Exploring Masculinities and Perceptions of Gender in Online Cybercrime Subcultures

Maria Bada; Yi Ting Chua; Ben Collier; Ildikó Pete
University of Cambridge, Cybercrime Centre

Abstract

While there is now a substantial literature on the role played by online forums in cybercrime economies, there has been little research which accounts for the role played by gender in these communities. We study the role of gender in cybercrime communities, using an innovative research methodology which makes use of both qualitative and data science elements approaches to analyse a very large sample of posts on a cybercrime forum. Our findings suggest that a substantial sub-section of cybercriminal activity associated with these forums is deeply tied up with ideas about gender. A significant number of the actors we studied first became involved in these forums in an attempt to hack, stalk, or blackmail an intimate partner (usually a woman). Additionally, once involved in these communities, the performance and commodification of femininity was a key part of many of the 'less-technical' or 'entry-level' forms of cybercrime which we observed. Finally, despite the low technical skill of most of these actors, we found that they still had a deep connection to the 'hacker' identity, using misogyny to legitimise their position within this subculture and construct hacking as intrinsically masculine. We conclude by reflecting on the potential relevance for these findings for policy and intervention approaches in low-level cybercrime communities.

Keywords – *Gender, Cybercrime, Underground forums, Grounded theory, Data Science, Natural Language Processing*

1 Introduction

Underground forums have been hypothesised to play a key role in cybercrime ecosystems, providing a platform for various activities, such as the exchange of illegal products and services, and facilitating the sharing and spreading of ideas, specialist knowledge and norms. These forums represent an enormous corpus of data, often comprising millions of posts over several years, and as such present unique challenges to criminological research. In this paper, we engage in innovative mixed-methods research to study a particularly prominent cybercrime forum, in order to improve academic understanding of the activities of forum members, their interactions and social dynamics, and the role these forums play in criminal offending.

We focus on the role which gender perceptions play in these forums, and in cybercrime more broadly. Although there is a wide body of research literature on gender and crime more generally (Messerschmidt, 2009), the application of this perspective to cybercrime has been limited. Some findings show that men are generally more likely to engage in cybercrime (Barlett & Coyne, 2014; Bachmann, 2010), although a few studies yield the opposite conclusion (Marcum, Higgins, Freiburger, & Ricketts, 2012), or indeterminate results (Bossler & Burruss, 2012; Holt & Morris, 2009). Understanding why women are not only less likely to commit cybercrime offences, but also why some do, represents an important step towards deepening academic understanding of cyber-offending, the pathways through which individuals become involved, and why they remain part of these communities (Hutchings & Chua, 2016).

This study investigates a very popular cybercrime forum which is largely dedicated to ‘low-level’, less-technical forms of cybercrime. Forums such as this represent an important gateway to cybercrime subcultures for many individuals (Holt, 2007; Jordan & Taylor, 1998). We expand on the work by Steinmetz and colleagues (2019) to understand and examine the gendered structures and dynamics within cybercrime subcultures.

This paper is structured as follows. Section 2 reviews current literature on hacking and cybercrime communities. In Section 3, we describe the data and the methodology followed. Section 4 presents the findings of the qualitative and quantitative analysis. In Section 5 we discuss our findings and conclusions.

2 Literature Review

2.1 Hacking

The figure of the 'hacker' looms large in research on cybercrime. There is a wide research literature on hackers and hacker culture, which documents the rise of this subaltern subculture of deviant technologists (Taylor, 1999). What unites hackers is a set of sensibilities towards technology and creative technical practices. Hackers engage with technological and social systems through 'creative disruption', cleverly repurposing systems, finding flaws, and subverting mechanisms of control (Coleman & Golub, 2008; Holt, 2007; Jordan & Taylor, 1998). This can take a range of distinct forms - while the stereotypical 'hacker' involves high-skilled technical exploitation, a 'hack' can involve many different kinds of knowledge and practices. These might be maintenance workers finding creative solutions to keep things running, activists using existing technologies in powerful new ways to achieve political goals, 'social engineers' who exploit human psychology in order to gain access to secure buildings or carry off scams, and 'culture-jamming' art all have a 'hackery' quality.

Rather than a single 'hacker ethic', research on hacker communities suggests a wide diversity of perspectives, often, though not exclusively, expressive of different variants of technological liberalism (Coleman & Golub, 2008). Coleman's work (e.g., Coleman, 2014; Coleman & Golub, 2008; Coleman, 2012, 2011, 2004) in particular documents a range of different sensibilities and politics within hacker communities, from privacy–enthusiasts who develop encryption protocols, to developers working in the Free and Open Source Software movement, to the 'underground' hackers who prefer to utilise their skills for disruption,

amusement, and to demonstrate their technological skill (Coleman & Golub, 2008). All share an ethic of creativity and experimentation, and a deep immersion in technological practice of different kinds.

While hacker communities are increasingly becoming involved in actively political struggles (Coleman, 2017; Steinmetz et al., 2019), they often (though again, not exclusively) share a distaste with overt politics, preferring to focus on the practicalities of technical work (Coleman, 2004). This is also implicated in some of the documented issues of misogyny in, for example, open source communities, where it is used as a veil for the silencing of women's voices and contributions (Nafus, 2012). Hacking culture is also associated with a privileging of 'hard mastery' and domination of technology, which plays into the misogyny often ascribed to these communities (Yar, 2005). These intense interests with various aspects of technology tended to develop at an early age (Holt, 2007; Steinmetz, 2016). In addition to the closeness to technology, this norm encompasses an appreciation toward various sets of technical skills and applying these skills in innovative manners (Holt, 2007; Jordan & Taylor, 1998). The appreciation for secrecy is reflected via hackers' need to keep their activities relatively unknown without compromising one's reputation (Jordan & Taylor, 1998). Specifically, 'underground' hackers need to avoid detection by the law enforcement agencies. However, hackers also need to share their accomplishments with fellow hackers for status and reputation management. This particular norm thus deems hackers as a group of "odd criminals, seeking publicity" (Jordan & Taylor, 1998, p. 765). Such reputation is of interest and importance to hackers as it dictates the types of interactions received from fellow hackers (Holt, 2007).

2.2 Gendering cybercrime

As is evident from this brief summary, many (if not most) hackers are not involved in cybercriminal activities, and many people involved in cybercrime are not hackers. Due to the 'technical deviance' which hacking embodies, these have often been conflated in the research

literature, however they are distinct sets of cultures and practices. Nevertheless, hacking provides an important cultural reference point for many cybercrime communities as part of their foundational self-mythology (Holt, 2007; Steinmetz, 2016; Taylor, 1999; Turgeman-Goldschmidt, 2008).

