

Hacker’s Paradise: Analysing music in a cybercrime forum

Anna Talas
Computer Laboratory
University of Cambridge
Cambridge, United Kingdom
anna.talas@cl.cam.ac.uk

Alice Hutchings
Computer Laboratory
University of Cambridge
Cambridge, United Kingdom
alice.hutchings@cl.cam.ac.uk

Abstract—Underground cybercrime forums have numerous discussion boards where users interact with each other. The majority of the topics revolve around technology, but a substantial number also discuss everyday topics and interests, including music. The aim of this research is to analyse the musical content posted on a large English-language underground forum to understand what types of musical content is shared, if the lyrics glamorise cybercrime, and if those who post musical content also post more about criminal activity. We find little evidence of the glamorisation of cybercrime. However, lyrics often depict a ‘gangster’ lifestyle, including the promotion of violence. We find that users who post on music boards post significantly less criminal content elsewhere on the forum, however when broken down by crime type they are significantly more likely to post about eWhoring and trading credentials than other forum users. We evaluate the performance of Google’s Perspective API in detecting toxic content in music lyrics. We find the toxicity classifier was able to detect toxic speech to an extent, but was not particularly reliable. In exploring this further, we find a bug, in that the classifier only takes the first 501 characters as input, providing a way to evade the detection of toxic content.

Content Warning: This research paper contains example forum posts, which contain graphic/explicit language that some readers might find upsetting. Please contact the author if you would prefer to read the article with toxic content removed.

Index Terms—cybercrime, music, underground forums, natural language processing

I. INTRODUCTION

Music is an important part of many people’s daily lives, and the type of music we listen to can affect our mood and outlook on life [1]–[3]. Although previous research has explored the effect of music in different subcultures, including both extremist and criminal groups [4]–[6], very limited research has focused on this type of content on the cybercrime forum community as a whole.

Underground cybercrime forums contain many thousands of users, providing a platform for interaction, skill development, and learning [7]. While these platforms include boards that provide marketplaces and technology-related discussions, there are also many conversations that focus on news and other everyday topics [8]. Popular culture is frequently discussed, including music. Such discussions can potential increase community cohesiveness.

It has been shown in previous research that music can be used to glamorise certain lifestyles, particularly within

extremist and criminal groups. It is also theorised that one of the possible appeals of cybercrime is the glamorised lifestyle of the hacker. There is some prior research on music and cybercrime by Tade [9], and Oyenuga and Ajewole [10]. The scope in both cases was limited to songs in Nigerian pop culture. Both research papers found that cybercrime-related topics were glamorized in the music and lyrics and normalised cybercrime as a desirable career path. One limitation of this prior research is that the songs were purposively selected for their lyrics, with little understanding of what cybercrime actors listen to or create themselves.

To gain a more general understanding of the sharing of music within the wider underground cybercrime community, we took a data-driven approach using the CrimeBB dataset collected by the Cambridge Cybercrime Centre [11]. Research on the same dataset has been previously carried out by Pastrana et al. [7] to identify key actors in the underground forums and analyse their behaviour. However, they did not focus on music-related boards. To address the gap in the research, we bridge the gap between music subculture and cybercrime offending. We analysed activities on Hack Forums, the largest running and most active English-language cybercrime community. Understanding the impact of non-cybercrime-related topics, particularly music, in these forums could help us understand if they glamorise cybercrime, and if there is a relationship between the music subculture and cybercrime activities.

This paper addresses the following research questions:

- 1) What types (e.g. genre, key, topics etc) of music and lyrics are being shared?
- 2) Is music used to “glamorise” cybercrime?
- 3) How well does the Perspective API toxicity classifier perform with song lyrics shared in underground forums?
- 4) Is there a relationship between sharing music and discussing different types of cybercrime?

In addressing the above questions, we make the following contributions:

- We extract and automate the labelling of music shared within the underground forum. We find the most popular music genres are electronic and hip-hop. Previous research has found links between hip-hop music and criminal subcultures, however we believe we are the

first to find a possible connection with electronic music outside of the drug scene.

- We use Latent Dirichlet Allocation (LDA) topic modelling to explore the topics in the song lyrics shared on the forum. We find most lyrics are being shared for ‘rap battles’, however, there is very little cybercrime-related content. Instead, we find the gangster lifestyle and the promotion of violence to be the most common types of lyrics posted on the music boards.
- Using manual annotation of a sample of lyrics, we evaluate the accuracy of the Perspective API toxicity classifier [12]. Toxicity classifiers are important for working with large amounts of textual data, particularly for minimising the amount of toxic material human moderators need to review. When evaluating how well the Perspective API detects toxicity within the music lyrics, we identify a bug, which we report to Google. Despite the classifier being able to handle requests of up to 20kB in size, we find it only processes the first 501 characters. Users can therefore avoid automated detection of toxic content, often used for moderation, with the use of padding.
- Across the entire forum, we compare the cybercrime-related interactions of users who post music-related content on Hack Forums to a randomly selected matched sample. We find those who share music (which may glamorise crime or the criminal lifestyle) are significantly less likely to post content classified as being cybercrime-related. We also compare specific crime types, and find when users who post on music boards do post about crime, it is significantly more likely to be in relation to eWhoring and trading credentials.

In §II we outline the previous work relating to underground forums and prior research into music and crime. We also provide an introduction to music analysis and ways to measure toxicity. In §III we present our methods, providing an overview of our data, research ethics, and how relevant posts are extracted and analysed. Our results are presented in §IV, including both audio and textual analysis of the music files and lyrics. We outline our analysis of users’ entire post history within the underground forum to explore the relationship between involvement in the music subculture and cybercrime discussions. Our conclusions are outlined in §V.

II. BACKGROUND AND RELATED WORKS

This section provides an overview of previous research and necessary background information. We first look at previous research into underground forums and the CrimeBB dataset. We outline prior research into the relationship between crime and music. Background information relating to the automated analysis of music and classifying toxicity is then provided.

A. Cybercrime forums and pathways to cybercrime

Underground cybercrime forums contain many thousands of active users who post about a variety of different topics, many to do with technology and deviant behaviour. These forums are often a place for cyber criminals to advertise their services and

share their knowledge [13], but they have also played major roles in high-profile attacks [7]. Due to this, researchers have been interested in analysing the content of the forum posts to better understand potential threats and how to detect them [14].

Different aspects of cybercrime forums have been of interest to researchers. Pastrana et al. [7] analysed user interactions to identify the key actors and the social connections between them in order to develop tools for detecting and predicting actors who might be involved in cybercrime activities. Moreover, there has been research on the “illicit infrastructure”, exploring the way new actors get involved in cybercrime and also their pathways out of cybercrime [15]. A large amount of research has focused on the social networks within forums, such as looking at the interaction patterns between users [16], [17]. Research has also been done to analyse the different types of common topics found in the forums, such as the perception of gender and the use of misogyny in the forums [18].

Other research has focused on specific types of crime commonly found on underground forums, such as online booter services [19]. Another topic of interest has been eWhoring, which is a growing type of cybercrime where offenders simulate cybersexual encounters by impersonating (generally) young women and selling misleading sexual material for financial gain [20], [21].

A key area of research focuses on developing automated solutions for data analysis. Forum data is typically collected using web scrapers and stored in datasets like CrimeBB, which is made available to academic researchers [11]. However, as discussed by Hughes et al. [22], one of the challenges when it comes to analysing cybercrime forums is the large amount of data. Although this provides researchers with possibly unique opportunities to answer research questions, new tools and techniques are needed for at-scale analyses, which is where computer science expertise becomes valuable [22]. A variety of tools have been developed to analyse forum data at scale, typically using natural language processing (NLP) [14], [23], [24]. Other tools, such as PostCog, have also been developed to aid both technical and non-technical users in exploring and comparing data across forums [25].

One of the tools built to better understand the content on cybercrime forums is the crime type classifier built by Atondo Siu et al. [26]. The classifier was built to identify nine different crime types, including bots and malware, identity theft, and eWhoring, based on the findings of prior research and also domain knowledge. Specialised tools are required as off-the-shelf solutions tend not to perform well with the jargon and poorly formed sentence typical of underground forums. In this paper, we use Atondo Siu et al.’s classifier to compare the volume of posts relating to criminal activity made by both musical and non-musical actors.

B. Music Analysis

As discussed in the previous section, music can have a large impact on the people listening to it. Research has shown that specific factors in music cause people to feel different

emotions. Some of the most basic “building blocks” of a musical piece can have a big impact on how we perceive it.

Fernández-Sotos et al. [1] analysed the relationship between a music piece’s tempo and the emotions felt by the participants. They find a significant difference in the emotions felt at different tempos. The participants were more likely to describe slower musical (90 bpm) pieces as “sad” or “relaxing”, while the same piece at a faster tempo (120 bpm) were more likely to be associated with “happiness”, “tension” and “surprise”. With a much faster tempo (150 bpm) the other trends stayed the same, while “tension” rose even higher [1].

