# Chapter 28
# Hidden Platforms for Cybercrime: Experiences Investigating Darknet Services

**Matthew Ball and Roderic Broadhurst**

## 1 Experiences Investigating Darknet Services

The criminal use of the darknet entered the public consciousness in 2011 following a media publication about the infamous Silk Road 1.0 (Chen, 2011). The merging of sophisticated technologies created the perfect infrastructure for the development of illicit online markets and discussion spaces. Combining anonymous Internet browsing, cryptocurrencies, and global delivery systems, darknet markets (i.e. illicit cryptomarkets) emerged and transformed the retailing of illicit products and contraband, as well as the discussion of illicit activities. Doing research into this 'hidden' universe is fraught with challenges and the risk of exposure to abhorrent materials, hacks, and scams. Security and ethical requirements add further demands. This chapter provides a summary of what these darknet services (i.e. darknet markets and darknet forums) look like, the methodological techniques a researcher may use to study them, and the ethical and emotional challenges associated with studying these illicit spaces. The emergence of reliable and replicable data collection methods able to fully capture these evolving darknet markets and forums remains a work in progress.

This chapter describes our team's research on illicit darknet markets and darknet forums. Each market or forum exhibits their own design variations, quirks, and challenges that often require a customised approach to data collection. While darknet markets are rich in quantitative data, and we can produce descriptive statistics representing those data, it is difficult to reason *about* those data. This is due to the nature of the darknet platform, tracking users is inherently flawed, tracking purchases is likewise difficult, although in both cases, some data can be approximated. Darknet forums are qualitatively rich in textual data. However, observing and

M. Ball · R. Broadhurst (✉)
Australian National University, Canberra, ACT, Australia
e-mail: roderic.broadhurst@anu.edu.au



A. M. Díaz-Fernández et al. (eds.), *Fieldwork Experiences in Criminology and Security Studies*, https://doi.org/10.1007/978-3-031-41574-6_28

interpreting these forums is very labour-intensive, but reveals attitudes, trends, and habits among these 'hidden' populations.

The online world has long had a history of illicit transactions. The first reported online transaction took place in 1972 and involved a small amount of cannabis sent between students at Massachusetts Institute of Technology and students at Stanford University using their respective ARPANET accounts (Bartlett, 2014). This technological evolution reached its peak in February of 2011 with the launch of Silk Road 1.0 – introducing the modern darknet market.[1] These online marketplaces, accessible only through specialised software which connects to 'hidden' networks, are e-commerce style websites specialising in the sale and distribution of illicit products and contraband. Typical products offered on darknet markets include prohibited drugs, pharmaceuticals, fraudulent/identity documents, malware and exploit kits, counterfeit goods, and other contraband. Specialist markets offering pornographic material (including child sexual abuse material [CSAM]) and weapons occur. Darknet markets feature three main actors: vendors, buyers, and market administrators (Celestini et al., 2016). Buyers can leave reviews, message vendors, and dispute transactions. Vendors provide product listings with basic descriptions and details such as quantities, prices, and shipping services. Market administrators provide overall supervision of the service (e.g. dispute resolution) and market operation. Darknet markets levy a commission, typically ranging from 3% to 8%, from each sale, and provide escrow services.

Darknet markets have created problems for law enforcement agencies (LEAs) across the world. The combination of online anonymity, pseudo-anonymous transactions, and sophisticated stealth packaging has created a novel environment that inhibits LEAs ability to investigate the activities of darknet markets. In the past decade, LEAs have attempted to curb development and expansion of these darknet markets, and at least 20 (publicly acknowledged) law enforcement operations have focused on their suppression and remain to be fully explored by qualitative researchers. Due to the worldwide connectivity of the Internet, these operations have typically involved cross-jurisdictional cooperation between LEAs and collaboration in the form of intelligence sharing and joint operations.

## 2 History of Darknet Platforms and Research

The most common 'version' of the darknet is that found via the Tor[2] service network, although other networks exist for hiding communications and services such as i2p[3] (Invisible Internet Project) and Freenet.[4] Though research into these alternate

[1] Arguably the first iteration was 'The Farmer's Market' – launched around 2006 – but this market did not utilise cryptocurrencies.

[2] See: https://www.torproject.org/

[3] See: https://geti2p.net/en/

[4] See: https://freenetproject.org/index.html

darknets does exist (e.g. Hoang et al., 2018; Figueras-Martín et al., 2022), due to the limited size and diversity of these alternate platforms, the focus of this chapter is on the Tor service network.

## *2.1 The Darknet: The Onion Router (Tor)*

The Tor network was launched in September 2002 as 'The Onion Routing' network[5] (Dingledine et al., 2004). Since the Tor browser became available in 2008, the Tor network has grown rapidly. By 2022, each day between 2.04 and 3.6 million clients (not unique persons) accessed Tor via standard relays and about 150,000 accessed Tor via bridges. The Tor network hosts an estimated 700,000 unique onion addresses/onion services[6] (formerly hidden services). Although mostly associated with *web* services (i.e. web pages), onion services may offer SSH, BitTorrent, and SMTP/POP3 services, to name a few. Many onion services have multiple (unique) onion addresses, which direct users to the same service, so a unique onion address does not necessarily reflect a unique service.

