Dark Web Markets: Turning the Lights on AlphaBay

Andres Baravalle(✉) and Sin Wee Lee

University of East London, London, UK
andres@baravalle.it

Abstract. Dark web markets have been, for a while now, at the centre of the attention of governmental bodies. After the rise and fall of The Silk Road, more and more marketplaces have followed the same path. Alphabay is the last of these markets to come under the spotlight, for its staggering amount of products on sale (estimated value of around $88M) during our analysis, and $590M of successful transactions during its existence.

In this paper we describe the spider that we developed for this project, our data analysis pipeline and the key findings of our data analysis.

The data analysis focuses on quantitative research, covering Alphabay's adverts, its sellers and its categories.

The outcomes of the research provide an insight on Alphabay, its general trend and the nature of its ads, which appear being a majority of drug products. United States and the United Kingdom are the most active countries on the website.

Keywords: Dark web · Dark web markets · Data wrangling · Text mining
Spidering

1 Introduction

Since the beginning of the web, illegal activities have been widespread. Popular software such as Napster, Kazaa, Gnutella, uTorrent and more recently Megaupload [1] have often being used for illegal file sharing. On-line auctions web sites and marketplaces are rife with fake goods and goods of questionable source. Web sites selling (and not delivering) tickets for event, planes or holiday packages is a recurring feature in newspapers.

In the modern world security and privacy are playing an increasingly important role, with numerous privacy related issues, such as governmental surveillance and corporate tracking facing modern Internet users. Such occurrences have led Internet users and software developers to deploy technology in order to protect their freedom and anonymity online.

The most widely known examples of such technology are Tor (The Onion Routing) and I2P (Invisible Internet Project), both volunteer maintained open source project which aims to protect end users by anonymising their Internet traffic by forwarding user data through a series of anonymous nodes.

H. Hacid et al. (Eds.): WISE 2018, LNCS 11234, pp. 502–514, 2018.
https://doi.org/10.1007/978-3-030-02925-8_35

Sites hosted on Tor (or I2P) can be visited by users, but it is hard to identify where they are hosted and who hosts them. This is what we call the dark web.

Anonymity and privacy are key characteristics of the dark web. Although dark websites are publicly visible, the IP addresses of the servers that run them are hidden (Egan n.d.). As a result, these websites can be visited by users, but it is hard to identify where they are hosted and who hosts them, and who are its users.

This is in contrast with the "surface" web and on the "deep web", where both users and servers are easily identified by their IP address. Content on the surface web is easily reachable with search engines; content on the deep web is not (e.g. password protected fora), but there is no anonymity in either case.

Dark web markets operate in anonymous networks, protecting the identity of all participants, buyers and sellers.

In the last years, dark web markets have quickly come to prominence as a safe haven where to carry out illegal transactions. Due to the difficulty of the Internet regulation and the convenience of online transaction, more and more illegal activities occur on dark web [20]. However, due to the initiative of the enforcement, these dark web sites, especially dark web markets, frequently shut down, and players come to prominence.

The size of the dark web is currently estimated to be between 7,000 and 13,000 websites "up and running" [17, 19]. The research carried out have identified about 30,000 websites, but most of them are short-lived (hence the lower number of web sites "up and running").

However, the estimation of the size of the dark web can be varied, depending on the method used, and researchers typically distinguish between websites online consistently enough to be investigated and short-lived websites.

In comparison, ISC reports 1.07 billion hosts in July 2017 (http://ftp.isc.org/www/survey/reports/current/) on the surface web.

Regarding the type of content hosted, according to INTELLIAGG's analysis, just under 70% of the dark web content is illegal (under UK and US law), while the remaining content would be legal. 30% of the illegal dark web are dark web markets.

The first dark web marketplace went online in 2011, and several dark web markets have risen to fame and have fallen later (notably: The Silk Road, The Silk Road 2.0, Evolution, Hydra, Agora).

Alphabay and Hansa were the largest dark web market places in 2017. Both have been shut down at the same time (July 2017) as a part of a law enforcement operation by the Federal Bureau of Investigation, the Drugs Enforcement Administration and European law enforcement agencies acting through Europol [14, 18]. On July 12, 2017, Alexandre Cazes, the founder of AlphaBay, allegedly committed suicide in the prison.