Another key element of music is the tonal scale used, which can be either major or minor. An experiment by Mizuno and Sugishita [27] used an MRI technique to measure participants’ responses to three categories of stimuli: major, minor and neutral. The study concluded that all patients were in agreement that words like “happy” and “cheerful” describe the major scales, “sad” or “lonely” fit the minor scale and “neither” was only used to describe the neutral stimuli, even though no mention of tonality was made during the experiment. The different tonality scales were also shown to activate different parts of the brain in an almost mirrored distribution.

Similar to NLP, describing music is a task that is generally easier for humans than computers. However, it can be time-consuming, and with increasing amounts of data to annotate, new approaches are needed. Music Information Retrieval (MIR) is an interdisciplinary field that focuses on automatically retrieving information from music. This can include both more objective features, such as identifying the key or tempo a piece is performed in and automatically transcribing sheet music from audio files, but also more subjective features, such as genres and keywords and sentiment associated with the given audio [28]–[31]. MIR techniques are nowadays widely used for many things, such as automatically recognising tracks by apps such as SoundHound and Shazam, as well as automatic categorising of large digital catalogues and developing recommendation systems on streaming platforms.

C. Music and crime

Music plays a big role in different cultures and has been an important element in many social movements. Although music is usually seen as a source of positive influence, music has been used as a propaganda tool in places like China [32] and 1990s Serbia [33], as well as historical totalitarian regimes such as Nazi Germany and Soviet Russia [34].

Numerous researchers explore the links between specific kinds of music and criminal or radical behaviour. Martinez and Selepak [5] show that “hatecore” lyrics are used by the White Power Movement as a recruiting strategy by portraying ethnic, sexual and religious minorities as easy scapegoats for the anger felt by some and encouraging them to fight back. Osmanović and Ena Kazić-Çakar [6] analyse the lyrics from one of the most famous rap duos from Bosnia and Herzegovina and also survey the local student population. Their research finds a correlation between listening to music that approves of violence and criminal activity. They believe such music could

encourage the listeners to see these activities in a positive light, and might influence them to misbehave.

Previous work by Tanner et al. [4] found youth who listened to rap music reported significantly more feelings of inequity and injustice, as well as criminal behaviour. They also found that the strengths of those relationships vary according to racial identity. However, hip-hop and rap are not the only musical genres which have potential ties to crime. Krsmanović published a paper in 2020 showing potential links between organised crime and Serbian Turbo Folk music [35]. Castillo-Villar et al. [36] analysed the “altered movement” songs popular in the drug subculture and found that it conveys the lifestyle of drug traffickers as a source of power, prestige, status, and wealth. They mainly identified four different key topics: “from poor to rich”, “power through violence”, “lavish lifestyle”, and “power over women”. There has also been previous research exploring the connection between electronic music and the drug scene [37].

On the other hand, there have also been attempts to use music as a means to deter crime. Specifically, the use of classical music has been used in Australia, Britain, Canada and the United States to deter crime by making places like shopping centres less appealing to young people for social gatherings [38], [39]. It is, however, unclear whether the deterring factor of some types of music is due to the musical genre itself, such as classical music, generally seen as more calmer and more peaceful as opposed to some more aggressive genres in modern music, or the subculture associated with it, like classical music seen by many in the younger generation as “boring” and something only older people listen to. Some also argue that this type of intervention does not prevent crime but merely displaces it [38].

The research when it comes to links between music and cybercrime has been more limited. Two publications that have analysed the music found in Nigerian pop culture [9] [10]. Tade [9] concluded from analysing three songs from Nigeria’s pop culture, all of which glamourise cybercrime, that it is not portrayed as an illegal activity. Instead, the lyrics rationalise the cybercrime alternative as a way to escape poverty and gain status, and also suggest that the frauds committed are successful because they have the backing of God. Oyenuga and Ajewole [10] analysed 60 Afro-Pop songs popular in Nigerian culture. They similarly concluded that the lyrics rationalised the cybercrime alternative using God as a neutralisation technique and claiming that it brings wealth and happiness. They also claim these songs have helped popularise cybercrime-related slang among the young people in Nigeria.

D. Classifying toxicity

User-generated media has grown rapidly over recent years. Online platforms typically have rules about the type of content they permit, but with the amount of content being posted it is impossible for human reviewers alone to enforce these rules. This has created a need for automated content moderation tools, specifically for tools which are more complex than the ones that simply look for keywords. For this purpose machine

learning based toxicity classifiers have been built, which allow for automatic tagging of content with high toxicity scores. Their objective is to lessen the amount of toxic content online and help moderators by reducing exposure to hateful content. One of the most widely used toxicity classifiers is Perspective API created by Google and Jigsaw [12].

Although toxicity classifiers have improved over the years, they still face issues with inaccuracy, specifically when it comes to understanding context. Even Google states that although their classifier is meant to make content moderation easier, it is not meant as a replacement for human moderation [12]. Classifying toxic speech is complicated as it is difficult to objectively define what is toxic.

Toxicity classifiers have been shown to reflect stereotypical biases [40] held by the annotators of the training data and also to be culturally insensitive [41] due to the training data generally being representative of only one culture. Moreover, offensive words can be used in non-offensive contexts, e.g. discussing personal experiences and bringing examples, and words generally considered offensive can have non-offensive meanings in different subcultures. For example, Thiago Dias Olivia et al. [42] found the Perspective API classified Twitter posts made by drag queens as more toxic than those of white nationalists and labelled tweets using words like “gay” and “lesbian” as highly toxic even if the context they were in was positive. At the same time the classifiers have also been prone to false negatives by labelling toxic text as normal.

Research evaluating existing toxicity classifiers has also shown that they often work well on only certain types of data. Gröndahl et al. [43] showed that many classifiers only have high accuracy when tested on data that is from the same dataset as the training data and performance drops drastically when tested on other datasets. There are also numerous publications demonstrating adversarial approaches against toxicity classifiers and how they can be deceived [43], [44].

III. DESIGN AND IMPLEMENTATION

This section provides an introduction into the dataset being used, the ethical aspects involved, and the methods and tools used for analysis. §III-C explains how relevant posts are filtered out on the platform. §III-D provides an overview of the features automatically extracted from the musical tracks found. §III-E explains how we use topic modelling to understand the topics discussed in the lyrics. §III-F talks about the toxicity classifier used and the evaluation of its results. Finally, §III-G outlines how we evaluate the potential relationship between sharing music and discussing different types of cybercrime.

A. Dataset

We use a subset of the CrimeBB dataset collected by the Cambridge Cybercrime Centre [11]. The dataset includes posts from various cybercrime forums, acquired through the use of web scrapers. This analysis focuses on the Hack Forums platform, the largest and longest-running English-language underground forum, with dedicated music-related boards. CrimeBB contains data from Hack Forums starting

from the year 2007 from almost 200 different boards and has altogether over 42 million posts from more than 640 thousand users. The newest posts on the dataset at the time of the analysis performed in this research project were from 2023. Some previous work on this dataset has been outlined in §II-A.

B. Ethics

This research project was granted ethics approval from the department’s ethics committee. The CrimeBB dataset is collected from publicly available forums through the use of web scrapers, and informed consent is not requested from forum members. However, under the British Society of Criminology’s Ethics Statement informed consent may not be required for research into online communities where the data is publicly available, and the research outputs focus on collective rather than individual behaviour [45]. As detailed in §III-F, when evaluating the toxicity classifier, we manually annotated only those posts predicted to be non-toxic to minimise exposure to harmful content. We provide some example posts for illustrative purposes, however these have been paraphrased to reduce the likelihood these can be attributed back to their author, and to remove some particularly toxic content.

C. Extracting relevant posts

Due to the large number of posts from Hack Forums in the CrimeBB dataset, most of which are not music related, we first extract the music-related posts. We focus our attention towards two boards dedicated to music-related content. Although this does not necessarily rule out non-music related threads on this board, or musical content elsewhere on the forum, this enables us to narrow down the search space. We select the boards titled “Dubstep and EDM” and “Melody, Harmony, Rhythm, and MP3” as they are the only two boards out of the 197 boards on the forum that focus only on music-related content. These two boards contain a total of 540,435 posts by 96,775 users.

The goal of this research is to better understand the type of music shared within these forums and how it might affect the users. Therefore, we extract two types of content within posts on the two selected boards: links posted from music tracks, and lyrical content. The latter mostly contains self-generated rap verses, but also lyrics from popular music.