The Tor system is an overlay network which uses 'mix network' routing between the website (or onion service) and user, ensuring neither knows the identity of the other. As a result, the identities and locations of the users of the network cannot be easily tracked due to the layered encryption systems used (Maras, 2014; Platzer & Lux, 2022). Originally designed by the United States Naval Research Laboratory for the purpose of protecting government communications, Tor is now '…used every day for a wide variety of purposes by the military, journalists, law enforcement officers, activists, and many others.'[7] The Tor project has acknowledged the potential for criminal abuse.[8] That said, while access to Tor for illicit products is frequent, it is not the most common use; many users access Tor to find clearnet (or open) sites that, for example, may be censored by their communications providers or prohibited by the jurisdiction in which they reside (Gehl, 2018). Jardine et al. (2020) estimated that about 6.7% of Tor users accessed the platform for illicit purposes, but note that this varies highly with the political context and location of the users.

[5] Although the project no longer uses this acronym and is just referred to as 'Tor' now.

[6] The Tor metrics website provides useful statistics on users and services (see: https://metrics.torproject.org/).

[7] See: https://www.torproject.org/about/torusers.html

[8] See: https://support.torproject.org/abuse/#abuse_what-about-criminals

### 2.1.1 Darknet Markets

Darknet markets are classified as either *omnibus* or *niche* markets. An omnibus or general market is defined as containing over 1000 product listings (or advertisements across different product categories) that are available on any given day, offered by at least several vendors. A niche or specialist market, on the other hand, is typically run by a single vendor and offers specific products. Omnibus markets are considered like a 'main street' or 'high street' market in the real world and offer a rapid assessment of the availability of illicit products and their prices. Niche markets focus on a limited set of products but may also offer 'specialist' products that are typically banned by omnibus markets, such as: CSAM and weapons.

Most darknet platforms require a user account to interact with the site's services and content. Current, 'third generation' darknet markets, have strong 'completely automated public Turing test to tell computers and humans apart' (CAPTCHA) protection and limit the duration of their market's login timeframes. Provided these logins can be refreshed, it is possible to automate the data collection process (see: *Automated techniques* below) to reduce the researcher interaction and associated time demands. This data capture process is limited by various factors and is further described in the automated techniques section below.

Collection of sales data remains a problem and often researchers need to undertake the recording of this via a time-consuming manual process. Nevertheless, evidence of actual sales transactions may be absent or scant, although when available offer further insight about markets, vendors, and product popularity. It is also apparent that some markets and vendors manipulate buyer feedback to create a favourable impression of the service provided by the vendor and/or market. This 'gaming' of buyer reputation or sales records and satisfaction feedback may be commonplace and can be difficult to detect.

Like onion services in general, the number of active darknet markets (or forums, see below) is unknown. Between 2010 and 2017, 103 darknet markets had been identified (Europol, 2018, p. 16), and between 2018 and 2020, an additional 43 markets have been reported (Ciphertrace, 2020). However, neither of these snapshots of the darknet market ecosystem was complete because of volatility and most ceased operations within a year (Europol, 2018, p. 16). In short, the volatile and dynamic nature of the Tor platform makes sampling imprecise and generalisation about darknet markets or platforms require caution.

There are a few notable websites and services (available in either the 'clear' or open net and the darknet) which act as 'dictionaries' listing the onion addresses for current high-street omnibus and niche darknet markets (Platzer & Lux, 2022). These services typically 'validate' the markets which they list and help guide both novice and regular users to the relevant active Tor address. Some of these sites, such as the notorious 'DeepDotWeb', took commission (i.e. affiliate marketing) from sending visitors to the respective darknet markets. We may never know the exact number of active darknet markets at any one time; however, we have a good idea of the most popular markets at that time.

### 2.1.2 Darknet Forums

Online discussion forums have existed in one form or another since the advent of the Internet. Prior to the introduction of the world-wide-web (WWW), the Internet featured online 'bulletin board systems' (BBS). A user could connect to a BBS server and exchange news, information, messages, and software with other users. By the early 1990s, with the launch of the WWW (Berners-Lee, 1989), BBSs rapidly evolved into what is now known as online discussion forums.

Virtual communities are congregations of individuals in the online world (Rheingold, 1993). These community groups are often formed around shared ideological, or specific, interests. Forums may also be generally offering a range of topics and discussion formats. 'Lurking' among, observing, or participating in these communities enable qualitative digital ethnographic study of the group or subculture (i.e. the study of the virtual community and the respective virtual culture(s) this platform has created, through a detailed analysis of the computer-mediated social interactions).

Since the mid-2000s, deviant communities have come together on anonymous platforms, often hosted on privacy-oriented protocols (e.g. Tor). These communities discuss illegal topics, such as narcotics and drugs, cybersecurity and hacking, and child sexual abuse topics. The illicit nature of their discussion often means that these communities are 'hidden' and that studying their behaviours is difficult; data captured from these darknet forums provide an insight into these otherwise elusive communities. Darknet forums are also sites where the attitudes and emotions of users are observable and can be traced over time. The inherent anonymity of these forums allows less inhibited discussions.

Online discussion forums are asynchronous (i.e. not in real-time), and their structure is hierarchical in nature: an index page lists one or more subforums, each subforum contains one or more threads, and finally, each thread contains one or more posts (Holtz et al., 2012). The asynchronous nature of communication enables users to not only respond at a time which is convenient for them, but also enables them to carefully consider and construct their response. The content (i.e. messages) from these online communities is 'relatively authentic natural data' (Holtz et al., 2012, p. 56). Specifically, this is the community behaving in an unobscured fashion; when dialogue between users is encouraged, these forums enable the community to discuss their lifestyles, share information (i.e. experiences), and facilitate communication in less inhibited ways.

Often arranged thematically, a subforum will contain many different conversations (or 'threads'). A thread consists of an initial message (a 'post') where a user starts a new discussion by asking a question, describing an experience, or requesting advice. Other forum users contribute to this thread by posting replies. Each thread contains one or more posts, a topic for discussion (a title), and a unique thread ID. Each forum post contains a message together with the post's metadata – the forum user who authored the post, the date and time it was posted, the unique post ID, and so on.

