Trends in the availability and type of drugs sold on the internet via cryptomarkets: Findings from the Drugs and New Technologies (DNeT) project, 2018-2021

Pedersen, M.[1], Man, N.[1], Sutherland, R.[1], Bruno, R.[1,2], Barratt, M. J.[1,3] & Peacock, A.[1,2]

[1]National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia
[2]School of Psychology, University of Tasmania, Hobart, Australia
[3]Global, Urban and Social Studies, Royal Melbourne Institute of Technology University, Melbourne, Australia

Background

Online drug markets operating on the darknet, known as cryptomarkets, facilitate the trade of illicit substances at an international level. The Drugs and New Technologies (DNeT) project at NDARC forms part of Drug Trends, a national illicit drug surveillance program that has been running for two decades. DNeT has been monitoring cryptomarkets since 2014 to assess their growth, the range of products sold, and impact of major disruptions such as law enforcement operations and 'exit scams'.

Aim

- To describe trends in drug listings on cryptomarkets over time, focusing on the period September 30 2018 to September 30 2021 inclusive.
- To explore strictly unique listings (listings at their first point in time), unique vendors, and the prevalence of vendors who sell across multiple markets in the specified time period.

Methods

There are 5 major steps to the process of collecting relevant information from cryptomarkets for analysis.

1. Cryptomarket identification:
Through monitoring of darknet forums and information sites such as Dread, Darknetlive, DarkNetStats and dark.fail, new cryptomarkets are identified. Cryptomarkets that are in the English language with ≥100 listings and ≥1 vendor qualify for monitoring.

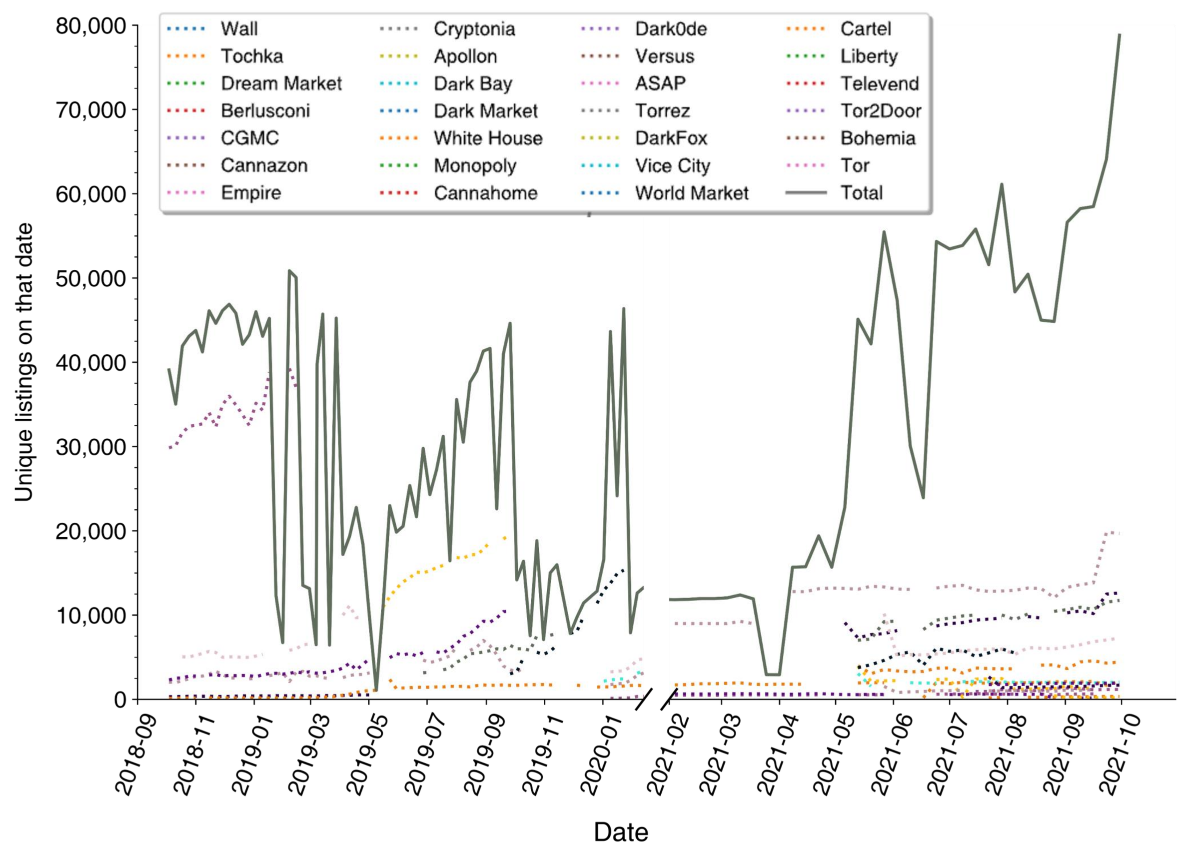
2. Cryptomarket scraping:
Through the use of Selenium re-packaged for the Tor browser (tb-selenium) in Python, web scraping scripts are developed for each qualified cryptomarket. These scripts save the raw HTML files for drug listings across markets every Thursday.