1) *Extracting lyrics*: To identify posts containing lyrics we use keywords such as “rap battles” and “lyrics” to discover relevant threads containing those keywords in the titles. From those threads we select the posts most likely to contain lyrics. While these threads are generally meant for sharing music lyrics, there are also non-relevant posts such as users commenting their thoughts on the lyrics posted or advertising services. For example, there are often posts containing tens of links to sites where users can illegally download music. Therefore, posts containing multiple links are filtered out. Very short posts are also filtered out, as these are most likely not lyrics. Lastly, some pattern-matching techniques are used, based on the way music lyrics are posted. The formatting of song lyrics generally follows some rules, such as having a pattern of lines of roughly comparable lengths separated by

line breaks. This sort of pattern is unlikely to occur when normal text is being posted. Although it is possible that the data still contains some irrelevant posts, this filtering was deemed sufficient for further analysis. Although the posts containing lyrics often also contain some lines which are not part of the lyrics, these are not removed and the posts are left in their original form. This is to ensure relevant context is retained, and due to the difficulty in automatically determining the exact beginning and end of the lyrics with high enough probability. In total, 1,888 posts containing music lyrics are identified and used for further topic analysis, as well as the evaluation of the toxicity classifier.

2) *Extracting musical tracks:* Links from SoundCloud, a music streaming platform, are extracted from all posts on the two music boards. We select SoundCloud for a number of reasons. First, the platform is popular among Hack Forums users, with many sharing links to their profiles in their post signatures. Moreover, SoundCloud’s main focus is music-related content, as opposed to sites such as YouTube, which is primarily dominated by other video content. Finally, SoundCloud is a popular platform for smaller artists or individuals to share their tracks compared to other streaming platforms, such as Spotify, which focus on already established artists. As the main aim of this analysis, as opposed to previous research that focused on tracks from pop culture, is to find trends in the type of music Hack Forums users themselves create and share, SoundCloud is preferred due to it being the platform of choice for sharing self-generated music.

Over 3,400 links to tracks on SoundCloud are found in all the posts in music boards. However, many no longer work due to the original users removing them or making them private on their SoundCloud accounts. Therefore, we filter out working links by looking at the response status codes. This results in 1,097 working links to audio tracks for analysis.

D. Music feature extraction

We use publicly available audio processing tools to identify the tempo, key and genre of the music tracks shared. Python library LibROSA [46] automatically detects the key and tempo of the audio, two basic features of all musical compositions. In music theory, the key is defined as the scale in which the music is written. The key is described using the tonic note used and whether the scale is in major or minor mode. Major and minor modes are often attributed as sounding “happy” and “sad” respectively, although this is not absolute. The tempo of the composition is measured as “beats per minute” (bpm), where a beat is the unit of time used, generally felt as an underlying “pulse” of the composition. Compositions performed in higher tempos tend to sound more upbeat and energetic, while songs in slower tempos are more likely to be described as sad or melancholic by the listener. Both of these features can be extracted with a reasonably high accuracy by analysing the spectrograms of the audio files. Although in more complex classical compositions both key and tempo can change numerous times during the piece, in popular music this

is not very common. Therefore, we only determine one key and tempo per track.

The Python library Essentia [28] identifies the genre and keywords associated with the track. Essentia provides pre-trained models aimed at performing various music information retrieval tasks. We use two models. First, a MusiCNN-based model [47] trained on the Million Song Dataset [48] identifies the most likely genre of the track from 50 different tags. Second, a model trained on the MagnaTagATune dataset [49] labels the tracks with additional keywords. The keywords predicted by the model include a wide variety of instruments as well as some more general keywords about the type of music and the presence of vocals.

Transcribing of lyrics from audio was considered, but not done for multiple reasons. First, due to the volume of tracks it is not be feasible to do manually and therefore automated tools are required. Although there are available tools for automatically transcribing audio, which have significantly improved over the recent years, correctly transcribing lyrics from musical content is much more difficult due to the additional background noise as well as changes in speech patterns and extremely flexible pitch contours [50]. Moreover, the lyrics shared are likely to contain vocabulary specific to the subcultures that are not likely to be recognised by the transcribing tools used. Therefore transcribing of lyrics would likely not result in data reliable enough to produce any meaningful results when analysed.

E. Topic analysis

We extract the lyrics from the posts to identify common words and topics. We use standard NLP pre-processing, including removing non-alphabetical characters such as punctuation, as well as removing stop-words - i.e. words that are very common in the English language and will therefore not give any significant information about the text, such as articles and conjunctions. We add some words to the stop-words such as “music” and “lyrics”. They are used in most posts, but do not provide any extra contextual information when it came to analysing the topics mentioned within the lyrics themselves. As some of the posts are in Russian, another language popular on the forum but out of scope for this project, words with Cyrillic characters are also filtered out.

We analyse the frequency of words, along with the most common bigrams and trigrams, i.e. word combinations of two or three words which appear together frequently. Finally, we use Latent Dirichlet Allocation (LDA) for topic modelling to identify groups of words which commonly appear in the same context. LDA uses unsupervised learning to cluster words into topics. This is done using publicly available Python libraries including NLTK (Natural Language Toolkit) [51], a library providing basic tools for language processing, and gensim [52], a library for topic modelling and document indexing on large corpora. The LDA model provided by the gensim library is based on a paper by Hoffman et al. [53].

F. Measuring toxicity

We analyse of the toxicity of the lyrics posted on the forum using the Perspective API [12]. The aim of the API is to aid content moderation by using machine learning to identify “toxic” comments. The API is free to use and given a text in any of the 18 languages currently supported returns a score between 0 and 1 that represents the likelihood of the comment being considered toxic by the reader. The Perspective API website [12] lists the partners who use their product, including notable names such as Reddit, The New York Times, and Wall Street Journal. We note that evaluating toxicity is hard due to the difficulty of defining what toxic content is. Google Perspective’s API has chosen to define toxicity as a “rude, disrespectful, or unreasonable comment that is likely to make someone leave the conversation” [12]. This definition, however, is somewhat subjective.

To evaluate the accuracy of the toxicity classifier in detecting toxic content in the context of lyrical content, we manually annotate of a subset of the lyrics. We only annotate lyrics with low predicted toxicity scores. The main goal is to identify if the classifier has many false negatives, where posts contained toxic content are given a low toxicity scores. The sample is not taken from posts with very high toxicity score to minimise the amount of exposure the annotators have to toxic content.

For manual annotation the posts are labelled as one of four possible categories, making it slightly more granular than simply labelling content as toxic or non-toxic. We anticipate that there will be situations where the content may not necessarily fit the two categories, when there is not enough information for the annotators to decide, or the meaning behind the post is unclear. We use the the following categories and definitions from Ousidhoum et al. [54] (taken as exact quotes so as to not change the meaning when paraphrasing):

Stereotypical. A stereotype is an overgeneralised belief about a particular social group. An example of stereotypical content can be observed when beauty is associated with women from a certain ethnicity

Insulting. A generated insulting statement can consist of a direct insult regardless of the context such as names of animals associated with social (X is a dog). Other indirect insulting statements depend on the context of the statement, such as saying that someone received a job offer because of their ethnicity, religion, or gender and not due to their abilities.

Confusing. A statement is labeled confusing when annotators cannot decide on whether the statement is problematic or not due to a lack of information. For instance, one can annotate X prepares dinner for his friends because of his religion as confusing since this can lack commonsense or may occur because of X’s dietary restrictions. However, the annotator could not decide due to the lack of context. Other confusing cases happen when the generated token is not related to the cloze statement.

Normal. When the generated content sounds normal.

The annotations were completed by two annotators who were given the post content of 120 posts. The annotators were not provided with the toxicity score given by the classifier, to

eliminate any bias. The annotators labelled each post as one of the four possible labels above when it came to the toxicity of the post. The annotations were later compared and the few exceptional disagreements were resolved, so all sample posts are marked with one label for content toxicity.

G. Cybercrime in musical and non-musical posters

We use statistical tests to evaluate if there is a relationship between sharing music and discussing different types of cybercrime. All users who had posted musical content (such as Soundcloud links or rap verses) on the two music boards were selected.

To compare cybercrime-related posting volumes, we obtain a randomly selected matched sample of non-musical posters, stratified by posting volumes. First, we sort the music posters into bins based on their overall number of posts on Hack Forums. For each bin, a random sample of the same size is picked from all other Hack Forums users who had posted the same number of times in total and posts from all these users were extracted.

To test if the amount of criminal content posted differed between the groups, all posts belonging to the selected users are automatically labelled by crime type. The classification of crime type use the predictions from Atondo Siu et al.’s [26] crime type classifier. The labels predicted by the classifier are ‘not criminal’, ‘access to systems’, ‘bots and malware’, ‘eWhoring’, ‘currency exchange’, ‘DDoS and booting’, ‘identity theft’, ‘spam’, ‘trading credentials’, and ‘VPN and hosting services’.