An online discussion forum can be classified as *open*, *restricted*, or *closed*. An open forum allows users to participate and observe material without providing any registration details (in common usage, users post under the alias 'Anonymous'). Restricted and closed forums require user registration. Registered users will 'log-in' to the forum, confirming their identity with a username and a password. A restricted forum allows a user to observe discussions but requires user registration to participate (i.e. create new posts) in the forum. A closed forum requires user registration to observe and participate in the discussions. Restricted and closed forums typically allow private messages (PMs) between registered users (Holtz et al., 2012).

A user hierarchy (i.e. structure) forms around online discussion forums. Accounts may be typified as: users; moderators/administrators; or owners. These latter user types have elevated privileges and may moderate content, ban users, and create new subforums (if necessary). Discussions on these forums are generally targeted and remain 'on-topic' due to the efforts of the administration and moderation team(s).

## *2.2 Previous Research of Darknet Services*

Research on the Tor network has focused on de-anonymisation and security concerns (Saleh et al., 2018), although, more recently, there has been a focus on illicit drug markets by criminologists and drug and addiction scholars (e.g. *International Journal of Drug Policy*). The proportions of onion services that are potentially criminal have been periodically estimated and generally about a third of potential onion services cannot be classified either because they were down, empty, or locked (i.e. a login was required). One study estimated that 20% of 10,367 known Tor addresses are 'suspicious', 48% 'normal' (offering hosting and cryptocurrency services), and 32% were classified unknown (Al-Nabki et al., 2019, p. 217). Biryukov et al. (2014) estimated that of 2618 English language Tor addresses, 44% of the onion services surveyed were platforms offering illicit drugs, pornography (including possible CSAM), various counterfeits products, or weapons; however, 31% could not be identified. Moore and Rid (2016) estimated 57% (from 2723 known Tor addresses) of the active onion services they identified were illicit services.

We also attempted to describe the 'public facing' (i.e. web-based) Tor services in our ongoing research on the scope of services on Tor and broadly classify 25,913 onion addresses identified by our crawler on August 16, 2021. This snapshot compares with the 47,230 onion addresses (not accounting for duplicate services) indexed on the popular Tor (user-submitted and compiled) listing service 'Fresh Onions', although a third of its links were estimated to be inactive. Just over half of the addresses we identified were distinct services (51.5%; $n$ = 13,342) and 61.3% were active ($n$ = 8184). Of the active distinct services, 6.1% could not be accessed (e.g. required a login or token) and classified. Darknet markets comprised only 5.7% ($n$ = 463) of the active sites, whereas unmoderated pornography (including CSAM and other illicit sexual images) accounted for 21.9% ($n$ = 1791), cryptocurrency (i.e. Bitcoin and other exchanges, mixers, mining, wallet services, money

laundering) 13.7% ($n$ = 1119), counterfeit products (i.e. fake currency, credit card markets, stolen data, credentials, passports, and other documents) 31.5% ($n$ = 2577), malware services (e.g. hacking, DDoS, botnets, ransomware, etc.) 15.2% ($n$ = 1247), and security (i.e. bullet-proof hosting and darknet market operational security) 1.7% ($n$ = 140) (Broadhurst & Ball, 2021).[9]

## 3 Methodological Techniques for Data Collection and Analysis

Both quantitative and qualitative analyses are useful in the study of darknet platforms. Analysis that describes the products listed on markets, sale volumes, prices, and quantities sold on or the content of forums may be augmented by a focus on vendors or forum posters and their diversity and cross-market or platform activity. Analysis may focus on the impact of police takedowns or sustained DDoS on market diversity, the behaviour of users and retailers (e.g. vendors), or attitudes revealed about illicit activity in forum posts. We noted above that the importance of the reliability and accuracy of the data captured depended on the efficiency and reproducibility of the data collection process, and we describe reliable and reproducible data collection methods.

The decision about what data to capture has implications for the data analysis phase as insights can be skewed based on the available data. We endeavour here to capture the entirety of a darknet market and describe all the available variables that can be identified such as: product descriptions, including quantities and prices; vendor information, including handles; PGP keys; shipping methods; and buyer feedback where available. Forums can also be crawled and offer a substantial corpus of text data that may reflect attitudes, emotions, grievances, illicit desires, and so on of the users and virtual community the forum represents.

### *3.1 Automated Techniques*

Data capture from websites is made easier when utilising automated 'crawler' and 'scraper' technologies. A 'crawler' is an automated script designed to search an entire website in a methodical manner and find as many unique pages as possible. This crawling process creates a static copy of the websites for later analysis. Copying the website in such a manner produces a timestamp for the time of capture, as well as retaining the structure of the page, allowing users to navigate the static site as if browsing in real time (Dolliver & Kenney, 2016). This is particularly important in

[9] A small number (201; 2.46%) of non-English onion services were identified and 1.76% ($n$ = 144) of onion sites comprised services not included elsewhere.

the case of volatile darknet markets. This method additionally creates historical records for later use. A 'scraper' implements a technique for extracting (i.e. 'scraping') data from hypertext markup (HTML) pages. This data is exported to a database for later analysis (e.g. a 'comma separated value' [CSV] file provides a 'flat database'). The database can then be used with different statistical programs (e.g. STATA, SPSS) to import and use the data.