Prior research has identified that the population involved in cybercrime consists primarily of boys and men (Bossler & Burruss, 2012; Hutchings & Chua, 2016; Jordan & Taylor, 1998; Marcum, Higgins, Ricketts, & Wolfe, 2014), although several notable women hackers have emerged throughout hacker history, such as Susan Thunder (Furnell, 2002). It is suggested that cybercrime subcultures operate within a 'boy culture' centred on demonstrations of mastery, wherein individuals show their ability to dominate their 'social and physical environment' (Thomas, 2002; Holt, 2007). The hegemonic masculinities thesis, developed notably in the context of criminology by Messerschmidt (2009, 2013), argues that although there are many different ways of 'doing' masculinity, they are all oriented around, towards, and sometimes in opposition to the 'hegemonic' form of masculinity which undergirds patriarchal structures of power. Men and boys are situated in structural and cultural relationships with one another, and within distinct groups can develop different orientations towards dominant masculinities.

In deviant subcultures, illegal activities can become bound up with masculinity - as a way of achieving hegemonic masculinity when other routes are blocked. Different subcultures and groups develop their own distinct constructions of masculinity, drawing on different forms of crime or deviant experience in accessing and performing masculinity and its attendant forms of structural power (Messerschmidt, 1994). Shared constructions of gender perform a role in constructing the shared social life of subcultures, underpinning crucial aspects of the systems of internal social capital which draw individuals in and bind them together. Oriented as these masculinities are towards broader regimes of gendered social power, these subcultures

therefore often reproduce gendered power relationships. These can mirror those which exist in mainstream society, with women's experiences and identities side-lined or subordinate (Messerschmidt, 1994).

Taylor (1999) argues that the gender gap in cybercrime may be explained in part by 'masculine environment' or computer culture 'dominated by men while women feel threatened or uncomfortable'. As with other deviant masculine subcultures, this can lead women who engage to be rendered invisible, whether intentionally by the individual, or marginalised by the culture. Women may experience a magnified misogyny deployed to reinforce gendered status hierarchies, rendered subordinate, silenced, or absent altogether. For online subcultures, with access to computer intrusion tools and automated posting capabilities, these harms may be compounded through organised "flaming" or "trolling" or even more harmful forms of harassment, such as the theft and publication of intimate photographs (Steinmetz et al., 2019).

In general, studies framed with a subcultural approach employ interviews with those involved in cybercrime communities and 'hacking'-related publications and web forums as sources of information (Holt, 2007; Jordan & Taylor, 1998). However, these studies on norms mainly focus on the perceptions of men and masculinity of those involved in cybercrime. To date, there is limited study on how those (of any gender) involved in cybercrime communities perceive women. A recent study with the inclusion of four women hackers among a sample of 33 is done by Steinmetz and colleagues (2019) who examined the cybercrime subculture and its provision of social structures that enable gender to be performed and reinforced in non-mainstream manners. Other studies that have included women generally examined cybercrime involvement among a college sample (Bossler & Burruss, 2012; Skinner & Fream, 1997), which may not be representative. In other words, women are silenced from the narrative of defining hacking and its subcultural norms. This illustrates a gap in understanding in the current research literature regarding the gendered dynamics of cybercrime, especially with regards to

pathways to initiation and cultural pull factors. Thus, our study aims to broaden the literature on gender construction by members of cybercrime communities to include how women are perceived.

3 Data and Methods

We perform a *qualitative* analysis of an underground forum complemented by a *quantitative* analysis through *Data Science* approaches, specifically *Natural Language Processing (NLP)*, which allows us to compare our deep qualitative analysis with a broader picture of activity on the forum over time. The motivation to combine these two methods stems from the fact that the size of the dataset and the purposive sample does not allow a manual analysis of the millions of posts which have been made on the forum over its lifespan. Using NLP techniques provides a means to perform exploratory analysis on the entire extracted dataset. Additionally, it serves the purpose of comparing the findings of the qualitative and quantitative analysis. In the following subsections we describe the dataset used to obtain underground forum data, the sampling and pre-processing approach, and the specifics of the analysis.

3.1 The CrimeBB Dataset

Underground forum data has been extracted from the *CrimeBB* dataset (Pastrana, Thomas, Hutchings, & Clayton, 2018), which is provided by the Cambridge Cybercrime Centre (CCC)[1]. CrimeBB, which is available through a legal agreement via the CCC, was created to allow large scale, longitudinal analysis of underground forums and cybercriminal communities. At the time of writing this paper CrimeBB contains over 84 million posts shared by over 2 million members, and data is scraped from a number of online communities. The forums are usually structured in a similar manner: they contain sub-forums, and posts written by members can be viewed in the context of threads, which is a collection of posts centred around one

[1] https://www.cambridgecybercrime.uk

discussion. Sub-forums can be categorised based on the topics they discuss, and they may serve as marketplaces. For this analysis we decided to focus on a single forum, which is one of the top low-level *English-language* cybercrime communities, with over 3 million members and 174 sub-forums.

3.2 Preprocessing

To address the research question, a *purposive sample* was generated with *keywords*, that is, relevant posts shared by members in the community in various sub-forums were retrieved using specific keywords. For the current analysis we deemed the singular and plural forms of the following keywords sufficient: *'girl'*, *'woman'*, *'female'*, and *'gender'*. The use of purposive sample is common within qualitative studies (Holt, 2007). Purposive sampling with these keywords allows us to examine topics associated with gender, especially with women, within the hacker subculture. As the keyword-based search also returned spam posts including code snippets, long lists of software, and miscellaneous links, the extracted posts went through further cleaning to filter these out. This resulted in a *total of 491 361 posts* authored between *2007* and *2019*, which served as the input for both the *quantitative* and *qualitative* analysis introduced in the subsequent sections.

3.3 Quantitative Methods

As mentioned in Section 3, our aim in applying NLP methods was to support the qualitative analysis by automating the analysis of the entire purposive sample and to provide insights to the themes discussed on the forum. There has been some work applying Machine Learning and NLP tools to research underground forums to date (Portnoff et al., 2017; Lui & Baldwin, 2010; Caines, Pastrana, Hutchings, & Buttery, 2018b; Overdorf, Troncoso, Greenstadt, & McCoy, 2018; Pastrana, Hutchings, Caines, & Buttery, 2018). Such techniques are especially useful for tasks where a manual analysis would prove to be cumbersome or impossible due to the size of the dataset used. In particular, these techniques can aid in

automating the identification of post types (Portnoff et al., 2017), the function and intent of posts (Caines et al., 2018b), to mention a few. The application of an NLP-based approach is however not without challenges due to the particular characteristics of posts on underground forums when compared to grammatically correct English language texts.