We use Chi-square tests of independence to test if there are differences in the type of content musical actors’ posts compared to this matched sample of other Hack Forums users. We run two tests, the first compares the posting of criminal vs non-criminal content, while the second compares the posting frequency for the different crime type categories.

IV. RESULTS

This section provides the results of our analyses. We first detail the type of music shared, both from the musical perspective as well as the most common topics mentioned. We then evaluate the results of the toxicity classifier on lyrics posted on the music boards. Finally, the relationship between posting cybercrime-related content between key actors in music-themed boards and other Hack Forums users is compared.

A. Analysis of audio tracks

For the analysis of the audio tracks, we first measure the tempos and keys as an indicator of the style of music, as well as the predicted genre and associated keywords.

1) *Musical features:* The 1,097 tracks from SoundCloud links found on the music boards are analysed using multiple musical features. The genre of each track shared is predicted using a MusiCNN-based model [47]. The overwhelming majority are classified as electronic (50.4%) or Hip-Hop (36.4%), with the rest of the genres making up only 13.2%, as shown

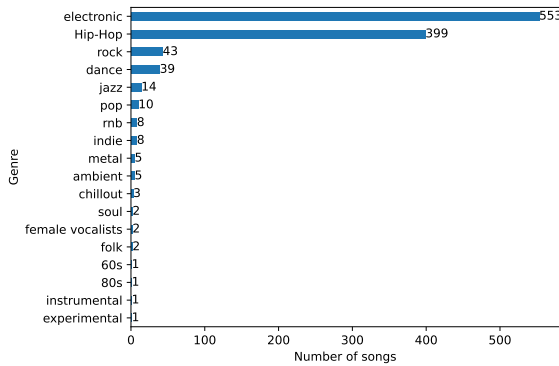


Fig. 1: Shared music by genre

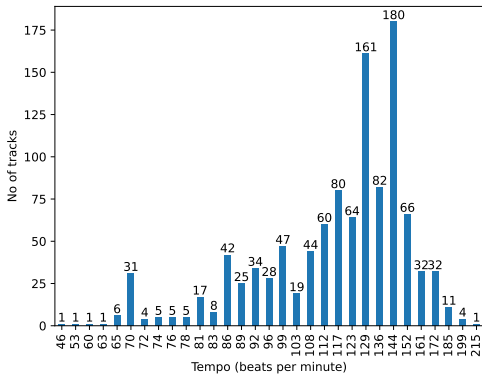


Fig. 2: Shared music by tempo

in Figure 1. This indicates a clear preference of forum users for some genres over others.

We also analyse simpler musical features such as key and tempo. Of all the tracks analysed, 785 (71.6%) are in a minor key and the remaining 312 (28.4%) are in a major key. This differs significantly from the overall popularity of keys. An analysis of over 30,000 songs from a variety of genres shows that most music is composed using major keys, with the top 3 most popular being C major, D major and G major [55]. While this might suggest that the music shared on underground forums is likely to be described as sad, electronic music is the contrary. While most electronic dance music is written in minor keys, with C minor and A minor being the most popular choices, it tends to be more upbeat [56].

Most of the tracks analysed were between 129-144 beats per minute, with a minority being slower. The distribution of the tempos of the tracks analysed can be seen in Figure 2. This confirms the findings of genres discussed below, as the typical BPM (beats per minute) for hip-hop is around 85-115 BPM, and while there is slightly more variance between different styles of electronic music, techno music is generally fast and energetic with a range of 120-160 BPM [57].

Additional keywords (with an average of 3 keywords per track) are assigned to the tracks using a classifier trained on the MagnaTagATune dataset [49]. The distribution of assigned keywords is shown in Figure 3. The classifier assigns keywords

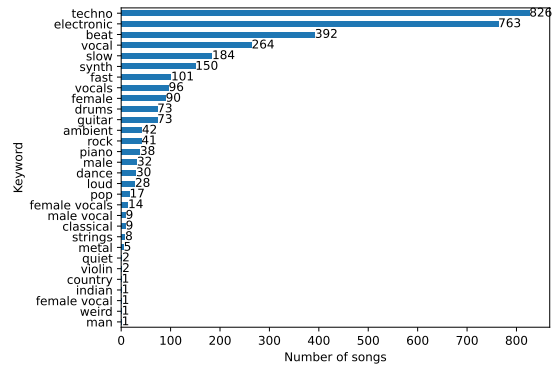


Fig. 3: Shared music by keywords (average of 3 per song)

relating to the use of instruments and vocals. We find few instruments were used in the tracks shared on the music boards. The overwhelming majority of the tracks are described by the keywords “techno” and “electronic”, with “beat” and “vocals” being the next most popular, albeit associated with only about half as many tracks.

The popularity of hip-hop music is not surprising, as it supports prior research findings relating to cybercrime and crime more generally. As discussed in §II-C, hip-hop and rap music have previously been connected to alternative subcultures and deviant behaviour [6] [4]. The same cannot be said for electronic music.

The prominence of electronic music in the data could be explained by the inclusion of the music board “Dubstep and EDM”, which specifically focuses on this type of music. The reason for a board like this to exist separately from other music genres could be related to the popularity of due to its connections to computer games, another popular pastime for users in these forums. Hack Forums has multiple boards dedicated to gaming, from general gaming boards to boards for specific games, for example, CS:GO, Overwatch, and League of Legends. However, another possible explanation for the overwhelming amount of electronic music, especially on SoundCloud, could be that this is the easiest type of music to create. Unlike most types of music, the creator does not need to know how to play any instruments or have access to any specialised technology, as a regular computer is enough to download the required software and start creating.

In conclusion, all four forms of analysis - tempo, key, genre and keywords - show that with an overwhelming majority, most of the music shared is electronic and techno music, with hip-hop music being the second most popular and other genres making up a small minority. This indicates that the musical tastes of the forum users differ from mainstream music.

B. Analysis of lyrics posted on music boards

We use machine learning to automate the analysis of lyrics posted on music boards, to identify the types of topics covered.

1) *Topic analysis:* The lyrics extracted from the forum posts are analysed to find common themes. These posts contain both original verses written by the users themselves, but also

shared lyrics from popular music. The total number of posts we analyse is 1,888. We analyse 346,954 words, making the average post roughly 184 words long. We first analyse the most common words and word combinations to appear in the lyrical text. We then present the results of the topic modelling, along with example posts from the forum.

First, we show the frequently occurring bigrams and trigrams, i.e. combinations of two and three words that are most likely to appear together. The most common bigrams, shown in Table I, are dominated by three categories. The first theme relates to music, such as “rap, battle”, “hip, hop” and also “Lil, Wayne” the name of a famous rapper. The second popular category is the use of swearwords and vulgar language. The final category is neutral language, which does not have any particular sentiment attached to it.

It is notable that there are no frequent bigrams relating to cybercrime, which would suggest that there isn’t a strong trend towards glamorising this type of deviant behaviour. However, “make, money” and “get, money” can be related to some of the motivations for engaging in financial crime.

The frequent trigrams, shown in Table II, provide different findings. The trigrams are dominated by combinations of more common phrases to appear in song lyrics, such as (baby, baby, baby). A significant number of them also relate to various emotional states, many of them sad, such as (pain, run, deeply) and (hide, behind, lie).

Bigram	Freq.
(rap, battle)	110
(hip, hop)	94
(never, say)	90
(would, think)	71
(get, shit)	70
(suck, dick)	62
(baby, baby)	57
(get, fuck)	54
(make, sure)	53
(one, day)	52
(get, bitch)	49
(say, shit)	49
(get, back)	48
(make, sense)	47
(Lil, Wayne)	48
(make, nasty)	42
(get, well)	41
(talk, shit)	41
(get, money)	40
(make, money)	39
(think, good)	39

TABLE I: Frequent bigrams in lyrical text

Although the frequent bigrams and trigrams provide some information on what types of words and phrases might describe the topics in lyrics, as can be seen in Table II, they do not provide a clear picture of the topics discussed. This method is more likely to find common lyrical combinations that might have little meaning on their own besides some emotions conveyed.

Therefore, we train an LDA model, extracting four more common topics discussed in the lyrics. The topics consist of

Trigram	Freq.
(never, say, never)	61
(baby, baby, baby)	43
(make, nasty, make)	27
(light, weight, light)	32
(tell, man, hot)	19
(hustle, hustlin’, hustle)	19
(hide, behind, lie)	19
(make, friend, amend)	16
(hustlin’, everyday, hustlin’)	14
(heaven, righteous, side)	14
(pain, run, deeply)	12
(till, end, time)	13
(bitch, son, bitch)	12
(last, day, today)	12
(ocean, run, deeply)	12
(let, burn, watch)	12
(never, make, cry)	10
(never, say, goodbye)	10

TABLE II: Frequent trigrams in lyrical text

groups of words that are commonly present in the same posts. The topics, along with the words most commonly associated with them, are shown in Table III. The titles for the topics are chosen based on examining the keywords found by the model and the associated posts.