### 3.1.1 Crawlers

Internet (or web) crawling has a long history (see: Mirtaheri et al., 2014). Arguably the first web crawler was implemented by Repository-Based Software Engineering the 'RBSE spider' in 1994 and rapidly became a widely used tool (Eichmann, 1994). The best-known Internet crawler was described by Brin and Page (1998) and implemented as the Google search engine. While crawling the clearnet is no longer a challenge, there are difficulties that arise in the context of crawling onion services and darknet markets or forums. Additionally, the intention of the crawler is also different. Most crawlers are designed to map out a domain of (unknown) sites. However, crawlers for darknet markets or forums typically identify only the domain to be crawled. The necessary components of a web crawler are (Alkhatib & Basheer, 2019a, p. 56): (1) The crawling space (i.e. the domain of information to be retrieved); (2) The processing of the website (i.e. the registration, login validation, and exporting of session cookies); and (3) The storage and analysis (i.e. processing) of static HTML pages.

Numerous studies have explored darknet platforms and focused on the products advertised (especially illicit drugs) and the ways in which sellers, market operators, and users interact through these onion services (Christin, 2012; Décary-Hétu & Aldridge, 2015; Soska & Christin, 2015; Aldridge & Décary-Hétu, 2016; Bancroft & Scott Reid, 2016; Barratt & Aldridge, 2016; Van Buskirk et al., 2016a, b). In earlier studies, data collection was not automated, but was copied manually from web listings, thus limiting the scale and frequency of data captured. Christin (2012; see also Soska & Christin, 2015) provided a thorough description of a simple method used to automatically collect data from Silk Road 1.0 by adapting a website copier 'HTTrack' (run through the 'torify' command) and exploiting the authentication cookies on Silk Road 1.0 to avoid CAPTCHA measures without the need to constantly provide a fresh login.[10]

The general challenges of crawling are scalability, management (i.e. login and cookie management), and obligations to avoid harm. The domain of websites to crawl is massive and accounting for this scope requires a crafted solution; additionally, when running a crawler, one must be cautious not to cause a DDoS attack against the website (Alkhatib & Basheer, 2019a, p. 56). Furthermore, there are unique challenges of crawling onion services. Onion services are typically volatile;

[10] HTTrack: http://www.httrack.com available since 1998 and cited in Christin (2012, p. 215).

there are accessibility concerns with CAPTCHA and automating the login/cookie retrieval process; and the operational security capability of the administrators who run these onion services (Alkhatib & Basheer, 2019a, p. 56). These problems are exacerbated in the context of darknet markets.

Daily collection or census 'snapshots' of darknet activity may be preferred but is less common than weekly or monthly data sweeps of known darknet markets. However, other than pragmatic reasons (i.e. time and cost), the rationale for different timeframes is unclear. Soska and Christin (2015) suggest that data capture should be complete, instantaneous, and frequent, but note several difficulties such as avoiding censorship, the variable time span different markets need to be fully or completely crawled (over hours to days), and the unpredictable availability of some darknet markets. More generally, problems can include the impact of law enforcement activity (e.g. takedowns or other disruption), DDoS attacks from competitors or other actors, and exit scams. Each of these events can have an impact upon data collection or even derail research, especially those that seek to describe trends. Service disruption is common and can lead to a sustained period of inaccessibility, or the end-of-life for the service. Consideration must be given for what period of inactivity represents an end-of-life – we arbitrarily determined that 2 weeks without activity represent end-of-life. From that moment on, the service is terminated, and we maintain the historical data for comparison purposes. It can be useful to find topical discussion about these ceased services on darknet forums. These discussions often explain the inactivity in detail (i.e. was it due to law enforcement takedown, a DDoS disruption, an exit-scam, or voluntary closure). Darknet crawlers, such as 'Datacrypto' described by Décary-Hétu and Aldridge (2013, 2015), may be available upon request. An early version of a crawler and scraper tool was described by Gwern Branwen (2019), who also provided the data obtained on several onion services between 2013 and 2015. Celestini et al. (2016) and Soska and Christin (2015) describe their methods in detail, but other studies do not always describe in sufficient detail the methods used, or problems encountered. Alkhatib and Basheer's (2019b) implementation ('Darky') does solve CAPTCHAs and automates the login step (where available), but does not download entire pages – the crawler simply downloads and processes the relevant data. However, 'Darky' was applied to a single (unnamed) market and included techniques for scraping.

This crawler and scraper implementation allows the extraction of information relevant to both products and vendors. The kind of information extracted includes: (1) products (e.g. title, vendor, country of origin, country of destination, price (often in USD), product views, product sales, product date (i.e. when it was put up as a listing), and product category); and (2) vendors (e.g. name/alias, level, trust, positive feedback, join date, and number of sales). In practice, not all markets feature the same elements, and specific designs may not easily generalise across markets.

Typically, an automated web-crawler will have a simplified user agent string to identify it and allow communication between a site that has been crawled and the users of the web-crawler. This user agent string changes with every other release of the Tor browser package and must be updated in the crawler accordingly. By convention, an automated crawler will feature the string 'bot', identifying it as an

automated bot.[11] An automated web crawler also abides with the rules known as the 'Robots Exclusion Standard' outlined in the markets' robots.txt file (if the market contains such a file). As this file is necessarily public knowledge, if a market does not contain a robots.txt file, the assumption is that those markets allow unrestricted crawling of their websites.