Data Cleaning. The first step of the quantitative analysis was data cleaning. The purposive sample was further processed using regular expressions to remove references to images and non-relevant characters from posts. Posts containing quotations, source code examples and links were removed. Similarly, one- and two-character long alphanumerical characters, and names were omitted as the exploratory analyses revealed that they were present in the data in large numbers, which would have significantly affected the results.

NLP Text Pre-processing. Next, the steps of a common NLP pipeline were applied to obtain frequency counts of keywords mentioned in the extracted posts, and to discover topics of discussions. The NLP pipeline starts with text cleaning. As part of this, stopwords, numbers, and punctuation were removed to exclude commonly used and non-relevant words, then the text was converted to lower case. Finally, posts were tokenised, which resulted in a list of tokens. Put simply, a token is a basic unit corresponding to a word. This pre-processing is of key importance due to the nature of the input text as mentioned above, and it was subsequently revisited based on initial findings of the analysis.

Frequency distribution. Frequency distribution simply provides the number of times each token occurs in a given text. The threshold used to decide whether a token is frequent in the given document is dependent on a number of factors, such as the number of tokens. Relevant most frequent terms include gender related words, such as ‘girl’, ‘girlfriend’, ‘women’, and ‘female’. This was expected since the original text was the result of a keyword search including these terms. Other frequent words provide initial insights to how these keywords are discussed. However, counting tokens is merely useful as a tool to gain an understanding of the data.

Frequent bigrams and collocations on the other hand provide meaningful pairs of significant keywords in the document.

Bigrams and Collocations. Thus, we continued by investigating *bigrams* - co-occurring terms - and *collocations* - strongly associated *ngrams*. An *ngram* is a sequence of n words. Similarly, a *bigram* is a sequence of two words, *x* and *y*. Since bigrams on their own do not provide meaningful insights in the given context, we selected frequent bigrams - frequently co-occurring terms. Frequent bigrams simply indicate how often word *x* is followed by word *y*. Next, meaningful collocations were discovered. Collocations are common phrases that consist of at least two words (bigram) and act like single words, such as 'machine learning'.

Topic Modeling. The aim of topic modeling is to discover abstract topics in a collection of documents. *Latent Dirichlet Allocation (LDA)*, the approach we adopted, is an example of topic models and is used to classify text in a document to a particular topic. LDA is widely used and has been applied in underground forum analysis (Pastrana, Hutchings, et al., 2018). Following the identification of significant collocations, topic modeling is performed on subsets of the purposive sample created based on a keyword search of the collocation of interest. These collocations are ('social' 'engineering') and ('thanks' 'advance').

3.4 Qualitative Methods

The qualitative analysis of the randomly drawn posts involved the use of a modified *grounded theory* approach (Bowen, 2006; Corbin & Strauss, 1990). Grounded theory methodology is common among studies on cyberspace subcultures using samples of web forums and online materials (Blevins & Holt, 2009; Holt, 2010, 2007). The methodology involves an inductive in-depth analysis of the sample, such as transcribed notes from interviews or texts from online discussions. That is typically conducted by hand on printed copies of the materials (Holt, Blevins, & Burkert, 2010). The main aim of using grounded theory methodology is to generate a theoretical model through detailed coding of data rather than

relying on existing concepts (Creswell, 2013). However, for the purpose of this study, the grounded theory methodology provides a framework for the inductive identification and categorization of the data.

To perform qualitative analysis, we extracted data from the pre-processed sample of *491,361* posts. Due to the size of the extracted data it is not feasible to perform manual qualitative analysis and coding, thus initially *40* posts were selected randomly, and subsequently *20, 200, 100,* and finally *300* posts were added to this sample. The final batch of samples covers posts from *2008-2019*, while the rest of the samples focus on the period of *2007* and *2008*. The reason for selecting a shorter time period was that the number of posts from the years *2007* and *2008* is significantly less compared to the total number of posts, which results in a sample that is representative of its respective time period. This sample is then compared against a sample drawn from a longer time period, which allows us to compare posts from *2007-2008* to posts from later years. Analysis was conducted at the post-level, rather than at the thread-level, to capture variations in the nature and content of posts.

For *open coding*, we utilized a sample of 60 randomly selected posts from 2007. Based on this initial sample, a list of 131 codes were generated by one of the authors. Some examples of the codes included pretending to be women to attack, sending nude pictures with malware embedded, questions on hacking into accounts, and self-declaration of using keylogger and malware. During *axial coding*, another author read and coded through the initial sample. The two authors discussed through discrepancies in codes and identified connected concepts that can be subsided under a broader category. This resulted in the collapse of the 131 codes from open coding to a list of 11 categories. To ensure the categories are validated and new categories are addressed, both authors coded through a randomly drawn sample of 300 posts from 2008. During this process, some new sub-categories emerged from the new sample and were added to the code list. For the last stage of coding, content labelled within categories was combined,

resulting in the collapse of the 11 categories into four core categories. To ensure the established categories are representative of the extracted dataset, the authors coded through a random sample of 300 posts between 2008 to 2019. This final sample yielded two core categories: discussion on hacking and attack methods and discussion on gender norms.

4 Results and Findings

4.1 Data Science Findings

Bigrams and Collocations. In our bigram analysis we selected terms that occur together more than 500 times. This reveals certain themes in the text. The bigram ('year', 'old') indicates that members often mention age, and post about women in different contexts (through bigrams, such as 'girl, know', 'girl, like', 'every, girl' and 'hot girls'). Another theme is related to members requesting help shown by bigrams ('need, help') ('like, know') ('thanks, advance'), and ('want, know'). Other themes are related to social media - through the bigrams ('facebook, account') ('social media') and social engineering is also frequently discussed - ('social, engineering'). The bigram ('make, money') occurs more than 500 times. *Table 1* shows these bigrams with their associated frequency counts. In the current analysis of collocations terms that occur together more than 500 times are considered and infrequent expressions are eliminated. Next, these frequently occurring terms are further filtered using the *Pointwise Mutual Information (PMI)* score, which measures how likely we are to see two events co-occur relative to when the two events are independent (Raviv, 2020). Thus, PMI allows us to keep only those terms that are more likely to occur together than as if they were independent. A disadvantage of this approach is that PMI is sensitive to rare combinations of bigrams where the individual terms do not appear frequently individually. These might be flagged up as a significant collocation. *Table 2* shows the most significant collocations and highlights a few key insights to the discussions. We can observe that members discuss age and gender related topics, social engineering, and they also request help. The same idea was revealed by the