This paper presents research carried out on the data collected on the last weeks of life of AlphaBay, between June and September 2017.

The closure of Alphabay effectively ended the data collection.

Based on past experiences, it is widely believed that other web markets will take the place of AlphaBay (although that is not yet the case as we write in August 2018).

2 Experimental Environment

The first step in data collection from the dark web markets involved spidering the dark web markets [8]. The first proof-of-concept spider was developed with a few lines of code to simulate human authentication on the market. This was followed by a full spider, developed using Python and Amazon Web Services (AWS).

Building on top of lessons learnt and on expectations of an evolved landscape, we decided to:

- Implement a Wizard of Oz approach: non-headless browser (Selenium); to speed up the development loop and to reduce detection risks
- Agnostic software (doesn't care on what operating system runs), building on top of cloud technologies. Amazon Simple Storage was used to store (some) binary data, and MySQL was used to store textual data
- Use of Computer-to-Human interfaces
- A higher degree of anonymisation: disposable Alphabay accounts were created and used for the spidering, with a new account created after each spidering session.

Regarding architecture, we have developed a focused, vertical spider.

The spider run over a period of 6 weeks, with incremental improvements. Screenshots were taken automatically, which greatly helped in the ex-post examination, once Alphabay was down.

Critical issues around performance were immediately evident. Multi-thread execution and horizontal scaling are typically options used to increase the performance of web spiders; one of the shortcomings of our system was its limited possibilities for horizontal scaling (Fig. 1).

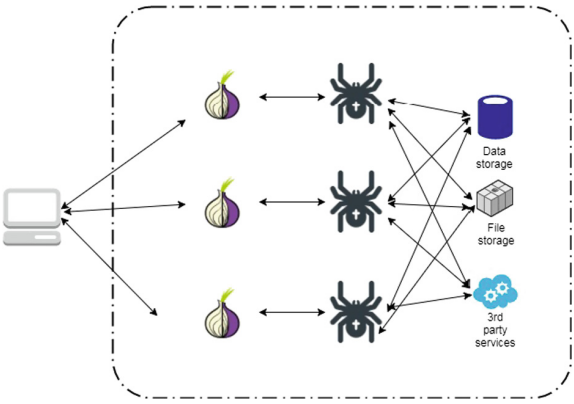

Fig. 1. Spider architecture

A single spider in our development box collected about 2,000 ads per day (using a single instance; more using multiple instances), which means that a full spidering if no new products added, would have taken about six months. Surface web spiders using

multiple threads and multiples proxies can easily reach 200 pages per minute, with just 5 proxies [1].

Spidering dark web markets requires researchers to heavily invest time to avoid detection during the spidering. Anti-DOS protections and anti-spidering protections (e.g. CAPTCHAs) limited heavily the speed of spidering.

Disposable accounts were created for each spidering session, and the IP address was always refreshed. In our project, each spider used a different Tor connection, and a different disposable user.

In order to address performance issues, multiple spiders were running at the same time as soon as the spider stabilised. No local storage was used; all data was stored on AWS in order support quick deployment and data collection.

2.1 Data Analysis Tools

Different technologies and programming languages that can be used for data analysis. Python, SAS and R (Anon 2017) are all valid and popular solutions, typically attracting support from different communities.

The data analysis presented in this paper has been performed using a combination of Python and R, with some exploratory data analysis done in SAS and Microsoft Access.

R and Jupiter (Python) Notebooks have been used through the data analysis to discuss the results and draft this paper (Anon n.d.).

2.2 Data Retrieval

It is important to highlight that Alphabay had a limitation on the number of pages that can be visualised per category (50). Essentially, this means that older ads are typically hidden (unless the user would perform a keyword search rather than browse the side).

Just over 380,000 ads were recorded to be present on Alphabay, both from screenshots and from other sources. Given the technical constraints, the number of ads that could be spidered at the time was just over 180,000. In most categories, the spidering could not retrieve ads before 2015 (Table 1).