Researchers at the University of New South Wales National Drug & Alcohol Research Centre (NDARC) have developed an automated crawler to capture relevant and monitor trends (Mathur et al., 2020). This focuses on a subset of products; in particular, 'illicit drugs (e.g. heroin), key licit drugs (e.g. alcohol, tobacco, e-cigarettes), and pharmaceutical medicines, as well as drug-related paraphernalia (e.g. needles and syringes, reagent test kits)' (Mathur et al., 2020, p. 1). This method includes an unsupervised artificial intelligence (AI) model for product classification (according to the Anatomical Therapeutic Chemical classification system devised by the World Health Organisation) and solves the onion service CAPTCHA automatically (i.e. robot exclusion), while offering researchers the capability of observing trends in real time. Such enhancements demonstrate the feasibility of monitoring darknet platforms at scale. NDARC hosts an interactive dashboard which features and discusses their findings.[12]

### 3.1.2 Scrapers

The previously defined crawlers have a corresponding scraper to collect the information from the captured hypertext markup (HTML) files and export the information to a human readable format (e.g. CSV format). Once the data are exported to the database format, a 'data-cleaning' process is undertaken to remove superfluous material and standardise text and other data values before the analysis phase.

Although scrapers are often implemented ad-hoc, depending on the darknet market or forum, a subset of common features routinely appear, which can be recorded in the corresponding database. At a minimum, each record ought to include the: product name, vendor name, price in BTC,[13] price in $CURRENCY (calculated from the value of BTC at the time of capture), product URL, and vendor URL. Other information such as quantity sold or a vendor's feedback ranking can sometimes be present on specific markets and scrapers can be adjusted accordingly to include these data. Due to the non-standardised formats in darknet market design, data from multiple markets may feature incomplete or missing values (for instance, a data field from one market is absent in another). These missing values should be included in the data analysis process.

[11] The user agent string should provide not only the bot's name (identifying it as a bot/web-crawler), but also contact details (an email).

[12] See: https://ndarc.med.unsw.edu.au/resource-analytics/trends-availability-and-types-drugs-sold-internet-cryptomarkets-october-2021

[13] Bitcoin is no longer the cryptocurrency of choice among darknet markets due to the ability to trace Bitcoin transactions.

### 3.2 Quantitative Techniques

Darknet markets provide researchers with quantitative data that can be used to observe trends in products, prices of illicit drugs, and other contraband. Most research into darknet markets has featured quantitative modelling techniques (including our own; see reference list). In our experience, this is a rich source of information about the *supply* side of the illicit darknet trade (i.e. product availability). However, without a qualitative perspective, this information may lack context about product trends or demands and what products are actually sold or transacted in the darknet trade system.

### 3.3 Qualitative Techniques

Qualitative research has been undertaken on darknet markets, with a particular emphasis on the forums and the information centres that focus on onion services (Kamphausen & Werse, 2019). The most common qualitative data sources include comments and user ratings (Aldridge & Décary-Hétu, 2016), total numbers and categories of listings (Van Buskirk et al., 2016b), and a combination of these, considering listings and estimates of transactions on a particular market or several markets (Aldridge & Décary-Hétu, 2016). This analysis provides useful insight about profitability, availability, and the extent of repeat business for a particular market(s) and/or products. Although samples are usually small (and potentially unrepresentative), interviews with or surveys of darknet users (as well as vendors and market operators) can also provide insights as well as potential verification of activities, logistics, and the role of trust (Bancroft & Scott Reid, 2016; Barratt et al., 2014; Kamphausen & Werse, 2019; Martin et al., 2020).

Discussions about darknet markets frequently occur on market-related forums (or, in certain cases, related to the whole ecosystem). Topics discussed often focus on a market or vendors' reliability, market closures, law enforcement countermeasures, and product purity or value, among others. These platforms provide a rich source of qualitative textual material (amenable to some forms of discourse and narrative analysis) which can be used to help predict trends in products and vendor activities for specific darknet markets.

Examining the content of forums allows an approximate validation (or proto triangulation) of quantitative data acquired by a crawler. Distilling the often-vast text corpus of forums into content themes, poster or participant activities, and priorities is time consuming. We proceed by creating a coding template after close reading of the text using a narrative analysis framework attending to the functions and substance of the text narrative. Dialogic and structural elements in the text corpus are also present but vary with the format of forums and the researchers' inductive reflections are decisive as to relevance (see e.g. Mishler, 1995; Parcell & Baker, 2018; Holtz et al., 2012). Market-oriented forums tend to be instrumental, but

performative elements may also be present and can be central to forums that focus on moral or ideological content.

Coding data is a method for systematically categorising (i.e. labelling, organising) data to extract themes and patterns. While technological solutions do exist to extract patterns in the data (e.g. word clouds, word frequencies), manual coding is still required. The grounded, or inductive, approach to coding suggests that the researcher should start with their data to explore content or thematic analysis. When the researcher finds an element of data they wish to categorise, they create a new code and document the data as such. By the end of the analysis process, the researcher has created a coding scheme.

Technological tools, such as natural language processing (NLP) techniques, face difficulty with the 'informal language' commonly used in computer-mediated communications (CMC). Netspeak (e.g. emoticons and abbreviations), accent, non-fluencies, and filler words are challenges for NLP algorithms. NLP techniques such as topic modelling can be used to reduce large corpora to a set of common or repeated 'abstract topics'. However, the number of topics is coded manually.

## 4 Ethical and Emotional Issues Related to Darknet Research

This methodology for collecting data from darknet markets requires that ethical consideration of the implications and risks for darknet users, sellers, and others (including researchers) is merited (Barratt & Aldridge, 2016; Martin & Christin, 2016). Darknet forums present further difficulties for researchers; gaining consent of participants and maintaining the confidentiality and anonymity of the participants are paramount for the integrity of research. Researchers engaging in participant observation approaches with deviant subcultures on forums, especially anonymous forums, face the risk of being exposed to abhorrent content (see below for an example). The ethical issues associated with web scraping are briefly noted here; however, a broader discussion is offered by Martin and Christin (2016) and Brewer et al. (2021).