analysis based on bigrams. For example, the collocation ‘social engineering’ reveals posts sharing knowledge about social engineering, which provide definitions, discuss purpose and basic methods. A subset of the posts covers topics related to human manipulation. These are opinions expressed about women who are ‘experts at social engineering’, advice given to women on how to apply social engineering techniques against their partners, and some members posting about applying social engineering to find a partner or to influence women to give out private information. Social engineering is also discussed as a technique applied in eWhoring, which is detailed in Section 4.2.2. Video games are mentioned in multiple contexts including it being an interest of members similarly to hacking, opinions on whether women play or do not play them, and their effects on children and social skills. The collocation ‘United, States’ reveals discussions and opinions about politics and history. Analysing the context of the collocation (‘thanks’, ‘advance’) shows that help is requested not only in technical questions. Most posts are requests for download links of images of women and image packs related to eWhoring.

Topic Modeling. We chose a 10-topic model for both of the documents based on the collocation (‘social’, ‘engineering’) and the collocation (‘thanks’, ‘advance’). *Table 3* and *4* present the respective topics that were selected, and the words associated with each topic. The topics were established using domain knowledge and by reading related posts. The words we deemed most significant are highlighted in bold. The first topic we identified involved individuals discussing joining up to a particularly popular forum group. A second set of topics contained posts concerning *eWhoring* (a low-level form of cybercrime where stolen or purchased nude photographs and videos of women are used to scam men) - a form of human manipulation and social engineering. Members share tutorials and experiences including the amount of money they make. eWhoring methods and using remote access trojans (RATs) in particular present topics on their own within this set. The word “traffic” is mentioned in the

context of social engineering, in particular generating eWhoring traffic, or a steady flow of men to scam: "*You are lazy to catch a guy via Social Engineering mate. No one willing to buy from you if you say ""sell"" on traffic sites because people think you are fake, bot or scammer. If you can verify yourself you are a real girl you can sell directly on sites*" (Quote 1).

A third topic we identified is 'Information gathering', which discusses techniques hackers can use to obtain information on their targets (generally women) as part of doxxing: "*...Through this simple search I found out enough information to steal this girl's identity*" (Quote 2).

The topics presented in *Table 4* centre around requesting either technical or relationship advice. As shown by the following quotes, while some of the advice is specific to eWhoring, hacking techniques related to accessing partners' accounts are also discussed.

> *So my request is for someone to edit pictures of hot girls i know into making it look like they are wearing a particular strip...* (Quote 3).
>
> *hello, im new to hacking and i want to know how to spy on my girlfriends computer. I want her passwords and anything she types...shes very easy to trick so... Also would love to know about spying on her cell phone* (Quote 4).

Pulling this analysis together, we argue that these 'data science' methods are a useful complement to more targeted qualitative research on very large datasets, clearly identifying a series of important categories of discussions as a set of initial concerns and themes to act as sensitising concepts for our qualitative investigations. Accordingly, we now move on, with these discussions in mind, to our more detailed qualitative analysis.

4.2 Qualitative Results

The modified grounded theory method was used to explore general topics and discussion with a gendered lens. Two core categories emerged from the inductive approach: discussion of hacking and attacks and discussion of gender norms. These categories provide an

overview on the discussion of gender within this cybercrime subculture, with an emphasis on the role of women in general and illegal online activities. Quotes from the randomly drawn samples are included when appropriate. We flag up here as a note to readers that (although we have redacted swearing), the quotes we use contain some offensive and violent imagery.

4.2.1 Stalking and gendered victimisation as a point of initiation

Our first finding was particularly striking, and came out immediately as we engaged in qualitative analysis of the data. It was identified that there are a large number of people whose first interactions with the forum are in the service of attempting to harass, stalk, or bully women, who were often, but not always, intimate partners. The majority of the posts which we observed were men asking for advice or help in hacking a woman's account. We provide some illustrative examples of the posts below:

> *I wanted to hack my girlfriends AIM MSN accounts because, sadly I think she's cheating on me...*(Quote 5).

Very occasionally, we observed cases of a woman wanting to hack a man's account. While hijacking and stalking fulfils a legal definition of cybercrime (and is clearly harmful), these efforts are generally not technically skilled and as a result, they are considered a fairly low-status activity on this forum:

> *The other day i was looking at the member list and i realized how many members just left [Forum Name] , Just Like That, and i dont like that!...most new members just go on and start posting spams,like demands and hack this, hack that.. when you start spamming you'll see the taste of hacking , i mean of the real hacking, not Real Hacking , but the hacking that you've never seen before...*(Quote 6).
>
> *You are the 10 000 000th f****** person who just wants to come and hack his girlfriends myspace or some shit and run off* (Quote 7).

From the first user's perspective, such a request is not "real hacking". This is consistent with current understanding on hackers' norms of having an intimate relationship with technology and maintaining a high level of curiosity and desire for new knowledge (Holt, 2007; Jordan & Taylor, 1998). Simply asking for help without demonstrating interest is inconsistent with the norms of the hacker subculture and therefore such behaviour is looked down upon and discouraged within the community. These posts suggest that requests for help on harassment and stalking are points of introduction to hacking for some forum members. However, with the lack of respect from more senior members of the community, these beginners and newcomers are confronted with the possible anxiety of their identity as hackers, especially since these practices were not 'real' hacking practices.

4.2.2 Gender as a resource and a risk in cybercriminal practices

When it comes to the types of attacks and practices discussed in the forums, our research has found that there is a substantial subcategory of 'low-level', or less technically-skilled cybercrimes which involve either targeting women, or the use of performed femininity as a resource for the completion of illegal or harmful activities. These are generally well documented forms of cybercrime which have been well-covered in the resource literature, though generally not in the context of a set of gendered cybercriminal practices. The main categories which have been identified are presented below.

Harassment. Harassment was one of the main types of attack described in forum discussions. It is notable to mention that this type of attack is mostly used against women (Caines, Pastrana, Hutchings, & Buttery, 2018a). One of the most characteristic ways of harassing women is described as blackmailing after taking screen shots on webcam: "*I use to have so many girls on webcam and the only thing I could do was take a screen shot . . . Blackmailing them is the best"* (Quote 8). Another example is using a picture to spread rumours

about a girl: "*this guy has a picture of my girlfriend and he is spreading the picture and telling everyone that my girlfriend is his own*" (Quote 9).

