers/v3/blei03a.html
7. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment **2008**(10), P10008:1–P10008:12 (2008). https://doi.org/10.1088/1742-5468/2008/10/p10008
8. Bostock, M.: d3-hierarchy: 2D layout algorithms for visualizing hierarchical data (2022), https://github.com/d3/d3-hierarchy
9. GeoNames: Cities (2022), http://www.geonames.org/
10. Girvan, M., Newman, M.E.: Community structure in social and biological networks. Proceedings of the National Academy of Sciences (PNAS) **99**(12), 7821–7826 (2002). https://doi.org/10.1073/pnas.122653799
11. Hipo: University domains (2022), github.com/Hipo/university-domains-list

12. Hutto, C.J., Gilbert, E.: VADER: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media **8**(1), 216–225 (2014). https://doi.org/10.1609/icwsm.v8i1.14550
13. Jones, K., Nurse, J.R.C., Li, S.: Behind the mask: A computational study of Anonymous' presence on Twitter. Proceedings of the International AAAI Conference on Web and Social Media **14**(1), 327–338 (2020). https://doi.org/10.1609/icwsm.v14i1.7303
14. Jones, K., Nurse, J.R.C., Li, S.: Out of the shadows: Analyzing Anonymous' Twitter resurgence during the 2020 Black Lives Matter protests. Proceedings of the International Conference on Web and Social Media **16**(1), 417–428 (2022). https://doi.org/10.1609/icwsm.v16i1.19303
15. Kigerl, A.: Profiling cybercriminals: Topic model clustering of carding forum member comment histories. Social Science Computer Review **36**(5), 591–609 (2018). https://doi.org/10.1177/0894439317730296
16. Lambiotte, R., Delvenne, J.C., Barahona, M.: Random walks, Markov processes and the multiscale modular organization of complex networks. IEEE Transactions on Network Science and Engineering **1**(2), 76–90 (2014). https://doi.org/10.1109/tnse.2015.2391998
17. Loria, S.: TextBlob: Simplified text processing (2022), https://textblob.readthedocs.io/en/dev/
18. Mahaini, M.I., Li, S.: Detecting cyber security related Twitter accounts and different sub-groups: A multi-classifier approach. In: Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. pp. 599–606. ACM (11 2021). https://doi.org/10.1145/3487351.3492716
19. Moscato, V., Sperlì, G.: A survey about community detection over on-line social and heterogeneous information networks. Knowledge-Based Systems **224**, 107112:1–107112:13 (2021). https://doi.org/10.1016/j.knosys.2021.107112
20. National Cyber Security Centre (NCSC), UK: Academic Centres of Excellence in Cyber Security Research (2019), https://www.ncsc.gov.uk/information/academic-centres-excellence-cyber-security-research
21. Newman, M.E.: Equivalence between modularity optimization and maximum likelihood methods for community detection. Physical Review E **94**(5), 052315:1–052315:8 (2016). https://doi.org/10.1103/PhysRevE.94.052315
22. Newman, M.E., Girvan, M.: Finding and evaluating community structure in networks. Physical Review E **69**(2), 026113:1–026113:15 (2004). https://doi.org/10.1103/PhysRevE.69.026113
23. NLTK Team: NLTK: Natural language toolkit (2023), https://www.nltk.org/
24. Nouh, M., Nurse, J.R.C.: Identifying key-players in online activist groups on the Facebook social network. In: Proceedings of the 2015 IEEE International Conference on Data Mining Workshop. pp. 969–978. IEEE (2015). https://doi.org/10.1109/icdmw.2015.88
25. Pattnaik, N., Li, S., Nurse, J.R.C.: Perspectives of non-expert users on cyber security and privacy: An analysis of online discussions on Twitter. Computers & Security **125**, 103008:1–103008:15 (2023). https://doi.org/10.1016/j.cose.2022.103008
26. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011), https://jmlr.org/papers/v12/pedregosa11a.html