Given that potentially identifying data is not available to researchers of darknet platforms, the risk of legal harm to subjects observed is minimal; however, a full ethical review is recommended given incidental risks to researchers such as exposure to CSAM. This process can be laborious and time-consuming for the researcher(s). However, it is crucial considering the potential risk of exposure and the need to comply with laws regarding the mandatory reporting of CSAM. Collecting data from CSAM media sharing platforms requires special ethical approval due to the risk of moral injury to researchers exposed to abhorrent materials (e.g. sexual violence, torture, lethal violence). Even in the absence of visual media (e.g. videos and images), these communities share fictional and non-fictional stories of their desires and behaviours. This exposure can be quite confronting to even the most

seasoned researcher. For example, there are CSAM media-sharing communities hosted as darknet services. While these services trade in media (e.g. images, videos, stories), they also feature community discussion on the topic. Access to these communities may be blocked by a CAPTCHA which itself features CSAM; thus, any researcher(s) willing to engage with this material should be aware they will likely encounter this material, despite using image obfuscation. This highlights the importance of a support network for the researcher(s) as well as institutional ethical guidance and approval for such research. Our ethical protocol includes psychological support (a 'de-briefing' process), and if encountered, specific arrangements to enable safe and secure referral of abhorrent images to law enforcement.

Investigators must inform participants of their involvement in research, as their willingness to participate may depend on whether they know they are involved. In the context of a darknet forum, providing this information to participants is difficult. Open forums (where registration and login credentials are required) can be viewed as being in the public domain. This removes the requirement that the researcher obtain an informed consent before collecting data. In this way, individual contributions (i.e. the 'posts') can be viewed in a similar manner to individual naturalistic observations in a public space. This mitigates issues around participant agreement on open forums. The data 'captured' is the public record of the onion service contents at a particular time and protocols for human research subjects are not required. However, obtaining data from a closed forum without prior consent could be unethical and researchers should seek ethical approval for covert observation when this is necessary. In our research, we have not investigated closed forums. Obtaining consent from participants in a closed forum context is difficult (e.g. not all original users are likely to be present, yet their data will still be available).

Maintaining the three fundamentals of participatory research, namely, anonymity, confidentiality, and privacy, is important for methodological reasons. Researchers of darknet forums achieve this by (1) *Anonymity:* rebranding the users' alias (i.e. given a pseudonym in the write-up). This would ensure that no link can be established between specific responses and a specific forum user. (2) *Confidentiality*: as above with anonymity a forum users' identity and IP address is unavailable and unattributable in the Tor network; and (3) *Privacy:* data can be 'scrubbed' before being recorded (e.g. usernames, etc.), this ensures that anonymity is maintained. Literal (or direct) quotes attributable to a specific user are avoided in publication.

Christin (2012) also raised ethical concerns such as whether CAPTCHA bypass or 'workarounds' are 'hacks' and prevent platform operators from becoming aware of the crawl by re-pooling the Tor relay if regularly accessed over time. Christin reported the Silk Road 1.0 platform administrators did not contact him: either because they did not detect the crawl (nor was the possibility discussed in the relevant forums) or if they did so they did not regard it as a problem. Our approach was to include a user-agent string, a means of enabling a website to identify the device (and users) accessing that website and like Christin no market administrators contacted us.

## 5 Conclusion

Darknet platforms are volatile and can experience significant or prolonged downtime due to load on Tor relays and bridges. Soska and Christin (2015), among others, note the ubiquitous connection failures typical of data collection on Tor. The absence of extensive cataloguing or indexing of darknet markets and other onion services prevents accurate sampling and inhibits generalisation about all markets or forums. Additional periodic reconnaissance and qualitative investigation are required to address representativeness and validity. We have found frequent attention to the content of relevant forums helpful in guiding research questions and identifying new and emerging markets and contraband products. Darknet platforms may be unavailable due to DDoS attacks (Cimpanu, 2019), exit scams (Greenberg, 2015), voluntary closures (Power, 2019), and occasionally hacks or de-anonymisation or seizure by law enforcement (Van Buskirk et al., 2016a, b; Europol, 2018). This often cannot be predicted ahead of time and can cause data collection processes to experience significant delays or fail altogether.

For general data capture on darknet services, there has been growing concern for the 'human' aspect as researchers are required to interact with illicit items. The scale of data from these onion services has increased significantly as the popularity of these darknet markets increases, resulting in a requirement for automation and classification methods for data analysis. While many classification models exist, specialised classifiers have become important due to the absence of a complete taxonomy capable of recording such data (Dalins et al., 2018b). Adding to this, illegal information such as CSAM can be prevalent on onion services and the use of automated classification methods could significantly reduce the number of images viewed by individuals who classify and investigate CSAM (Dalins et al., 2018a) and thus minimise moral injury. Regarding other illicit services available on Tor, it is also usual to program crawlers to avoid image capture where the risk of CSAM capture is likely and to inspect text with images obscured. However, in jurisdictions such as Australia, the collection (i.e. possession) of textual descriptions of CSAM may be unlawful and additional ethical and reporting arrangements are necessary to ensure the research is both lawful and ethical.

Researchers operating in this environment need to be aware of the risks associated with investigating onion services. These risks can potentially limit the kind of research which can be undertaken, but simultaneously highlight just how important it is to perform this kind of research.

**Acknowledgements** We gratefully acknowledge the assistance of Alexander Niven and Harshit Trivedi. In addition, we record our thanks to the Australian Federal Police division of the Australian Cyber Security Centre and the Australian Institute of Criminology.