Privacy breach. Breaching privacy is also a type of attack discussed in hacking forums. It is quite interesting that one of the main topics that emerged in relation to privacy breach is that of hacking into the account (mostly Myspace, Facebook, Ebay and email accounts) of the significant other. The main reason for that is either to identify if the significant other is unfaithful or as a type of revenge. An example is:

> *You should have used the last girls' myspace account and sent spam messages to all 289 of her friends to get her account suspended for TOS [Terms of Service] abuse* (Quote 10).

eWhoring. eWhoring is an emerging form of less-technical cybercrime, as documented by Hutchings and Pastrana (2019). It involves the use of stolen or purchased images and videos of women to defraud (largely) men online, who are tricked into handing over money in exchange for what they believe is an online sexual encounter with a woman. These men are occasionally blackmailed by the person committing this crime once the deceit has been unveiled. We observed a considerable amount of references to this activity. For example, "*I ewhored on a page, for almost 6 months now, every time there is retard who writes to me and its the same person*" (Quote 11).

Photobucket hack. One of the most discussed hacking techniques is stealing user information and photos from Photobucket, a website which hosts private photo albums of customers including sexually explicit images. Thus, attackers can potentially use this as a tool for sextortion or blackmail as shown by the following post: "*Blackmailing them is the best, I use to trade nudes off a website with other guys*" (Quote 12)

Social engineering. While media accounts of hacking focus on its technical dimension, in fact, much of hacker practices are targeted at weaknesses in social systems and human

psychology (Rusch, 1999; Hancock, 1998). Known as *social engineering*, this method has been a common way to exploit systems, organisations, and institutions, using subterfuge, blackmail, and manipulation to gain access. We found a range of examples, which explicitly mention women, either through the affectation of femininity in the pursuit of attracting or manipulating a target, or with women as explicit targets: "*Seems like social engineering might be your best shot. You have to take every bit of knowledge you have about this girl and use it"* (Quote 13) and "*Just try to link something on your profile to attract people. Like cute girls is my favorite acc [account]for phishing" (*Quote 14). One of the main techniques of social engineering we identified is creating a fake account and pretending to be a girl. As a forum member describes: "*for example you could get close to someone pretending to be a woman interested in him or something... and then offer to send him a nude picture of yourself and send him a picture you found on the internet but bind that picture with a keylogger. That's just an example to give you an idea of what social engineering is*" (Quote 15).

We also identified advertisements for fake women's accounts. For example: "*Looking for a stat instagram with a girl or luxury niche with around 10k followers my budget is around 80-100 dollars"* (Quote 16).

Therefore, gendered performances of stereotypical femininity are used here as a way to attract someone's interest and manipulate them, becoming a victim of social engineering. Interestingly, it is identified that also within the hacking forums, someone might use gender pretending to be a girl in order to convince hackers to provide their help in hacking an account, as shown in this post:

> *No one would ever help a guy out because they think he was some lamer' Or something along those lines* (Quote 17).

Additionally, using a woman's name is quite common such as "Cutie" or "Baby girl" as a way to convince others of being a woman. This has been previously identified in Hutching's et al.'s (2019) studies of eWhoring.

Keylogger. Keylogger, a type of surveillance technology used to monitor and record each keystroke typed on a specific computer's keyboard, has also been identified as one of the main methods in hacking (Creutzburg, 2017). As mentioned above, social engineering might be used as a method facilitating keylogging. Often, a keylogger is used in order to hack the account of current or previous significant others. For example:

> *send him 'pics of yourself' that you will merge with a keylogger formerly. Search around for some new keyloggers*. . . (Quote 18).
>
> *yea dude bind your keylogger to a picture or song, go google for a file binder and then bind the two together and send it to her via e-mail.....I got my girlfriends email password without even trying that way* (Quote 19).

Remote Access Trojans (RATs). As described in the type of attack section, Remote Access Trojans (RATs) were suggested as a method for breaking into an account and breaching privacy. Attackers use RATs not to hijack their targets' computers and, in some situations, sextort their victims. First, these are inexpensive and simple to deploy, no sophisticated technical skills are necessary, and a variety of forums and websites provide tutorials on the subject. Thus, this method is suggested to newcomers as a method to breach their significant other's account or computer. RATs are commonly used to target women, either for abuse, stalking, harassment, or to collect images used for blackmail or eWhoring (Hughes & DeLone, 2007; Halder, Jaishankar, & Jaishankar, 2012; Southworth, Dawson, Fraser, & Tucker, 2005).

4.2.3 Gender norms and access to 'hacker' identity

In addition to sharing tools and knowledge, posting requests for services, and discussing cybercrime or hacking, the forum also contained a substantial amount of general discussion,

acting as a place where the community could interact with one another. Gender, in particular the reinforcement of norms about the social construction of men's and women's identity, played an important role in these discussions. This sub-category encompasses a variety of posts. The first group of posts centred around the topic of gender and sex. The majority of these posts were from the year 2008 and onward. In these posts, there was an openness in the discussion of sexual acts and fetishes. A small number of these posts were explicitly homophobic, identifying undesirable acts as 'gay' in the puerile, 'locker-room' use of the term. Interestingly, this notion of 'gay behaviour' was challenged by a woman member: '*And btw, since I'm not too homophobic and a girl, I can handle the gayness of the post. And what does it even mean, "looks gay"? When did you become an expert?'* (Quote 20). Overall, though, posts were male-centric where discussions of gender roles and sexual acts were in relation to men. This lean towards men in content was evident in the second group of posts focusing on the discussion on social interactions between gender ranging from friendships to relationships.

The content of posts prior to 2008 mainly revolved around the discussion of relationships, with an emphasis on trust. This is expected given the high numbers of requests on hacking into significant others' accounts. However, the content of the posts from the year 2008 and onward shifted towards requests and advice on social interactions: "I*f you "want a girlfriend over the summer", we have to be clear that you shouldn't like her solely because you want to be able to say that you have a girlfriend....Now on to actually asking her out: Girls enjoy confidence in guys. So, don't even think about asking her out via a telephone, etc...."* (Quote 21).

This advice and content demonstrated the negotiations and discussions on appropriate social interactions. On one hand, the poster of the quote indicated appropriate behaviours within relationships, such as the right reason for starting a relationship. On the other hand, posters often drew on generic and stereotypical views on gender, as illustrated in the above

poster's mention of girls' preference on confidence in guys. Some of this advice were rooted in negative assumptions, such as women's preference for the 'bad boys' (Quote 22), or when discussing men's requirements of girlfriends (Quote 23).