The coverage of the data extracted is well over 10% in most categories, with few exceptions. One is “Other listings” (very small category, 3% coverage) and the others are Drugs and Chemicals (6%) and Fraud (7%). This coverage is based on the theoretical maximum (380,000). However, in reality, the users could not access more than 180,000 ads, which would give nearly double coverage.

For each advert, title, description, price (in USD), URL, seller, payment, origin, destination, category, collection timestamp, date of posting and number of products sold were collected.

Before analysis, data was prepared, by:

1. Removing special characters and switching to lowercase
2. Finding in the title or description of the ads the amount (number and mass) of the product

Table 1. Coverage per category

Category	Sample	Products	Coverage
Carded items	504.00	3,997.00	12.61%
Counterfeit items	2,658.00	10,098.00	26.32%
Digital products	2,908.00	18,941.00	15.35%
Drugs & Chemicals	16,536.00	259,399.00	6.37%
Fraud	3,725.00	48,046.00	7.75%
Guides & Tutorials	3,531.00	16,802.00	21.02%
Jewels & Gold	524.00	1,894.00	27.67%
Other listings	138.00	4,344.00	3.18%
Security & Hosting	264.00	903.00	29.24%
Services	1,376.00	8,182.00	16.82%
Software & Malware	1,205.00	3,653.00	32.99%
Weapons	614.00	5,562.00	11.04%

3. Calculating the price of one unit of one dose (1 g) for each transaction for the "Drugs and Chemicals" category.

A small number of outliers had drastically affected our initial analysis. To start, a "manual" review of the worst outliers showed some issues around data quality.

First of all, sellers wanting to retire a product typically increased the price to absurd, non-market values (an indicated in the ad that the product was not on sale any more). Manually analysing all products on sale for over 8,000 USD resulted in identifying 535,000 USD worth of retired products. Secondly, a number of "real" products where on sale at 0 USD – possibly as "feedback factories" and this had to be considered when deciding what to do with null values. Finally, a number of extreme values where actually genuine products.

While deciding to identify and remove extreme values, likely to be "dirty" data, we had to deal with the fact that the distribution of the data across most subcategories in the dataset is very long-tailed.

Different techniques have been tried on the data: z-score, boxplot (using 2 and 3 times the interquartile range) and percentiles.

For this analysis, we have decided to remove as little data as possible, slicing (by price) at the top 0.5^{th} percentile and at the bottom 99.5^{th} percentile.

3 Alphabay

At the time it was shut down, AlphaBay was the largest dark web marketplace. Its reputation can be confirmed by referring to the Google search statistics with the keywords AlphaBay and Dream Market (another popular dark web market) between January 2015 until June 2017 (Anon n.d.) (Fig. 2).

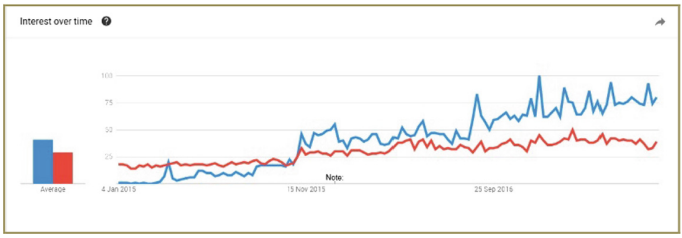

Fig. 2. Evolution of AlphaBay and Dream Market Google researches (Color figure online)

AlphaBay (in red in the graph) has become more and more popular since the demise of Agora, and before being shut down, it was the most popular dark web market (Anon 2017).

This assumption can be reinforced by looking at the data collected on AlphaBay. The number of ads per month (in the samples collected) from January 2015 until June 2017 shows the growing popularity of Alphabay (Fig. 3).

Fig. 3. Evolution of AlphaBay web market

Between 2015 and 2016, there was a significant jump, the number of ads (in our sample) rose from 7,712 up to 14,161. The number of ads that have been posted during the six first months of 2017 (before the closing) is 12,878, which is almost the same that in the whole 2016. If AlphaBay would have stayed online, the number of ads would reached a peak of 4,000/month by the end of the year.

Estimating the value of the products posted in the market and of each of the categories is key to a better understanding of dark web markets

As it can be seen in Table 2 we have mined products for a total value of over 6M USD (after removing extreme values); our estimates put the total value of the

individual products on sale at the time between 79 and 88 million USD (depending on the modelling technique).