3. Data extraction:
Data from the raw HTML files is accessed using Python extraction scripts. These scripts utilise XPath selectors from the Scrapy library to extract information of interest, such as listing prices, country of origin, and vendor details.

4. Data cleansing:
Cryptomarkets vary in their ability to show differing quantities of the same product. To account for multiple listings of the same product varying in quantity in a given week, the listing with the lowest quantity is kept, and listings of additional quantities are removed. This removal reduces the number of listings by 30% on average. In addition, vendor names are de-duplicated where possible, through use of Levenshtein distance, listing style and geographical comparison to identify slight variations of vendor names that are ultimately the same vendor.

5. Drug classification:
A LSTM (long short-term memory) neural network was implemented in TensorFlow, that enables mapping of drug listings to 1462 Level 1 drug categories, under 46 broader Level 2 drug categories. Listings that are categorised to a Level 1 substance with a prediction accuracy >= 90% qualify for further analysis.

Figure 1: Cryptomarket listings across period

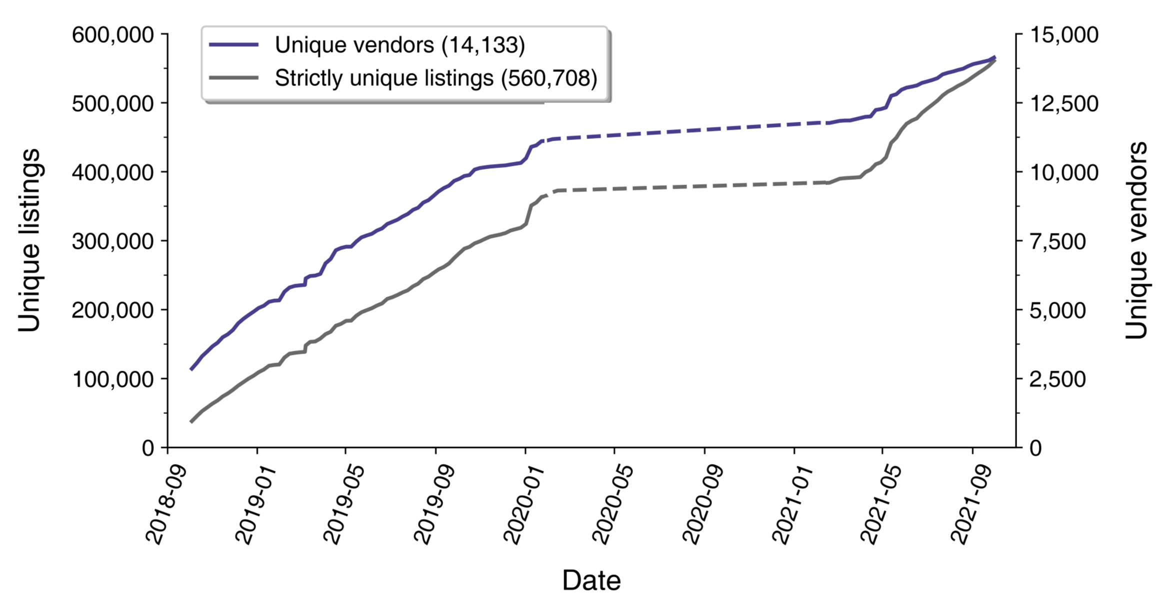


Note: Dotted lines indicate data for single market. Diagonal lines indicate gap.