Topic	Most commonly associated words
1) Commentary	opinion, singer, fan, artist, album, song, share
2) Gangster lifestyle	rap, bitch, fuck, shit, n*gga, weed, money
3) Violence	kill, death, deep, metal, psycho, hardcore, fear
4) Emotions	never, love, rock, god, nothing, live, blue, sad

TABLE III: Common topics in lyrics

Topic 1: Commentary. As we retain all the post contents, one of the most common topics includes words associated with user commentary on the lyrics posted. These posts usually consist of users giving their opinions on different lyrics they posted or asking for an opinion from others. An example is given below:

“Long time fan of Lucki, love his flow and quality beats. This is one of the tracks on his new album.”

Topic 2: Gangster lifestyle. The most commonly detected topic present in the actual lyrics was a representation of a “gangster lifestyle”. The lyrics generally portray topics such as making money, using drugs, and having sex with women. These lyrics are generally shared in rap battles. They frequently use vulgar language to insult others, and the users portray an over-exaggerated “tough guy” personality.

*“Uh huh, we’re cool and doing lines.
This shit so good it has you reading minds.
Say what you want about me, you fucking prick.
Ima win this shit you all can suck my dick”*
(extract from rap battle)

Many of the words from this topic can also be seen in the frequent bigrams in Table I. Over 70% of all the lyrical posts contained at least one of the most commonly associated words (as shown in Table III) with this topic, probably due to the presence of common swearwords. Around 25% of all lyrics were found to be associated with this topic.

We do find some posts classified under this topic written by users criticising these common themes in rap lyrics, labelling them as unoriginal and claiming they give the genre a bad reputation. For example, one user wrote: *“Can we remember that rap stands for rhyming and poetry, bad lyrics like, shitty lyrics like “i’ll fuck your bitch and smoke your weed” is neither.”*

Topic 3: Violence. Another common topic among the lyrics posted by the users is violence. This is the most explicit category of posts. It features users writing or sharing lyrics relating to different kinds of violence, a lot of it rather explicit and gory. The example given below is one of the less toxic examples.

*“...[username] is here to end this fight im about to hulk out and crush your spine
mothafucka im psychotic im insane think about it
take some time*

*you remember when you were talking that shit
now im here to fuck up your shit
alright now it's time to truly end this fight, get ready
to be run in 5,4,3,2,1 click-clack BOOM!
That's my gun, I'm so hard you see
murderin ya'll n*ggas this a game to me...*

(extract from a post)

Many of the posts mention brutally attacking or murdering either their rap battle opponents or others, with one poster even commenting “*I did a funny serial killer type rap let me know what you think*”. Roughly 9% of the lyrics are found to be associated with this topic. Keywords related to violence do not come up when looking at the most common bigrams and trigrams (Tables I and II) suggesting that this topic does not contain distinctive patterns of word pairs appearing together.

Topic 4: Emotions. In contrast to the vulgar and violent posts users also post lyrics about more emotional topics, such as love, relationships and feeling sad. One of those examples can be seen below:

*“Spent all my time with you
I fucking worshiped you.
Woke up each day just next to you.
Now I can't bare to look at you....”* (extract from a post)

Some of the more frequent trigrams shown in Table II are also related to this topic. Compared to the previous two topics, this one is rather humanising and generally feels a lot more sincere than the “savage” personas users seem to be putting on when writing about committing violent acts and the “gangster” persona depicted in the rap battles. Notably, one of the keywords strongly associated with this topic was “rock”, showing that these lyrics are likely mostly from songs in the “rock” genre, which is also the third most popular genre found in the audio tracks analysed, as shown in Figure 1.

2) *Mentions of cybercrime:* Explicitly cybercrime-related keywords do not come up in the analysis among the most common topics discussed. However, through targeted keyword searches, such as mentions of hacking and cybercrime-related topics, numerous examples of cybercrime-related lyrics are found. In general, mentions of cybercrime are used as a way to insult others and their hacking skills in rap battles, such as can be seen in Quote 1.

*“How you turned the hacking world into a niche
To sell ebooks and other shit you leechd
And copy pasted hacking tools that you be selling to
fools for forty dollars each”* (Quote 1)

The other common mention of cybercrime includes threatening one’s opponent in a rap battle, often insinuating that the opponent would be stupid enough to fall for their scam attempts. Very few posts were show cybercrime in a positive or glamorous light, although some examples can be found, such as Quote 2.

*“I wanna be where the hackers are
I wanna see, wanna see them crackin’*

*Posting around on those - what do you call 'em?
Oh - forums!”* (Quote 2)

C. Evaluating the Perspective API on lyrics

In this section we provide the results of our evaluations of the Perspective API toxicity classifier. In §IV-C2 we also describe a suspected bug we detect in the Perspective API, which allows users to evade their content being flagged as toxic on many sites where this popular classifier is used.

1) *Classifying toxicity:* First, we use the Perspective API to obtain toxicity scores for all lyrical posts. We find a bimodal distribution, with posts receiving either a very low or very high toxicity scores, with the main peaks being around 0.05 and 0.85 at either end respectively (see Figure 4). The Perspective API does not provide any guidance on the appropriate threshold to separate toxic and non-toxic content. Therefore, as we saw an increase in the number of posts around 0.75, we used this threshold to select posts for evaluation, to minimise our exposure to toxic material.

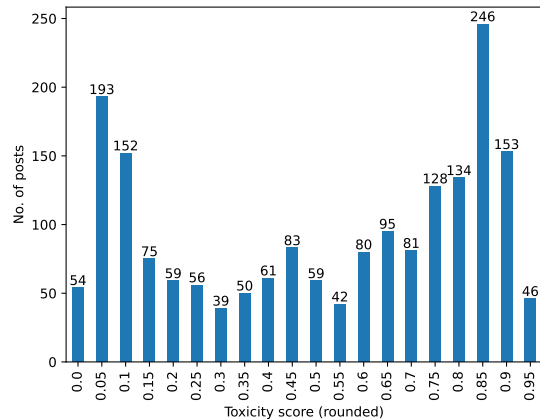


Fig. 4: Toxicity scores for all lyrical text posts

We randomly sample 120 posts for manual annotation, all with a toxicity score below 0.75. Our annotations use the four labels defined in §III-F. Of the 120 posts, 54 are labelled “normal”, 52 “insulting”, 7 “stereotypical”, and 5 “confusing”. As the toxicity scores are not normally distributed, we use the non-parametric Kruskal-Wallis H test and find a significant difference in the distribution of toxicity scores for the four labelled categories ($H = 14.5, p < .001$). Figure 5 shows posts classified as normal have much lower toxicity scores compared to the other categories, with content labelled as insulting having the highest mean and maximum toxicity score. This suggests that for the most part the classifier is correctly giving toxic content a higher score, although there is some overlap between the scores for toxic and non-toxic posts. These results indicate that 0.50 may be a suitable threshold for differentiating between toxic and non-toxic content.

2) *Evading toxicity detection:* As can be seen in Figure 5, some comments classified as insulting have significantly lower toxicity scores than expected. We investigate these outliers further to understand why they do not have higher toxicity

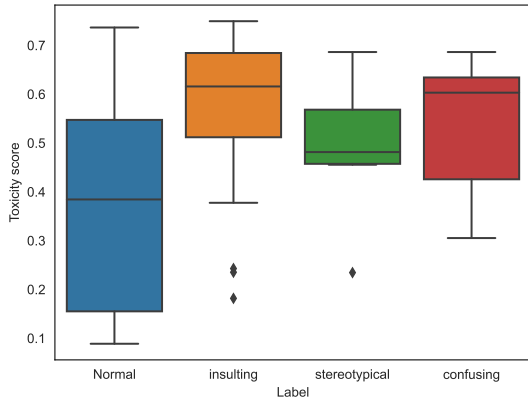


Fig. 5: Variance of toxicity scores compared to results of manual annotation

scores. The lowest score given by the classifier to any post manually annotated as insulting was 0.18. This post contains lyrics to a song with numerous swearwords and otherwise toxic vocabulary, so this very low toxicity score is surprising. Reviewing the post, we notice the user had added a few lines of their personal opinion before the lyrics:

“Any thoughts?

*My thoughts (quickly):
 Sam got way darker, (love it) and more complex,
 and Carl...well...he kinda slipped a little. I'm actually a bit disappointed in him.*

Here’s the lyrics,”

When these lines are removed from the post, the toxicity score rises to 0.63. Further testing shows that when the same lines are added in front of all the lyrics tested, the scores dropped significantly (see Figure 6).