Table 2. Market estimates (in USD)

Category	SumOfPriceSample	SumOfPRiceEstimate
Carded items	24,940.00	197,788.00
Counterfeit items	431,625.00	1,639,785.00
Digital Products	30,404.00	198,034.00
Drugs & Chemicals	4,413,189.00	69,229,367.00
Fraud	134,297.00	1,732,197.00
Guides & Tutorials	44,168.00	210,170.00
Jewels & Gold	197,899.00	715,307.00
Other listings	3,986.00	125,472.00
Security & Hosting	10,522.00	35,990.00
Services	227,700.00	1,353,955.00
Software & Malware	246,919.00	748,544.00
Weapons	402,712.00	3,648,020.00
	6,168,361.00	79,834,629.00

3.1 Ads Distribution by Category

Alphabay included 12 main categories; "Drugs and Chemicals" is the largest one, representing about 45% of the global market (Fig. 4).

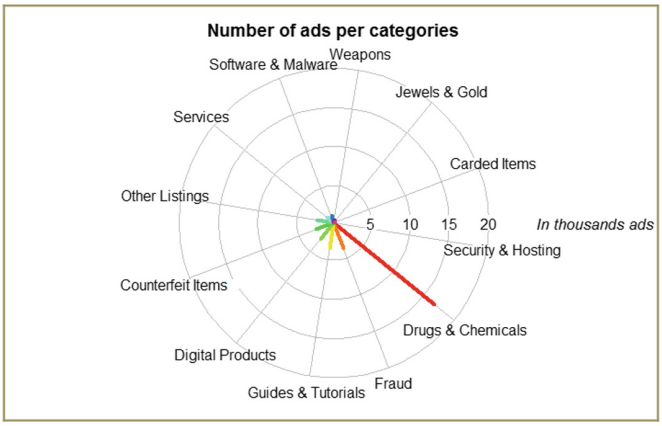

Fig. 4. Distribution of the market

The second most popular category is "Fraud", including all the ads regarding impersonation, fake ids and accounts, and representing about 13% of the market.

Finally, all other items (digital product, weapons, jewellery etc.) represent a small portion of the marketplace. It is worth mentioning that although weapons contributed to just a small portion of the listings, the total value of the listings is about $3.6M, with the United States as the main sellers, followed by UK. These two countries together, representing 97% of the global market.

3.2 Ads Distribution by Country

Figure 5 represents the 10 main countries in the world based on the number of adverts. As we can see, United States is the largest country by number of adverts, with more than twice as the number of ads than the second one, United Kingdom.

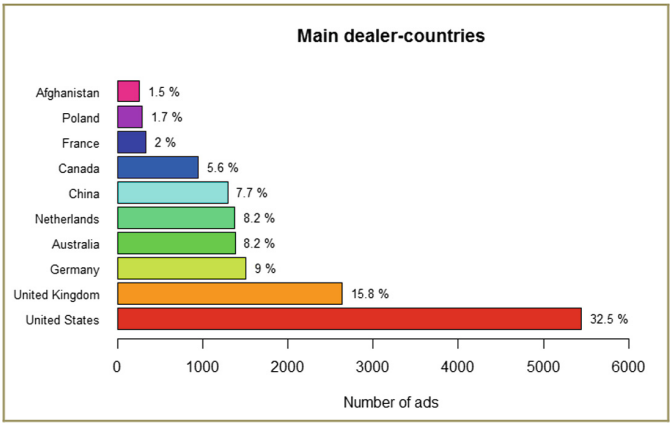

Fig. 5. Countries by number of adverts

Moreover, it is noticeable that most of these countries have strong economies. Five of the top 10 countries belong to the Group of Seven (G7), only Japan and Italy are not present. Top markets

In the next section, a more in-depth products analysis is presented on the top three sellers countries, United States, United Kingdom and Germany, plus a country of interest, Afghanistan.

A high-level analysis of some peculiarities of the specific markets is also included at the end of the section.