## References


Aldridge, J., & Décary-Hétu, D. (2016). Hidden wholesale: The drug diffusing capacity of online drug cryptomarkets. *International Journal of Drug Policy, 35*, 7–15. https://doi.org/10.1016/j.drugpo.2016.04.020

Alkhatib, B., & Basheer, R. (2019a). Crawling the dark web: A conceptual perspective, challenges and implementation. *Journal of Digital Information Management, 17*(2), 51. https://doi.org/10.6025/jdim/2019/17/2/51-60

Alkhatib, B., & Basheer, R. (2019b). Mining the dark web: A novel approach for placing a dark website under investigation. *International Journal of Modern Education and Computer Science, 11*(10), 1–13. https://doi.org/10.5815/ijmecs.2019.10.01

Al-Nabki, M. W., Fidalgo, E., Alegre, E., & Fernández-Robles, L. (2019). ToRank: Identifying the most influential suspicious domains in the Tor network. *Expert Systems with Applications, 123*, 212–226. https://doi.org/10.1016/j.eswa.2019.01.029

Bancroft, A., & Scott Reid, P. (2016). Concepts of illicit drug quality among darknet market users: Purity, embodied experience, craft and chemical knowledge. *International Journal of Drug Policy, 35*, 42–49. https://doi.org/10.1016/j.drugpo.2015.11.008

Barratt, M. J., & Aldridge, J. (2016). Everything you always wanted to know about drug cryptomarkets* (*but were afraid to ask). *International Journal of Drug Policy, 35*, 1–6. https://doi.org/10.1016/j.drugpo.2016.07.005

Barratt, M. J., Ferris, J. A., & Lenton, S. (2014). Hidden populations, online purposive sampling, and external validity. *Field Methods, 27*(1), 3–21. https://doi.org/10.1177/1525822x14526838

Bartlett, J. (2014). *The dark net: Inside the digital underworld*. Random House.

Berners-Lee, T. J. (1989). *Information management: A proposal* (No. CERN-DD-89-001-OC). https://cds.cern.ch/record/369245/files/dd-89-001.pdf

Biryukov, A., Pustogarov, I., Thill, F., & Weinmann, R. (2014). Content and popularity analysis of Tor hidden services. In *2014 IEEE 34th international conference on distributed computing systems workshops*. IEEE. https://doi.org/10.1109/icdcsw.2014.20

Branwen, G. (2019, May 22). *Darknet market archives (2013–2015)*. Gwern.net. https://www.gwern.net/DNM-archives

Brewer, R., Westlake, B., Hart, T., & Arauza, O. (2021). The ethics of web crawling and web scraping in cybercrime research: Navigating issues of consent, privacy, and other potential harms associated with automated data collection. In A. Lavorgna & T. J. Holt (Eds.), *Researching cybercrimes methodologies, ethics, and critical approaches* (pp. 435–456). Palgrave Macmillan.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems, 30*(1–7), 107–117. https://doi.org/10.1016/s0169-7552(98)00110-x

Broadhurst, R., & Ball, M. (2021). *Tor's underworld, 'onion services' and child sex abuse material: Submission to the Australian Parliamentary Joint Committee on Law Enforcement inquiry into 'Law enforcement capabilities in relation to child exploitation'*. SSRN: https://ssrn.com/abstract=3927628 or https://doi.org/10.2139/ssrn.3927628

Celestini, A., Me, G., & Mignone, M. (2016). Tor marketplaces exploratory data analysis: The drugs case. In *Global security, safety and sustainability – The security challenges of the connected world* (pp. 218–229). Springer. https://doi.org/10.1007/978-3-319-51064-4_18

Chen, A. (2011, June 1). The underground website where you can buy any drug imaginable. *Gawker*. https://www.gawker.com/the-underground-website-where-you-can-buy-any-drug-imag-30818160

Christin, N. (2012). *Traveling the Silk Road: A measurement of a large anonymous online marketplace*. https://doi.org/10.21236/ada579383

Cimpanu, C. (2019, April 30). Dark web crime markets targeted by recurring DDoS attacks. *ZDNet*. https://www.zdnet.com/article/dark-web-crime-markets-targeted-by-recurring-ddos-attacks/

Ciphertrace. (2020). *Market timeline – Chart*. https://ciphertrace.com/wp-content/uploads/2020/10/CipherTrace-Market-Timeline-Chart.pdf

Dalins, J., Tyshetskiy, Y., Wilson, C., Carman, M. J., & Boudry, D. (2018a). Laying foundations for effective machine learning in law enforcement. Majura – A labelling schema for child exploitation materials. *Digital Investigation, 26*, 40–54. https://doi.org/10.1016/j.diin.2018.05.004

Dalins, J., Wilson, C., & Carman, M. (2018b). Criminal motivation on the dark web: A categorisation model for law enforcement. *Digital Investigation, 24*, 62–71. https://doi.org/10.1016/j.diin.2017.12.003

Décary-Hétu, D., & Aldridge, J. (2013). *DATACRYPTO: The dark net crawler and scraper. Software program*. https://www.research.manchester.ac.uk/portal/en/publications/datacrypto-the-dark-net-crawler-and-scraper-software-program(0d849f52-da3b-44bb-ab03-4dcf2a48f349)/export.html

Décary-Hétu, D., & Aldridge, J. (2015). Sifting through the net: Monitoring of online offenders by researchers. *European Review of Organised Crime, 2*, 122–141.