Figure 2: Top 10 drug categories across period

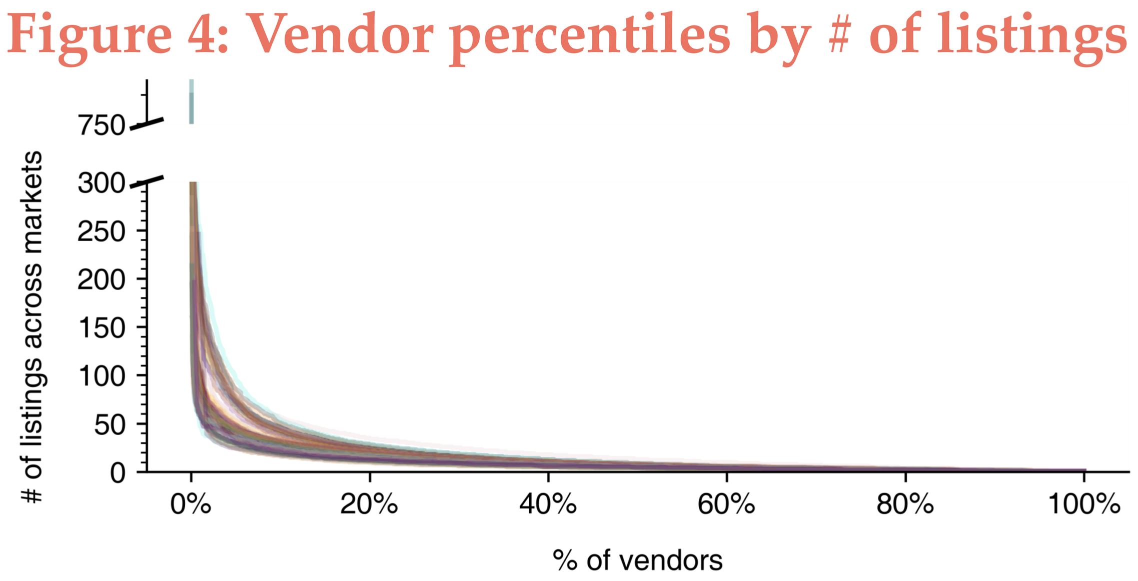
Level 2 Category	# of Listings	% of Total
Cannabis	1,103,965	34.34%
MDMA	306,498	9.53%
Cocaine	235,262	7.32%
Benzodiazepines	222,152	6.91%
PiEDS/weight loss	190,822	5.94%
Meth/amphetamine (illicit)	183,861	5.72%
Opioids	176,227	5.48%
NPS	126,432	3.93%
LSD	121,984	3.79%
Ketamine	107,908	3.36%

Figure 3: Cumulative unique vendors & strictly unique listings across period



Note: Dashed lines indicate missing data due to gap in completeness.

Figure 4: Vendor percentiles by # of listings



Note: Each line represents the data from a single weekly scrape in the time period.

Findings & Implications

The dataset comprised of 106 weekly scrapes, from 30 September 2018 - 13 February 2020, and 4 February 2021 - 30 September 2021. In this period, there were 3,214,985 listings across 27 unique markets (Figure 1). Data from late February 2020 to January 2021 was ignored due to issues with data quality.

Over a third of all listings in the period were cannabis. Together with MDMA and cocaine, these three Level 2 categories made up 51.19% of all listings (Figure 2).

560,708 'strictly unique' listings were identified, where only new listings on a given scrape date were counted (Figure 3). This indicates the longevity of listings, with a listing appearing for 5.72 weeks on average, and 25% appearing for 8 or more weeks. A tendency for vendors to list the same product across markets was also evident, with 30% of vendors having listings on two or more markets. Accounting for this by removing market information reduced strictly unique listings to 462,328.

14,133 unique vendors were identified across the timeframe, from a total of 17,591 distinct vendor names. Approximately 3,450 vendors had 2 or more accounts with slight name variations, that were detected using the vendor de-duplication methods outlined. On average, there were 2,054 unique vendors in a weekly scrape.

A considerable amount of listings are from a small group of major vendors, but beyond that, the market environment rapidly becomes fragmented. Whilst the median vendor had just 7 listings weekly, the top 1% of vendors had 175 or more in an average week. Across the 106 weeks, the fragmentation remained largely consistent (Figure 4). Averaged across the period, the top 5% of vendors made up 31% of listings, and the top 1% had near-identical listings to the bottom 50% (11.97% and 12.12% of listings respectively).

Limitations

There are a number of limitations with the dataset:

- Where possible, matching of highly similar vendor names has been applied. However, vendor duplicates may still remain for some markets due to varying accuracy. Hence, the true number of unique vendors is likely lower than 14,133.
- Identification of cryptomarkets was not exhaustive in the Feb-May 2021 period. 8 markets were identified and scraped from May 2021, but these markets existed for some months prior.
- Some cryptomarkets make use of random sorting algorithms to maximise 'fairness' of product discovery by users. Issues with implementation meant some historical scrapes did not save all pages, and hence underreported listings.

Further Information

Specific information on cryptomarket drug listings can explored using the Interactive Visualisation Platform, whilst further details regarding the underlying methods can be found in the links attached.

Acknowledgements: We would like to acknowledge the following individuals for their contribution to this project: Dr Amanda Roxburgh, Joe Van Buskirk, A/Prof Lucinda Burns, Dr Courtney Breen, Sundresan Naicker, A/Prof Timothy Dobbins, Anant Mathur and Rajat Katyal. We would like to acknowledge the Australian Institute of Health and Welfare. Drug Trends is funded by the Australian Government Department of Health under the Drug and Alcohol Program.

Contact: Max Pedersen | ✉ max.pedersen@unsw.edu.au | ✉ drugtrends@unsw.edu.au | ☎ +61 (2) 9385 0333