3.3 Products from the United States

The United States have the most diversity in terms of the products listings on AlphaBay. Its sellers trade products ranging from Drugs and Chemical (74%), Jewels and Gold (10%), Services (4%), counterfeit items (8%) and Weapons (2%).

Comparing with other countries, United States have the one of the highest concentration of Drugs and Chemicals. 29% of this are Cannabis and Hashish, followed by Psychedelic drugs.

The United States are also the largest market for weapons (Fig. 6).

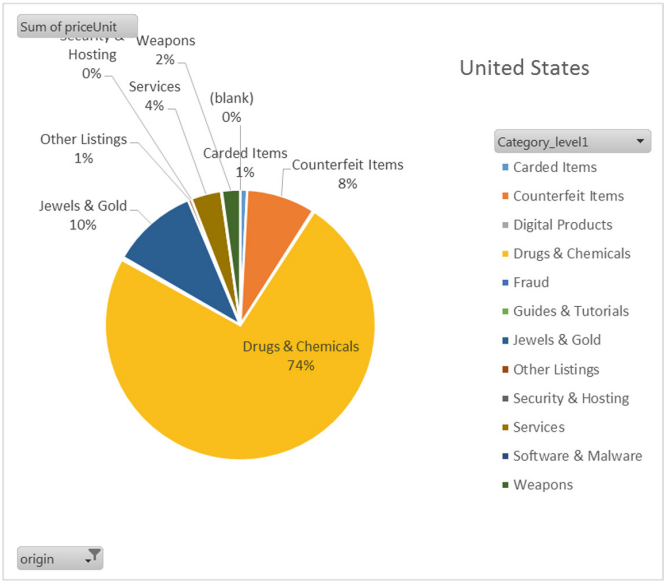

Fig. 6. Products from the United States, market distribution

3.4 Products from the United Kingdom

The United Kingdom, as the United States, have an extensive degree of diversity in terms of products listings. A huge portion of the products are in the "Cannabis & Hashish" category, but stimulants and other illegal drugs are significantly present as well (Fig. 7).

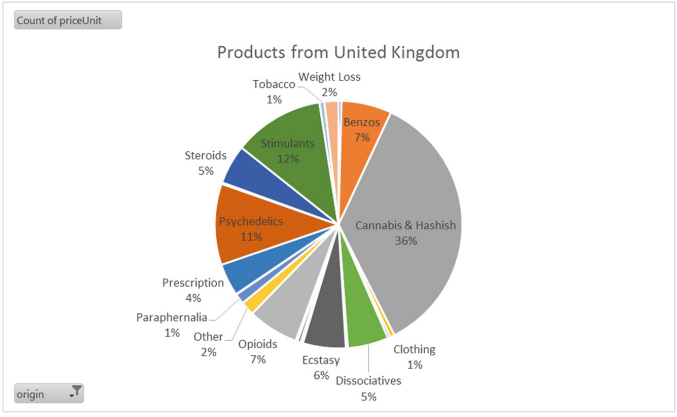

Fig. 7. Products from the United Kingdom, market distribution

3.5 Products from Germany

As for Germany, the country has very large number of products on sale in the category "Cannabis & Hashish", as the United Kingdom. Most of the European countries follow a similar pattern.

In Germany, the other most popular products are in the "Disassociatives" and "Psychedelics" categories (Fig. 8).

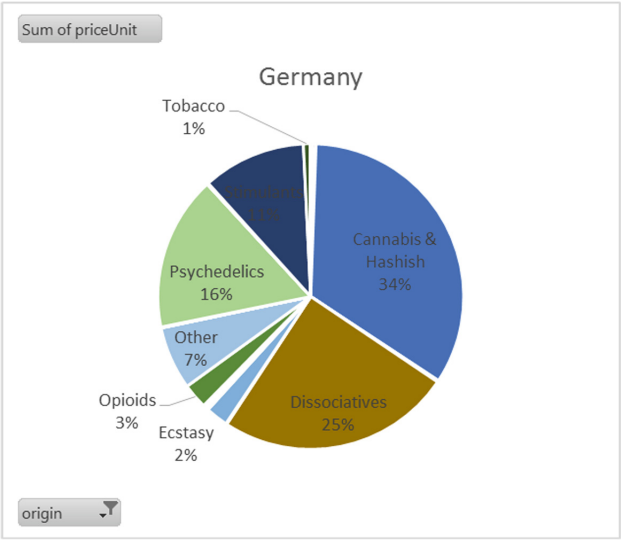

Fig. 8. Products from Germany, market distribution

3.6 Products from Afghanistan

It is interesting to notice that, unlike most of the countries, Afghanistan does not retail drugs on AlphaBay. The vast majority of the products advertised are false identity documents and fake accounts. Sellers from Afghanistan are also dealing in electronic devices and software (Fig. 9).