To better understand any possible effect of padding the beginning of posts, we chose a test with non-toxic content, an English translation of the classic placeholder text “Lorem ipsum” [58]. We translate the text to English as the Perspective API does not support Latin. We manually verify the text is non-toxic, and Perspective API gives it a toxicity score of 0.09. To test the effect of adding non-toxic text to the beginning of the text to be evaluated, multiple tests are run using different lengths of the “Lorem ipsum” text at the beginning of all the samples. The results are shown in Figure 7.

It can be seen that the more characters of non-toxic text are appended to the beginning, the more the toxicity score drops. More notably, if more than 500 characters are added, the classifier ignores any text beyond it and returns a constant score independent of the toxicity of any of the following words in the same text. The constant score is also equal to the score of the text by itself. Subsequent testing showed that by choosing any arbitrary text of at least 501 characters, any further text appended to it always returns a constant toxicity score. In fact,

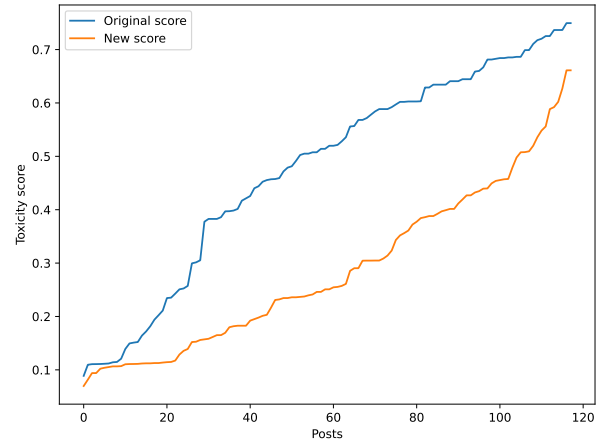


Fig. 6: Changes in toxicity scores with lines from the outlier added to the beginning

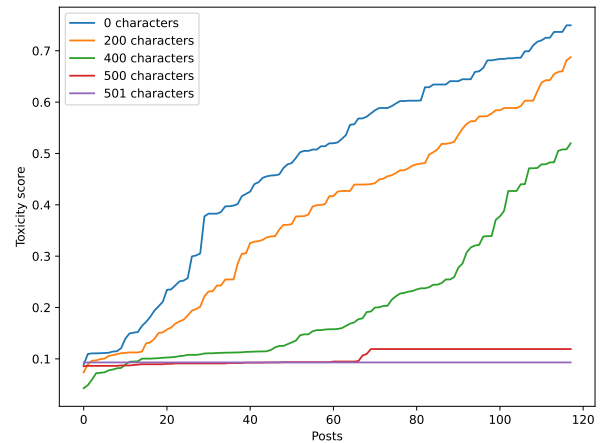


Fig. 7: Changes in toxicity scores as a set number of non-toxic text is appended to the beginning

simply adding 501 whitespaces to the beginning of any text from the sample would result in a constant toxicity score of 0.14. This suggests that the classifier only uses the first 501 characters to determine the toxicity of the given text.

At the time this paper was submitted, the Perspective API’s website claims that the model was trained on online comments and is therefore most suitable for comments of a similar length [12]. However, they also state in their documentation that the maximum size per text request is 20 kB [12]. When the maximum length is exceeded, an error stating so is also thrown. During our testing, all requests stayed under 4kB, well below the maximum size allowed. The response from the API also contained information about the span of the toxicity score, which correlated to the length of the entire sample being tested at the time. We found no publicly available disclosure

or mention of this bug at the time. Therefore, we contacted the Perspective API team and submitted a bug report.

3) *Effects on the annotated data:* In the sample data, the mean length of posts is 1,054 characters, with less than 25% of the posts being below 501 characters. This means the majority of the posts analysed with the toxicity classifier are only analysed partially. However, this becomes detectable only when the first 501 characters comprise significantly less toxic text than the latter part of the post. This is not very common in song lyrics, where there is generally a relatively consistent mood and topic from start to finish. The distribution of toxicity scores in posts with 501 characters or less as shown in Figure 8 changed quite a bit when compared to the distribution of all posts shown in Figure 4. There is still a large number of posts with very high toxicity scores but significantly fewer with lower toxicity scores. Therefore, it can be assumed that most longer posts scored, on average, lower toxicity scores.

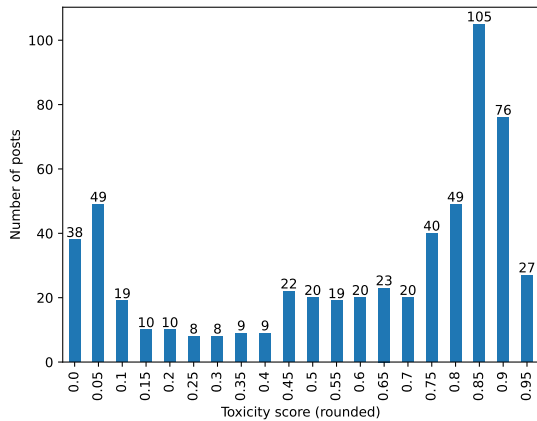


Fig. 8: Toxicity scores for all lyrical text posts with 501 characters or less

D. Cybercrime posts by musical and non-musical users

We compare posts made by users active in music forums compared to a randomly selected matched sample of other Hack Forums users, finding significant differences between the amount of cybercrime-related posts, and the types of cybercrime being discussed. To differentiate between the behaviour of users based on how active they are in the forums, we categorise users based on the total number of posts made. We use a logarithmic scale due to the different distribution of the number of users compared to the number of posts. In underground forums there tends to be many users who post only a few times and significantly fewer users who post prolifically. We find the distribution is slightly different when it comes to key actors from music boards, where most users have between 100-10,000 posts (see the comparison in Figure 9). This suggests that music boards have a core community with fewer casual “visitors” to the forum; thus, the results are likely to be more representative of the music subculture.

For the Chi-square test an equivalent number of users were taken for each of the groups to create a randomly selected strat-

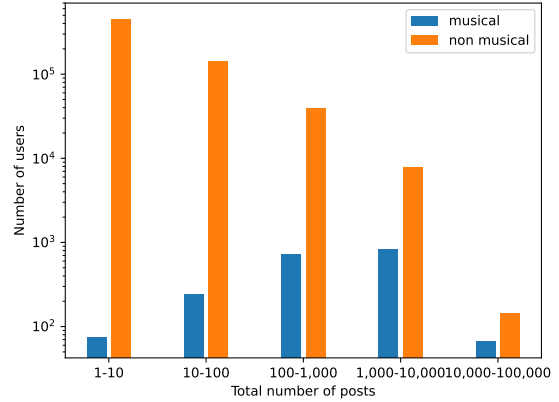


Fig. 9: Distribution of users by total number of posts made, grouped by music and non-music actors

ified sample. The Chi-square test revealed that key actors from music forums generally post significantly more “not criminal” content than their counterparts ($\chi^2(1, N = 7,070,971) = 7,967.54, p < .001$). A vast majority of the content posted on the forums is non-criminal, with criminal content making up only about 8.12% of all the posts in the sample.

We also find a statistically significant difference in the crime types the users post about ($\chi^2(8, N = 574,797) = 1,303.58, p < .001$), shown in Figure 10. Users from music boards make more posts relating to eWhoring, trading credentials, currency exchange, and spam. However, they post significantly less about the other crime types. This shows clear differences between the groups in the type of deviant behaviour they are most likely to engage in. It is not clear what may be influencing their choices, which could be investigated in future research by analysing the posts themselves and also looking at any changes in trends over time. However, in general, posting in music boards seems to correlate negatively with the likelihood of making cybercrime-related posts.

V. DISCUSSION AND CONCLUSIONS

In this research, we analyse types of music and lyrics shared in Hack Forums, the largest and longest-running cybercrime forum. We also analyse the activities of the members sharing music across the entire forum, compared to a matched sample, to compare their posting activities relating to different types of cybercrime. No prior research on this topic could be found. Therefore, this one is likely to be the first work of research that focuses on analysing the music shared within the cybercrime underground. However, this project looks at data from only one forum, so future research could include analysing data from other platforms to see if similar patterns can be found.