27. Řehůřek, R.: Gensim: Topic modelling for humans (2022), https://radimrehurek.com/gensim/index.html
28. Řehůřek, R., Sojka, P.: Software framework for topic modelling with large corpora. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. pp. 45–50. Elra (2010), http://is.muni.cz/publication/884893/en
29. Röder, M., Both, A., Hinneburg, A.: Exploring the space of topic coherence measures. In: Proceedings of the 8th ACM International Conference on Web Search and Data Mining. pp. 399–408. ACM (2015). https://doi.org/10.1145/2684822.2685324
30. Saura, J.R., Palacios-Marqués, D., Ribeiro-Soriano, D.: Using data mining techniques to explore security issues in smart living environments in Twitter. Computer Communications **179**, 285–295 (2021). https://doi.org/10.1016/j.comcom.2021.08.021
31. Schubert, E., Sander, J., Ester, M., Kriegel, H.P., Xu, X.: DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. ACM Transactions on Database Systems **42**(3), 19:1–19:21 (2017). https://doi.org/10.1145/3068335
32. Sievert, C., Shirley, K.: LDAvis: A method for visualizing and interpreting topics. In: Proceedings of the 2014 Workshop on Interactive Language Learning, Visualization, and Interfaces. pp. 63–70. ACL (2014). https://doi.org/10.3115/v1/W14-3110
33. Soni, K.: locationtagger (2022), https://pypi.org/project/locationtagger/
34. Tavabi, N., Bartley, N., Abeliuk, A., Soni, S., Ferrara, E., Lerman, K.: Characterizing activity on the deep and dark web. In: Companion Proceedings of the 2019 World Wide Web Conference. pp. 206–213. ACM (2019). https://doi.org/10.1145/3308560.3316502
35. Traag, V.A., Waltman, L., van Eck, N.J.: From Louvain to Leiden: guaranteeing well-connected communities. Scientific Reports **9**(1), 5233:1–5233:12 (2019). https://doi.org/10.1038/s41598-019-41695-z
36. We Are Social Inc.: DIGITAL 2023: What we learned. Special report, We Are Social Ltd (2023), https://wearesocial.com/uk/blog/2023/01/digital-2023/

A Evaluating Baseline/Individual Classifiers Performance

Classifiers Training: before using the classifiers reported in [18], we re-validated their performance with our ACEs-CSR dataset (i.e. about 42,000 Twitter accounts), which is different from the ones these classifiers were trained with originally. We utilised the same original labelled datasets and followed the same steps for the feature extraction phase from [18]. After that, we selected the best-performing feature sets according to the reported results: C, L, PBC, and PBCL (see the original study for more details on the feature sets). We re-trained the classifiers using the same original models, Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), SVM with linear kernel (SVM-L), and SVM with RBF kernel (SVM-R). To see if we could get better results, we added two more models: Extra Trees (ET) and eXtreme Gradient Boosting (XGBoost). The training process was also done using the Scikit-Learn library with 5-fold stratified cross-validation. The training results are shown in Table 1. We show only the best-performing feature sets.

Our results were similar to the original ones for the first five models. As for the ET models, we noticed a similarity in performance compared to the RF

Table 6: Re-validation results of the Baseline and Individual classifiers

Task	Samples	TP	TN	FP	FN	Acc	F1	Prec	Rec
Baseline	1,154	900	63	87	104	0.83	0.90	0.91	0.90
Individual	1,003	535	281	37	150	0.81	0.85	0.94	0.78

models. This was expected as they are quite similar methods. In some cases, the ET models performed slightly better than the RF models. The XGBoost models performed well for the Baseline classification task with the PBCL feature set, where the F1-score is 91%, similar to the RF and ET models. However, XGBoost was slightly ahead of all the other models (in terms of F1-score) using the PBCL feature set. To summarise the results, we noticed that RF and ET models performed well across all the classification tasks. As for the feature sets, we found that for both Baseline and Individual classification tasks, the PBCL feature set seemed to be a good and stable choice.

Manual Evaluation: to evaluate the performance of the trained classifiers on the prediction dataset, we had to manually verify the results by selecting a subset of Twitter accounts for each classification task and manually labelling them. After that, we compared the actual labels with the predicted labels to calculate the confusion matrix. Next, Accuracy, F1, Precision, and Recall were calculated. The results of the manual verification are shown in Table 6. For the Baseline classifier evaluation, we randomly selected 1,154 samples. The F1-score was 90%, which means a 2% drop in performance compared to the F1-score from the original training/testing results, reported in [18]. For the Individual classifier, we selected 1,003 samples, and the F1-score was 85%, representing a 5% drop in performance. However, considering the significant difference in size between the original training dataset and our prediction dataset (2k vs. 42k accounts) and the relatively small performance drop, we can confidently assert that both the Baseline and Individual classifiers are good enough for our case study.

B Issue with TextBlob Sentiment Analyser

Below are some example tweets that were wrongly classified by the TextBlob sentiment analyser as negative, while the VADER sentiment analyser classified them correctly as positive.

- *Our Academic Centre of Excellence in Cyber Security Research becomes active this week.*
- *Academic Centre of Excellence in Cyber Security Research Open Day @ucl: @uclisec hosting an open day at the ACE center November 15th #infosec #CyberSecurity.*
- *Congratulations to @UniKent @KingsCollegeLon and @cardiffuni who join @UniofOxford and 13 other UK universities as Academic Centres of Excellence in Cyber Security Research, announced recently by the National Cyber Security Centre @NCSC and @EPSRC.*