3.7 Other Markets

Here are some of the key findings and some of the peculiarities of other individual markets:

- **Netherland:** 100% of the products advertised by sellers in this country are Drugs and Chemicals. Two type of drugs, Ketamine and Crack are the top products advertised, with 36% and 38% respectively. This is similar for **Germany** (95%), **UK** (87%) and **Australia** (90%), **Canada** (97%), **France** (95%), **India** (nearly 100%).

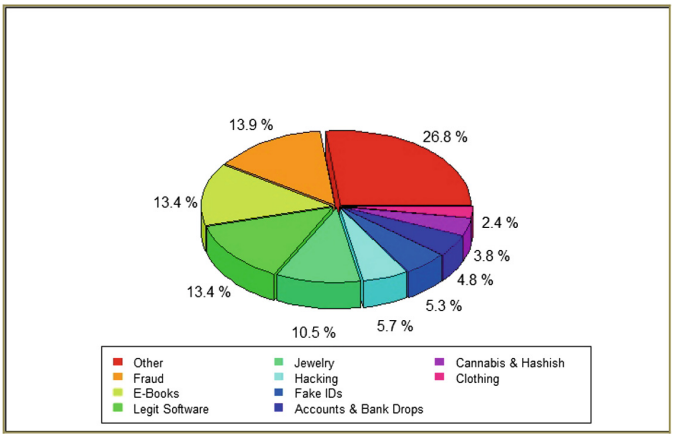

Fig. 9. Products from Afghanistan, market distribution

- **China:** 74% of the market is represented by drugs. This is a significant increase comparing with the data that was collected in Agora in 2015 [15]. It is worth noting that most of the drugs available are chemical. Counterfeit items are 18% of the market, and gold and jewels represent 16% of the market (much more than in all the other countries). That said, over 16% of the counterfeit items are again jewelry, which makes it a fairly unique market.
- **India:** India has been clearly identified as the main seller of Ketamine. Ketamine is very popular in several markets, as United Kingdom (15%), Czech Republic (68% of the market), France (58%), Germany (25%), India (96%), Netherlands (36%), Switzerland (36%). Worldwide, ketamine represents about 10% of the market. It is worth noting the presence of India – both a producer and consumer of Ketamine (while in most other countries is likely to be imported). It is also worth mentioning that Ketamine can and is also used as a date rape drug. In India's case it seems to be supply driven rather than just demand driven.
- **Japan:** Japan's market stands out; sellers mainly distribute digital products and online services. Drugs and Chemicals only contribute to 2% of the market. Examples of digital goods on sales include game keys, fraud softwares, malwares and security software.

4 Conclusion

The work presented in this paper is a follow up research from [15], reporting on the analysis of the largest dark web marketplace at the time.

The conclusion drawn from our research are alarming.

Dark web markets keep resurfacing regardless of the efforts made to shut them down, and Alphabay was much larger than any other marketplaces before.

Around $88M of individual products are estimated to have been on sale on AlphaBay at the time of our research. The total value for the transactions during the existence of the marketplace is in the region of $590M.

Drugs, fake IDs and weapons were readily available in a trans-national marketplace, just one click away and (fairly) anonymously. When it comes to counterfeit documents, any EU ID card would allow potential buyers to travel through any country in the EU, open bank accounts and in general create a new identity for himself/herself.

A new development is that dark web markets are now working very cautiously, implementing security measures and hacker avoidance updates regularly. The architectures are clearly more secure.

The understanding of illegal dark web markets is crucial. Information gathered on dark web markets can provide police forces and policymakers with greatly needed data to make informed decisions.