> *make her jealous... kiss ur new girl passionately infron [in front] of her, and talk to her about sex stories etc... she will hate it, but just seem oblivious to the fact u used to be together... she will be crawling back to u in no time...* (Quote 22).
>
> *lol, any chicks rule as long as they are not fat or sweaty* (Quote 23).

However, not all members of the forum shared such negative views on women and relationships. A small number of posts expressed their concern on the gender roles of women. These quotes showed that there was some pushback on the dominant public discourse in the forum, which reflected the masculine, even misogynistic, dynamics within the community: "*F*** those who think women are nothing but sex Machines. That you can beat them up anyway you want. They have an equal right in the society and I was never in that impression that they are weak*" (Quote 24). This speaks to current findings on the masculine nature of the hacker subculture (Holt, 2007; Jordan & Taylor, 1998; Steinmetz et al., 2019). However, there appears to be some negotiation and discussion on the definition of masculinity and acceptable masculine behaviours towards women.

Humour and sarcasm also played a prominent role within the samples (as has been well-established in studies of hacker communities) (Coleman, 2014). Some of these posts were misogynistic in character, such as sexualising women and making fun of violence against women. In her study of doing gender in online environment, Kendall (2000) found that participants expressed support towards hegemonic masculinity through the use of jokes and sarcasm about women's status as sexual objects. Hegemonic masculinity refers to the prevailing masculinity that legitimizes hierarchical gender relationships between men and women (Messerschmidt, 2009). More commonly, humour and sarcasm from the samples

portrayed women as the lesser gender. This is often tied to ideas about hacking and technical work, framing it as intrinsically masculine: "*A women meets god and after short conversation with him god grants her three wishes: She first wished to be the smartest women in the world... god granted this to her. Her second wish was to be 2 times smarter than she is... god granted this to her. Her third wish was to be 4 times smarter than she is... and god turned her into a MAN*" (Quote 25) and "*Hacking is like sex. You get in, you get out, and hope that you didn't leave something that can be traced back to you*." (Quote 26). With the first quote, men's intelligence, both general and technical, was framed as more superior than women's. In the second quote, women disappeared altogether when hacking and sex were associated together, suggesting the non-important role of women in the process and the masculine construction of hacking. In the posts we studied, users consistently framed hacking as masculine.

The association between gender and technology was present in normal discussions as well. There appeared to be the assumption that women are less knowledgeable with computer science and technical skills. This is evident when discussing women's roles as potential victims of hacking attacks: "*Chances are, your GirlFriend Is not a big Techy. And she wont know how they got deleted*" (Quote 27). In this sense, women were framed as perfect victims as they would fail to detect the attackers. Similarly, in a detailed tutorial on how to use social engineering, the poster made general assumptions on women's susceptibility to downloading external files: "*Claiming the exe is a picture (look above) will probably get most women*" (Quote 28).

The diversity in topics associated with gender suggest a shift in the focus and discourse of the forum and appears to be welcoming other views on masculinity. At the same time, discussions of these topics continue to sustain the biased views on gender. In general, there appears to be more tension between healthy masculinity and biased views on gender, suggesting that within an overall tone of misogyny in the community, there is minor push-back.

5 Discussion and Concluding Thoughts

Constructions of masculinity and femininity clearly play an important role in structuring the involvement and activities of at least a sub-group of the actors on the cybercrime forum which we studied. This research is both a deep qualitative study of attitudes, sensibilities, and norms, as well as a broad data science project which aims to establish these as reflective of broader dynamics and patterns within the forum community. We have engaged in this research as an interdisciplinary collaboration between computers scientists, psychologists, and criminologists. In this section, we discuss our findings and their broader relevance to criminological understanding of involvement in cybercrime.

How and why people become involved in cybercrime is an important topic of criminological research. Hacking forums have been identified as an important part of this ecology, where people can learn skills and become involved in communities, so studying them is of value. Previous studies have provided compelling evidence that online gaming may be a potential route into some forms of low-level cybercrime. This has gone on to shape policy responses, particularly in Europe, where organisations have developed programmes, such as the *National Crime Agency's (NCA's)* '*Cease and Desist*' project, which take a diversionary approach. These interventions target young people who start to show signs of engaging in cybercrime and low-level Computer Misuse Act offences.

Our findings suggest that, in fact, misogyny and toxic masculinity may be an even more important 'pathway' into cybercrime, and are clearly implicated in harmful behaviour within the gaming community as well (Quinn, 2017). We found clear evidence that the first interaction of many people with the cybercrime forum which we studied was spurred by a desire to spy on their intimate partner. These individuals come to the forum either asking for help, or trying to learn techniques and purchase tools to develop the capability to do this themselves. Although the discourse has later shifted towards request and help on social interactions with women in

the community, the nature of the discourse remained male-centric. Constructions of masculinity in these forums are clearly oriented towards dominant forms of masculinity and tied to the performance of male structural power. Once individuals become more acculturated to the life of the forum, there are a range of less-technical forms of cybercrime in which they can become involved. Few of these low-level or 'script-kiddie' kinds of cybercrime bear any resemblance to the technical work, which dominates their stereotyped view of hacking, instead making use of premade tools and basic social engineering.

Despite this, the idea of the 'hacker' as hard, technical mastery remains an important reference point for community self-identity, and one which is inherently bound to the construction of masculinity in this subculture. For these forms of crime, which can be moderately lucrative, gender and in particular the commodification of femininity, play an important role. That is in both initiation and monetisation well-beyond any technical elements. In many of these, women's sexuality and the performance of femininity is either directly exploited and commodified for financial gain (such as in eWhoring), used as bait for manipulation, or exploited for blackmail and harrassment (Hutchings & Pastrana, 2019). This sets up a striking tension within this forum community - they are bound to a self-image of heteronormative masculinity dominated by hard technical stereotypes of the male hacker. However, their everyday deviant activities often require the performance of femininity and a 'heightened' or stereotyped version of the heterosexual feminine role (and engaging in what might well be classed as sex work).