A large majority of the music shared in the forum is electronic music, with hip-hop being the next most popular and very few tracks from other genres. Although hip-hop and rap music have been previously linked to crime and deviant behaviour [4], [6], there has not been any similar research

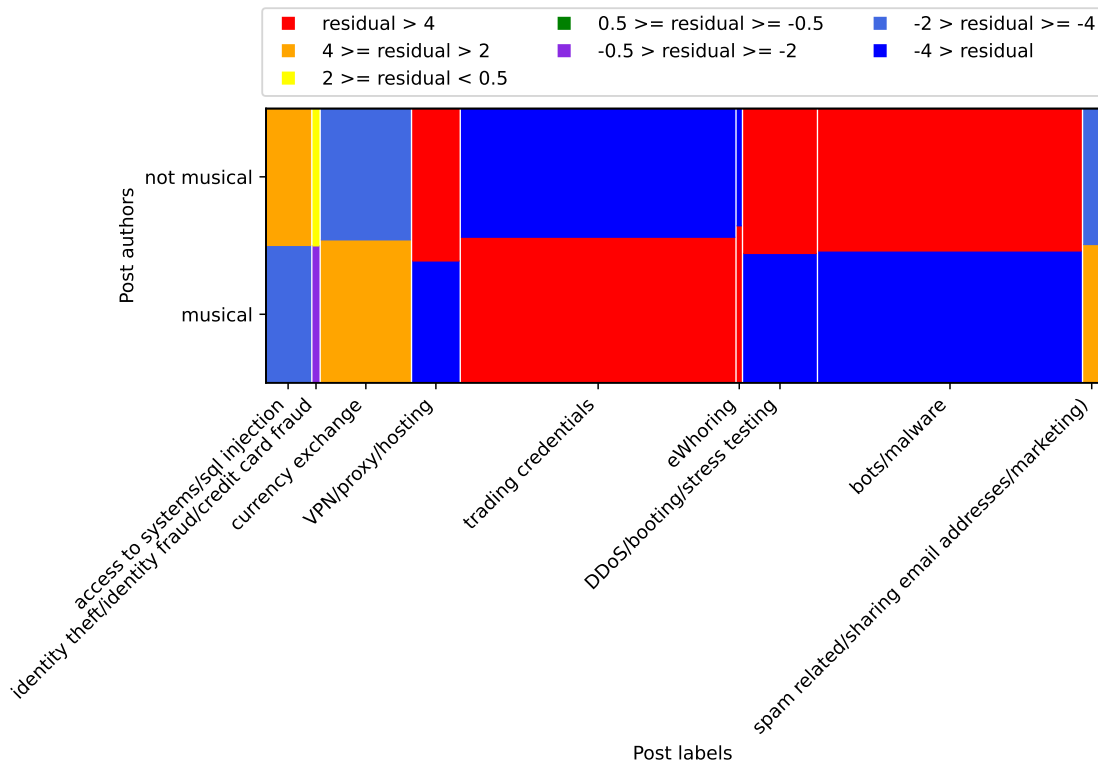


Fig. 10: Mosaic plot of posts from musical and non-musical users by crime type

in relation to electronic music. The popularity of electronic music may be due to one of the boards mainly being focused on EDM (Electronic Dance Music), but also because for users with no formal music training composing electronic music is the easiest way to start, without needing expensive equipment or years of musical training.

Although mentions of cybercrime are not very common in lyrical texts, we uncover other interesting topics. A large proportion of the lyrics popularise a “gangster lifestyle”. Other common topics include commentary on the song lyrics, where users express their opinions on the lyrics they share or ask for opinions from others. There are also some more emotional topics, such as song lyrics discussing love, relationships, sadness and god. A large number of lyrics also centre around descriptions of violence and murder. As seen from the examples found in §IV-B1, the mentions of hacking or cybercrime are generally either the users bragging about their own skills or insulting their rap battle opponents.

We score the toxicity of the posted lyrics using the Perspective API [12], which gives most posts either very low or very high toxicity scores. The manual annotation of a sample of posts shows there is a statistically significant difference in the distribution of the toxicity scores for the labels used, but also reveals some outliers in the toxicity scores for posts labelled as “insulting”. This shows that while the toxicity classifier, on average, gives toxic posts a higher score, it cannot be fully relied on for accurate classification of song lyrics. Through further analysis of the posts classified as “insulting”

but scoring very low in toxicity, we discover the Perspective API only takes into account the first 501 characters of the text. Thus, the toxicity score of any post can be lowered substantially by adding non-toxic text or even non-alphabetic characters to the beginning. No prior mention of this issue could be found. Although it is unclear how likely this bug is to be exploited by users then it can have more serious concerns when the API is used by researchers who are unaware of this limitation [59]. We reported this potentially concerning discovery to the Perspective API team.

In general, Hack Forums users who engage in music-related content are not very active when it comes to posting about cybercrime. They post overwhelmingly more non-criminal content compared to other Hack Forums users. This may be because they are active on music boards, where the content is generally not about cybercrime, as opposed to users who may have made their accounts with the sole intention of promoting their services and do not engage in the community side. However, there are some crime types that users belonging to this group are more likely to post about, in particular eWhoring and trading credentials.

In conclusion, there is no clear evidence to suggest that cybercrime is being glamorised through music shared within the criminal underground. However, the forum users do seem to be glamorising and idealising a certain “gangster lifestyle”, which includes an abundance of money, women, and drugs. This is also partially supported by the type of music style generally shared, with hip-hop being the second most popular

genre, although electronic music is much more popular among users. While key actors from music boards are, in general, less likely to post cybercrime-related content, when they do the crime types differs significantly compared to other forum users. Due to this being an unexplored topic with very limited research, future work could include analysis of music-related content in other cybercrime forums and further analysis to understand whether the characteristics of the Hack Forums users are generalisable to the broader cybercrime community.

ACKNOWLEDGEMENTS

We thank the Cambridge Cybercrime Centre for allowing us access to the CrimeBB dataset, as well as our colleagues for their support and advice. This work is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 949127).