Although this research presented in this paper provides a detailed analysis of AlphaBay market, it can be extended to other dark web markets as well.

Acknowledgements. The authors wish to thank Simon Delecourt and Edouard Donze, who spent a summer internship at UEL in June–August 2017 and contributed to the preliminary data analysis for this paper.

References

1. Achsan, H.T.Y., Wibowo, W.C.: A fast distributed focused-web crawling. Proc. Eng. **69**, 492–499 (2014)
2. Anon: Megaupload file-sharing Site Shut Down (2012). http://www.bbc.co.uk/news/technology-16642369. Accessed 07 Aug 2017
3. Anon: 2017 SAS, R, or Python flash survey results (2017). http://www.burtchworks.com/2017/06/19/2017-sas-r-python-flash-survey-results/. Accessed 08 July 2017
4. Anon: Forget silk road, cops just scored their biggest victory against the dark web drug trade (2017). https://www.forbes.com/sites/thomasbrewster/2017/07/20/alphabay-hansa-dark-web-markets-taken-down-in-massive-drug-bust-operation/#480a3cb05b4b. Accessed 07 Aug 2017
5. Anon: Cocaine – Trafficking and supply (n.d.). http://www.emcdda.europa.eu/publications/eu-drug-markets/2016/online/cocaine/trafficking-and-supply. Accessed 24 July 2017
6. Anon: Explore: AlphaBay (n.d.). https://trends.google.co.uk/trends/explore?date=2014-12-01%202017-07-01&q=alphabay. Accessed 22 July 2017
7. Anon: Explore: AlphaBay, Dream Market (n.d.). https://trends.google.co.uk/trends/explore?date=2015-01-01%202017-07-01&q=alphabay,Dream%20Market. Accessed 22 July 2017
8. Anon: Facts about drugs (n.d.). http://www.thestudentpocketguide.com/2012/01/student-life/health-and-relationships/facts-about-drugs/. Accessed 12 July 2017
9. Anon: How much do drugs cost? (n.d.). http://www.drugwise.org.uk/how-much-do-drugs-cost/. Accessed 12 July 2017
10. Anon: R Markdown tutorial (n.d.). http://rmarkdown.rstudio.com/lesson-1.html. Accessed 29 June 2017
11. Anon: The average cost of illegal street drugs (n.d.). http://www.rehabcenter.net/the-average-cost-of-illegal-drugs-on-the-street/. Accessed 12 July 2017

12. Anon: The cost of street drugs in Britain (n.d.). http://www.telegraph.co.uk/news/uknews/crime/11346133/The-cost-of-street-drugs-in-Britain.html. Accessed 12 July 2017
13. Anon: What illegal drugs cost on the street around the world (n.d.). http://o.canada.com/business/interactive-what-illegal-drugs-cost-on-the-street-around-the-world. Accessed 12 July 2017
14. Baraniuk, C.: Alphabay and Hansa Dark Web Markets shut down (2017). http://www.bbc.co.uk/news/technology-40670010. Accessed 24 July 2017
15. Baravalle, A., Lopez, M. M., Lee, S.W.: Mining the dark web: drugs and fake IDs. IEEE, Barcelona (2017)
16. Egan, M.: What is the Dark Web? How to access the Dark Web. What's the difference between the Dark Web and the Deep Web? (n.d.). http://www.pcadvisor.co.uk/how-to/internet/what-is-dark-web-how-access-dark-web-deep-joc-beautfiulpeople-3593569/
17. INTELLIAGG: Deeplight. (Shining A Light On The Dark Web), s.l., s.n (2016)
18. Kopan, T.: DOJ announces takedown of dark web market Alphabay (2017). http://edition.cnn.com/2017/07/20/politics/doj-takes-down-dark-web-marketplace-alphabay/index.html. Accessed 24 July 2017
19. Lewis, S.J.: OnionScan Report: July 2016 - HTTPS Somewhere, Sometimes, s.l., Mascherari Press (2016)
20. Zulkarnine, A., Frank, R., Monk, B., Mitchell, J.: Surfacing collaborated networks in dark web to find illicit and criminal. s.l., s.n (2016)