This tension is offset through the dominant discourse about women on this forum, in which misogyny appears to play an important part, with misogynist humour and 'locker room' jokes between boys. This fulfils a more important function in strengthening their involvement in the forum community and the hacker identity. Most of the actors on the forum, especially those who are newer to the site, have little technical expertise, and would struggle to fit their

own stereotypical construction of the 'hacker', at best using pre-made scripts created by others. Misogynistic humour plays an important role in buttressing their idea of themselves as a hacker and hence their claim to this subculture's construction of masculinity (despite their lack of technical ability). Construction of hacker identity within these communities is intrinsically gendered - many of the discussions we observed frame it as essentially masculine, the hard mastery and domination of technology. This hacker identity is put on a pedestal as an idealised construction of a particular (alternative) type of masculinity. Further, the objectification of women in this humour plays an important role in supporting the cognitive work and neutralisations required for them to turn their own constructions of femininity into a resource which they can exploit, through eWhoring and social engineering. As in Tanczer's extensive study of gender in hacktivist communities, this negation of women's hacking and technical ability serves to sideline them from 'hacker' identity (despite the many notable women hackers) and contribute to a restrictive construction of what forms of work constitute "real hacking" (Tanczer, 2016).

The misogynistic discourse and humour around femininity and women also ties well with Messerschmidt's (2009, 2013) development of hegemonic masculinity. In contexts where individuals or groups perceive their claims to masculinity to be threatened or challenged, there is a tendency to resolve to violence to assert masculinity over the perceived subordinate group. With the current study, misogynistic humour and discourse is an easier path towards the ideal identity of "hacker", especially for those less-skilled individuals in the community. For individuals who are more capable or closer to the "hacker" identity, such behaviours are unnecessary and may explain some of the push-back we observed in the results.

There are clearly many distinct ways in which gender is being drawn on by people in this community as a way of making sense of the kinds of 'technological' activity in which they are involved. It is important to note that this is only a partial perspective. We explicitly carried

out our qualitative and quantitative research on posts which mentioned women. Therefore, these findings are not necessarily reflective of the broader communities of interest and practice on this forum. What we have identified, however, is a substantial subsection of the life of this forum for which a 'toxic' form of masculinity is a crucial factor in initiation pathways, social norms, and the mechanisms through which they carry out their illegal activities. It is important to note that this is not reflective of 'hacking' in the round; in fact, very little of the phenomena we observe bear more than aspirational links to hacker practice. Crucially, our findings from the data science work establish that this is representative of a sizeable subgroup of activity on this forum, whose composition, kinds of activity, and internal differentiation accords with the characterisation we establish in our qualitative work.

Finally, we argue that the approach we have taken to this research represents a useful methodological innovation for conducting qualitative research on extremely large text datasets. Using a combination of data science and traditional qualitative methods, we took a very large database of several million posts and managed to conduct meaningful qualitative research on it. We suggest that our approach could prove useful for others attempting to do research on these large datasets of forum posts. These findings also have immediate policy relevance. While intervention approaches such as the NCA's Cease and Desist program attempt to divert people at risk of becoming involved in cybercrime, explicitly targeting gaming as a potential entry point, in fact, our analysis suggests that this may miss some important factors for the kinds of actors we observe. These interventions may not be addressing some of the root causes which lead to involvement in cybercrime (and which underlie much of the harms we observe) - misogyny and toxic masculinity.

We argue that, if such interventions are to succeed, it makes sense to include elements of relationship education and positive visions of masculinity and of women, explicitly

countering sexist narratives about women. Feminist pedagogy and activism may well be a crucial starting point for addressing low-level cybercrime.

References

Bachmann, M. (2010). The risk propensity and rationality of computer hackers. *International Journal of Cyber Criminology*, *4*(1&2), 643-656.

Barlett, C., & Coyne, S. M. (2014). A meta-analysis of sex differences in cyberbullying behavior: The moderating role of age. *Aggressive Behavior*, *40*(5), 474–488.

Blevins, K. R., & Holt, T. J. (2009). Examining the virtual subculture of johns. *Journal of Contemporary Ethnography*, *38*(5), 619–648.

Bossler, A. M., & Burruss, G. W. (2012). The general theory of crime and computer hacking: Low self-control hackers?. In T. J. Holt & B. H. Schell (Eds.), *Corporate Hacking and Technology-Driven Crime: Social Dynamics and Implications* (pp. 38-67). IGI Global.

Bowen, G. A. (2006). Grounded theory and sensitizing concepts. *International Journal of Qualitative Methods*, *5*(3), 12–23.

Caines, A., Pastrana, S., Hutchings, A., & Buttery, P. (2018a, October). Aggressive language in an online hacking forum. In *Proceedings of the 2nd workshop on abusive language online (ALW2)* (pp. 66–74). doi: 10.18653/v1/W18-5109

Caines, A., Pastrana, S., Hutchings, A., & Buttery, P. J. (2018b). Automatically identifying the function and intent of posts in underground forums. *Crime Science*, *7*(1), 19.

Coleman, G. E. (2004). The political agnosticism of free and open source software and the inadvertent politics of contrast. *Anthropological Quarterly*, *77*(3), 507–519.

Coleman, G. E., & Golub, A. (2008). Hacker practice: Moral genres and the cultural articulation of liberalism. *Anthropological Theory*, *8*(3), 255–277.

Coleman, G. E. (2011). Hacker politics and publics. *Public Culture*, *23*(3 (65)), 511–516.

Coleman, G. E. (2012). *Coding freedom: The ethics and aesthetics of hacking*. Princeton University Press.

Coleman, G. E. (2014). *Hacker, hoaxer, whistleblower, spy: The many faces of [a]nonymous*. Verso books.

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, *13*(1), 3–21.

Creswell, J.W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Sage Publications.

Creutzburg, R. (2017). The strange world of keyloggers - an overview, Part I. *Electronic Imaging*(6), 139–148.

Furnell, S. (2002). *Cybercrime: Vandalizing the information society*. London: Addison-Wesley

Halder, D., Jaishankar, K., & Jaishankar, K. (2012). *Cyber crime and the victimization of women: laws, rights and regulations*. Hershey, PA: Information Science Reference.

Hancock, B. (1998). Can you social engineer your way into your network? *Computer Fraud & Security*, *1998*(11), 12–13.

Holt, T. J. (2007). Subcultural evolution? examining the influence of on-and off-line experiences on deviant subcultures. *Deviant Behavior*, *28*(2), 171—198.

Holt, T. J. (2010). Exploring strategies for qualitative criminological and criminal justice inquiry using on-line data. *Journal of Criminal Justice Education*, *21*(4), 466–487.

Holt, T. J., & Morris, R. G. (2009). An exploration of the relationship between mp3 player ownership and digital piracy. *Criminal Justice Studies*, *22*(4), 381–392.

Hughes, L. A., & DeLone, G. J. (2007). Viruses, worms, and trojan horses: Serious crimes, nuisance, or both? *Social Science Computer Review*, *25*(1), 78–98.