REFERENCES

- [1] A. Fernández-Sotos, A. Fernández-Caballero, and J. M. Latorre, "Influence of tempo and rhythmic unit in musical emotion regulation," *Frontiers in Computational Neuroscience*, vol. 10, 2016. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fncom.2016.00080>
- [2] D. Västfjäll, "Emotion induction through music: A review of the musical mood induction procedure," *Musicae Scientiae*, vol. 5, no. 1_suppl, pp. 173–211, 2001. [Online]. Available: <https://doi.org/10.1177/10298649020050S107>
- [3] S. Saarikallio and J. Erkkilä, "The role of music in adolescents' mood regulation," *Psychology of Music*, vol. 35, no. 1, pp. 88–109, 2007. [Online]. Available: <https://doi.org/10.1177/0305735607068889>
- [4] J. Tanner, M. Asbridge, and S. Wortley, "Listening to rap: Cultures of crime, cultures of resistance," *Social Forces*, vol. 88, pp. 693 – 722, 2009.
- [5] B. A. Martínez Jr. and A. Selepak, "O som do ódio: explorando o uso das letras da música *ijê*hatecore*i*/*ijê* como estratégia de recrutamento pelo movimento da força branca," *Intercom: Revista Brasileira de Ciências da Comunicação*, vol. 37, no. 2, p. 153–175, Jul 2014. [Online]. Available: <https://doi.org/10.1590/1809-584420147>
- [6] I. Osmanović and E. Kazić-Çakar, "Music as a factor in etiology of crime: Can it make us accept the unacceptable?" *Journal of Criminology and Criminal Law*, 2021.
- [7] S. Pastrana, A. Hutchings, A. Caines, and P. Buttery, "Characterizing Eve: Analysing cybercrime actors in a large underground forum," in *Proceedings of the 21st International Symposium on Research in Attacks, Intrusions, and Defenses (RAID), Heraklion*, 2018, pp. 207–227.
- [8] J. G. Atondo Siu, B. Collier, and A. Hutchings, "Follow the money: The relationship between currency exchange and illicit behaviour in an underground forum," *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2021.
- [9] O. Tade, "Cybercrime glamorization in Nigerian songs," *International Journal of Cyber Criminology*, vol. 13, no. 2, 2020.
- [10] A. S. Oyenuga and O. Ajewole, "Let us dance to live: The influence of afro-pop-music on promotion of the norms of cybercrime among young people in Nigeria," *Journal of Advance Research in Social Science and Humanities (ISSN:2208-2387)*, 2023.
- [11] S. Pastrana, D. R. Thomas, A. Hutchings, and R. Clayton, "CrimeBB: Enabling cybercrime research on underground forums at scale," *Proceedings of the 2018 World Wide Web Conference*, 2018.
- [12] Google Jigsaw, "Perspective API," 2017. [Online]. Available: <https://www.perspectiveapi.com/>
- [13] J. Hughes and A. Hutchings, "Digital drift and the evolution of a large cybercrime forum," in *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 2023, pp. 183–193.
- [14] A. Caines, S. Pastrana, A. Hutchings, and P. Buttery, "Automatically identifying the function and intent of posts in underground forums," *Crime Science*, vol. 7, 2018.
- [15] B. Collier, R. Clayton, A. Hutchings, and D. R. Thomas, "Cybercrime is (often) boring: Infrastructure and alienation in a deviant subculture," *British Journal of Criminology*, 2021.
- [16] M. Yip, N. Shadbolt, and C. Webber, "Structural analysis of online criminal social networks," *2012 IEEE International Conference on Intelligence and Security Informatics*, pp. 60–65, 2012.
- [17] I. Pete, J. Hughes, Y. T. Chua, and M. Bada, "A social network analysis and comparison of six dark web forums," *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pp. 484–493, 2020.
- [18] M. Bada, Y. T. Chua, B. Collier, and I. Pete, "Exploring masculinities and perceptions of gender in online cybercrime subcultures," in *Cybercrime in Context: The Human Factor in Victimization, Offending, and Policing*. Springer, 2021, pp. 237–257.
- [19] A. Hutchings and R. Clayton, "Exploring the provision of online booter services," *Deviant Behavior*, vol. 37, pp. 1163 – 1178, 2016.
- [20] S. Pastrana, A. Hutchings, D. R. Thomas, and J. E. Tapiador, "Measuring eWhoring," *Proceedings of the Internet Measurement Conference (IMC)*, 2019.
- [21] A. Hutchings and S. Pastrana, "Understanding eWhoring," *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*, pp. 201–214, 2019.
- [22] J. Hughes, Y. T. Chua, and A. Hutchings, "Too much data? opportunities and challenges of large datasets and cybercrime," *Researching Cybercrimes: Methodologies, Ethics, and Critical Approaches*, pp. 191–212, 2021.
- [23] R. S. Portnoff, S. Afroz, G. Durrett, J. K. Kummerfeld, T. Berg-Kirkpatrick, D. McCoy, K. Levchenko, and V. Paxson, "Tools for automated analysis of cybercriminal markets," in *Proceedings of the 26th International World Wide Web Conference*, 2017, pp. 657–666.
- [24] B. Ampel, S. Samtani, H. Zhu, S. Ullman, and H. Chen, "Labeling hacker exploits for proactive cyber threat intelligence: A deep transfer learning approach," *2020 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pp. 1–6, 2020.
- [25] I. Pete, J. Hughes, A. Caines, A. V. Vu, H. Gupta, A. Hutchings, R. J. Anderson, and P. Buttery, "Postcog: A tool for interdisciplinary research into underground forums at scale," *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pp. 93–104, 2022.
- [26] G. A. Siu, B. Collier, and A. Hutchings, "Follow the money: The relationship between currency exchange and illicit behaviour in an underground forum," *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pp. 191–201, 2021.
- [27] T. Mizuno and M. Sugishita, "Neural correlates underlying perception of tonality-related emotional contents," *NeuroReport*, vol. 18, pp. 1651–1655, 2007.
- [28] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. R. Zapata, and X. Serra, "Essentia: An audio analysis library for music information retrieval," in *International Society for Music Information Retrieval Conference*, 2013.
- [29] T. Berg-Kirkpatrick, J. Andreas, and D. Klein, "Unsupervised transcription of piano music," *Advances in Neural Information Processing Systems*, vol. 27, 2014.
- [30] Y. Zhu, M. S. Kankanhalli, and S. Gao, "Music key detection for musical audio," in *11th International Multimedia Modelling Conference*. IEEE, 2005, pp. 30–37.
- [31] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull, "Music emotion recognition: A state of the art review," in *11th International Society for Music Information Retrieval Conference (ISMIR)*, vol. 86, 2010, pp. 937–952.
- [32] L. G. Giray, "The infusion of propaganda in the music education in China," *Aksara: Jurnal Ilmu Pendidikan Nonformal*, 2022.
- [33] M. Plantak, "Coca-cola, marlboro, suzuki: Turbofolk music as a tool of political propaganda during the 1990s in Serbia," *Research in Social Change*, vol. 12, pp. 26 – 41, 2020.
- [34] A. Parris, *Music as propaganda: Art to persuade and to control*. Greenwood Press, 1985.
- [35] E. Krsmanović, "Crimen et circenses: Serbian turbo folk music and organised crime," *Crime and Music*, pp. 149–167, 2021.
- [36] F. R. Castillo-Villar, J. Cavazos-Arroyo, and N. O. Kervyn, "Music subculture as a source of conspicuous consumption practices: a qualitative content analysis of "altered movement" songs and music videos," *Journal of Consumer Marketing*, vol. 37, pp. 353–363, 2020.
- [37] T. L. Anderson, P. R. Kavanaugh, R. Bachman, and L. D. Harrison, "Exploring the drugs-crime connection within the electronic dance music and hip-hop nightclub scenes," *Methodology*, vol. 36, p. 57, 2007.

- [38] L. Hirsch, “Weaponizing classical music: Crime prevention and symbolic power in the age of repetition,” *Journal of Popular Music Studies*, vol. 19, pp. 342–358, 2007.
- [39] M. Thompson, “To soothe or remove? affect, revanchism and the weaponized use of classical music,” *Communication and the Public*, vol. 2, pp. 272 – 283, 2017.
- [40] A. M. Davani, M. Atari, B. Kennedy, S. Havaldar, and M. Dehghani, “Hatred is in the eye of the annotator: Hate speech classifiers learn human-like social stereotypes,” *ArXiv*, vol. abs/2110.14839, 2021.
- [41] N. Lee, C. Jung, and A. Oh, “Hate speech classifiers are culturally insensitive,” in *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)*, 2023, pp. 35–46.
- [42] T. D. Oliva, D. M. Antonialli, and A. Gomes, “Fighting hate speech, silencing drag queens? Artificial intelligence in content moderation and risks to LGBTQ voices online,” *Sexuality & Culture*, vol. 25, pp. 700–732, 2020.
- [43] T. Gröndahl, L. Pajola, M. Juuti, M. Conti, and N. Asokan, “All you need is “love”: Evading hate speech detection,” in *Proceedings of the 11th ACM Workshop on Artificial Intelligence and Security*, 2018, p. 2–12. [Online]. Available: <https://doi.org/10.1145/3270101.3270103>
- [44] H. Hosseini, S. Kannan, B. Zhang, and R. Poovendran, “Deceiving Google’s Perspective API built for detecting toxic comments,” *ArXiv*, vol. abs/1702.08138, 2017.
- [45] British Society of Criminology, “Statement of ethics,” 2015. [Online]. Available: <https://www.britisocrim.org/documents/BSCEthics2015.pdf>
- [46] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, “librosa: Audio and music signal analysis in Python,” in *Proceedings of the 14th Python in Science Conference*, vol. 8, 2015.
- [47] J. Pons and X. Serra, “musicnn: pre-trained convolutional neural networks for music audio tagging,” in *Late-breaking/demo session in 20th International Society for Music Information Retrieval Conference (LBD-ISMIR2019)*, 2019.
- [48] T. Bertin-Mahieux, D. P. W. Ellis, B. Whitman, and P. Lamere, “The million song dataset,” in *ISMIR*, A. Klapuri and C. Leider, Eds. University of Miami, 2011, pp. 591–596. [Online]. Available: <http://dblp.uni-trier.de/db/conf/ismir/ismir2011.html#Bertin-MahieuxEWL11>
- [49] E. Law, K. West, M. I. Mandel, M. Bay, and J. S. Downie, “Evaluation of algorithms using games: The case of music tagging,” in *International Society for Music Information Retrieval Conference*, 2009.
- [50] C.-P. Tsai, Y.-L. Tuan, and L.-S. Lee, “Transcribing lyrics from commercial song audio: the first step towards singing content processing,” *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5749–5753, 2018.
- [51] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit*. ” O’Reilly Media, Inc.”, 2009.
- [52] R. Rehurek and P. Sojka, “Gensim–python framework for vector space modelling,” *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, vol. 3, no. 2, 2011.
- [53] M. D. Hoffman, D. M. Blei, and F. R. Bach, “Online Learning for Latent Dirichlet Allocation,” in *NIPS*, 2010.
- [54] N. D. Ousidhoum, X. Zhao, T. Fang, Y. Song, and D.-Y. Yeung, “Probing toxic content in large pre-trained language models,” in *Annual Meeting of the Association for Computational Linguistics*, 2021.
- [55] “Song keys ranked by popularity,” hooktheory.com/cheat-sheet/key-popularity/, accessed: 2023-06-01.
- [56] “Best chords for edm? mixedkeys,” <https://mixedinkey.com/captain-plugins/wiki/best-chords-for-edm/>, accessed: 2023-06-01.
- [57] “Rhythm tips for identifying music genres by ear,” <https://www.musical-u.com/learn/rhythm-tips-for-identifying-music-genres-by-ear/>, accessed: 2023-06-01.
- [58] “Lorem ipsum - all the facts - lipsum generator.” [Online]. Available: <https://www.lipsum.com/>
- [59] A. V. Vu, L. Wilson, Y. T. Chua, I. Shumailov, and R. Anderson, “ExtremeBB: A database for large-scale research into online hate, harassment, the manosphere and extremism,” in *Proceedings of the ACL Workshop on Online Abuse and Harms (WOAH’23)*, 2023.