Dingledine, R., Mathewson, N., & Syverson, P. (2004). *Tor: The second-generation onion router*. https://doi.org/10.21236/ada465464

Dolliver, D. S., & Kenney, J. L. (2016). Characteristics of drug vendors on the Tor network: A cryptomarket comparison. *Victims & Offenders, 11*(4), 600–620. https://doi.org/10.1080/15564886.2016.1173158

Eichmann, D. (1994). The RBSE spider – Balancing effective search against web load. *Computer Networks and ISDN Systems, 27*(2), 308. https://doi.org/10.1016/s0169-7552(94)90151-1

Europol. (2018). *Internet organised crime threat assessment (IOCTA) 2018*. https://www.europol.europa.eu/internet-organised-crime-threat-assessment-2018

Figueras-Martín, E., Magán-Carrión, R., & Boubeta-Puig, J. (2022). Drawing the web structure and content analysis beyond the TOR darknet: Freenet as a case of study. *Journal of Information Security and Applications, 68*, 103229. https://doi.org/10.1016/j.jisa.2022.103229

Gehl, R. W. (2018). *Weaving the dark web: Legitimacy on Freenet, Tor, and I2P*. MIT Press.

Greenberg, A. (2015, March 18). The dark web's top drug market, evolution, just vanished. *Wired*. https://www.wired.com/2015/03/evolution-disappeared-bitcoin-scam-dark-web/

Hoang, N. P., Kintis, P., Antonakakis, M., & Polychronakis, M. (2018). An empirical study of the I2P anonymity network and its censorship resistance. In *Proceedings of the internet measurement conference 2018*. ACM. https://doi.org/10.1145/3278532.3278565

Holtz, P., Kronberger, N., & Wagner, W. (2012). Analyzing internet forums. *Journal of Media Psychology, 24*(2), 55–66. https://doi.org/10.1027/1864-1105/a000062

Jardine, E., Lindren, A. M., & Owensen, G. (2020). The potential harms of the Tor anonymity network cluster disproportionately in free countries. *PNAS, 117*(5), 31716–31721.

Kamphausen, G., & Werse, B. (2019). Digital figurations in the online trade of illicit drugs: A qualitative content analysis of darknet forums. *International Journal of Drug Policy, 73*, 281–287. https://doi.org/10.1016/j.drugpo.2019.04.011

Maras, M. (2014). Inside darknet: The takedown of Silk Road. *Criminal Justice Matters, 98*(1), 22–23. https://doi.org/10.1080/09627251.2014.984541

Martin, J., & Christin, N. (2016). Ethics in cryptomarket research. *International Journal of Drug Policy, 35*, 84–91. https://doi.org/10.1016/j.drugpo.2016.05.006

Martin, J., Munksgaard, R., Coomber, R., Demant, J., & Barratt, M. J. (2020). Selling drugs on darkweb cryptomarkets: Differentiated pathways, risks and rewards. *The British Journal of Criminology, 60*(3), 559–578.

Mathur, A., Bruno, R., Man, N., Barratt, M. J., Roxburgh, A., Van Buskirk, J., & Peacock, A. (2020). *Methods for the analysis of trends in the availability and type of drugs sold on the internet via cryptomarkets* (Drug Trends Bulletin Series). National Drug and Alcohol Research Centre, UNSW Sydney. https://ndarc.med.unsw.edu.au/sites/default/files/ndarc/resources/Methods_0.pdf

Mirtaheri, S. M., Dinçktürk, M. E., Hooshmand, S., Bochmann, G. V., Jourdan, G. V., & Onut, I. V. (2014). A brief history of web crawlers. *arXiv preprint arXiv:1405.0749*.

Mishler, E. G. (1995). Models of narrative analysis: A typology. *Journal of Narrative and Life History, 5*(2), 87–123. https://doi.org/10.1075/jnlh.5.2.01mod

Moore, D., & Rid, T. (2016). Cryptopolitik and the darknet. *Survival, 58*(1), 7–38. https://doi.org/10.1080/00396338.2016.1142085

Parcell, E. S., & Baker, B. M. A. (2018). Narrative analysis. In M. Allen (Ed.), *The SAGE encyclopedia of communication research methods*. SAGE Publications Inc. https://doi.org/10.4135/9781483381411

Platzer, F., & Lux, A. (2022). A synopsis of critical aspects for darknet research. In *Proceedings of the 17th international conference on availability, reliability and security*. ACM. https://doi.org/10.1145/3538969.3544444

Power, M. (2019, April 11). The world's biggest online drug market has shut down – What's next? *Vice*. https://www.vice.com/en_au/article/wjmw3w/the-worlds-biggest-online-drug-market-has-shut-downwhats-next

Rheingold, H. (1993). *The virtual community: Homesteading on the electronic frontier*. Addison-Wesley Publishing.

Saleh, S., Qadir, J., & Ilyas, M. U. (2018). Shedding light on the dark corners of the internet: A survey of Tor research. *Journal of Network and Computer Applications, 114*, 1–28. https://doi.org/10.1016/j.jnca.2018.04.002

Soska, K., & Christin, N. (2015). Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. In *Proceedings of the 24th USENIX security symposium ({USENIX} security 15)* (pp. 33–48). USENIX Association. https://www.usenix.org/system/files/sec15-paper-soska-updated_v2.pdf

Van Buskirk, J., Naicker, S., Bruno, R., Burns, L., Breen, C., & Roxburgh, A. (2016a, October). *Drugs and the Internet* (Issue 7). National Drug and Alcohol Research Centre. https://ndarc.med.unsw.edu.au/sites/default/files/ndarc/resources/Drugs%20&%20The%20Internet%20Issue%204.pdf

Van Buskirk, J., Naicker, S., Roxburgh, A., Bruno, R., & Burns, L. (2016b). Who sells what? Country specific differences in substance availability on the agora cryptomarket. *International Journal of Drug Policy, 35*, 16–23. https://doi.org/10.1016/j.drugpo.2016.07.004