Hutchings, A., & Chua, Y. T. (2016). Gendering cybercrime. In T. J. Holt (Eds). *Cybercrime through an interdisciplinary lens* (pp. 181–202). Routledge.

Hutchings, A., & Pastrana, S. (2019, June). Understanding eWhoring. In *2019 IEEE European Symposium on Security and Privacy (EuroS&P)* (pp. 201-214). IEEE.

Jordan, T., & Taylor, P. (1998). A sociology of hackers. *The Sociological Review*, *46*(4), 757–780.

Kendall, L. (1998). Meaning and identity in "cyberspace": The performance of gender, class, and race online. *Symbolic Interaction*, *21*(2), 129–153.

Lui, M., & Baldwin, T. (2010). Classifying user forum participants: Separating the gurus from the hacks, and other tales of the internet. In *Proceedings of the Australasian Language Technology Association Workshop 2010* (pp. 49–57).

Marcum, C. D., Higgins, G., Ricketts, M., & Wolfe, S. (2014). Hacking in high school: Cybercrime perpetration by juveniles. *Deviant Behavior*, *35*(7), 581—591.

Marcum, C. D., Higgins, G. E., Freiburger, T. L., & Ricketts, M. L. (2012). Battle of the sexes: An examination of male and female cyber bullying. *International Journal of Cyber Criminology*, *6*(1), 904-911

Messerschmidt, J.W. (1994). Schooling, masculinities and youth crimes by white boys. i: Stanko, Elizabeth & Tim Newburn (red.): Just boys doing business. Men, Masculinities and Crime.

Messerschmidt, J. W. (2009). "Doing gender" the impact and future of a salient sociological concept. *Gender & Society*, *23*(1), 85–88.

Messerschmidt, J. W. (2013). *Crime as structured action: Doing masculinities, race, class, sexuality, and crime*. Rowman & Littlefield.

Nafus, D. (2012). 'Patches don't have gender': What is not open in open source software. *New Media & Society*, *14*(4), 669–683.

Overdorf, R., Troncoso, C., Greenstadt, R., & McCoy, D. (2018). Under the underground: Predicting private interactions in underground forums. ArXiv preprint arXiv:1805.04494.

Pastrana, S., Hutchings, A., Caines, A., & Buttery, P. (2018). Characterizing eve: Analysing cybercrime actors in a large underground forum. In *International Symposium on Research in Attacks, Intrusions, and Defenses* (pp. 207–227).

Pastrana, S., Thomas, D. R., Hutchings, A., & Clayton, R. (2018). CrimeBB: Enabling cybercrime research on underground forums at scale. In *Proceedings of the 2018 World Wide Web Conference* (pp. 1845–1854).

Portnoff, R. S., Afroz, S., Durrett, G., Kummerfeld, J. K., Berg-Kirkpatrick, T., McCoy, D., Levchenko, K. & Paxson, V. (2017). Tools for automated analysis of cybercriminal markets. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 657–666).

Quinn, Z. (2017). *Crash override: How gamergate (nearly) destroyed my life, and how we can win the fight against online hate*. Hachette UK.

Raviv, E. (2020). Understanding pointwise mutual information in statistics. https://eranraviv.com/understanding-pointwise-mutual-information-(Accessed: 2020-06-08)

Rusch, J. J. (1999). The "social engineering" of internet fraud. In *Internet Society Annual Conference*, http://www.Isoc.org/isoc/conferences/inet/99/proceedings/3g/3g 2.htm.

Skinner, W. F., & Fream, A. M. (1997). A social learning theory analysis of computer crime among college students. *Journal of Research in Crime and Delinquency*, *34*(4), 495–518.

Southworth, C., Dawson, S., Fraser, C., & Tucker, S. (2005). *A high-tech twist on abuse: Technology, intimate partner stalking, and advocacy*. Violence Against Women Online Resources.

Steinmetz, K. F. (2016). *Hacked: A radical approach to hacker culture and crime*. NY, New York University Press.

Steinmetz, K. F., Holt, T. J., & Holt, K. M. (2019). Decoding the binary: Reconsidering the hacker subculture through a gendered lens. *Deviant Behavior*, 1–13.

Tanczer, L. M. (2016). Hacktivism and the male-only stereotype. *New Media & Society*, *18*(8), 1599–1615.

Taylor, P. (1999). *Hackers: Crime in the digital sublime*. London: Routledge.

Thomas, D. (2002). *Hacker culture*. Minneapolis, MN: University of Minnesota Press.

Turgeman-Goldschmidt, O. (2008). Meanings that hackers assign to their being a hacker. *International Journal of Cyber Criminology*, *2*(2), 382—396.

Yar, M. (2005). Computer hacking: Just another case of juvenile delinquency? *The Howard Journal of Criminal Justice*, *44*(4), 387–399.

Tables

Bigram	Frequency	Bigram	Frequency
year, old	6285	need, help	1710
hot, girl	2316	like, know	1292
like, girls	2225	thanks, advance	1133
girls, like	4271	want, know	1406
every, girl	1407	facebook, account	1007
girl, know	1652	social, media	578
social, engineering	883	make, money	1468

Table 1. Frequent Bigrams.

Collocation	Collocation	Collocation
United, States	Little, bit	Male, female
Social, engineering	Long, term	Friend, zone
Thanks, advance	Big, deal	Last, night
Social, media	Story, short	Months, ago
Video, games	Gender, male	Days, ago
Justin, Bieber	Weeks, ago	Long, distance
Year, olds	Makes, sense	Every, single
Smoke, weed	High, school	Age, gender
Join, [name redacted]	Phone, number	Lost, virginity

Table 2. Significant Collocations.

Topics	Contributing Words
Join group '[group name redacted]	people, want, **account**, computer, [**group name redacted**], active, number, part, hacker, online
eWhoring	engineering, know, person, try, **whoring**, day, start, facebook, able, still
Social Engineering methods	social, something, money, method, way, give, methods, getting, **traffic**, around
Information gathering	also, **need**, **information**, used, **skills**, could, lot, hacking, email, look
eWhoring tools (kik)	like, first, group, always, many, **kik**, phone, [name redacted], hack, community
Using rats to create eWhoring pack	things, send, going, **rat**, tell, work, started, sure, back, pictures

Table 3. Topics related to `Social Engineering'.

Topics	Contributing Words
Relationship Advice	advance, girls, could, kiss, give, great, nice, advanced, opinion, still
Hacking into account	like, top, girlfriend, account, something, made, last, home, pretty, computer
eWhoring	please, help, love, back, way, pack, man, hot, going, without

Table 4. Topics related to `Thanks in advance'.