University of Public Service

Doctoral School of Law Enforcement

Ákos Szigeti

Innovative social research methods in criminology:

Text mining in darknet exploration

Theses of PhD dissertation

Supervisor:	**Co-supervisor:**
Dr Katalin Parti	Dr Gábor Héra
....	....

Budapest, 2024

TABLE OF CONTENTS

1 METHODOLOGICAL INNOVATION RESEARCH

„Researchers are in the process of making a change akin to the transition from photography to cinematography" (Salganik, 2019:5)

As society is increasingly digitized, the number of interconnected users grows year by year (International Telecommunication Union, 2019). Citizens use social media networks and the personal and business services of these with increasing regularity (Digital, 2020). The role of sociologists in this situation is twofold. On the one hand, they must follow and analyze the latest processes in society, and on the other, they must explore the future trends of societal problems. Social science can only make use of the opportunities of digitization "if they are capable of renewing their research methodology while maintaining their critical reflection" (Németh & Barna, 2019:121).

This self-reflexion is a prerequisite of the adaptation of digital methods to rethink sampling procedures and the interpretation of data (boyd & Crawford, 2012; Karpf, 2012) to strike a bridge between the strict research ethics rules of social sciences and the ad hoc approach of data science, and develop these further (Salganik, 2019). It also represents a challenge to acquire the technical expertise required for applying data science methods, for which reason social scientists typically conduct research involving large data sets in interdisciplinary cooperation (Metzler és mtsai., 2016).

The first sub-project of this PhD research was aimed to explore researchers' attitudes towards the use of innovative research methodologies at the intersection of social science and data science. We aimed to contribute to the development of social science research by providing recommendations for researchers on the use and adaptation of innovative research tools, methodological pitfalls in the use of transactional data, and research ethics issues.

Our exploratory research was based on expert interviews, which were aimed at an in-depth exploration of the willingness of social and data sciences teaming up in interdisciplinary teams. We also conducted an online survey based on purposive sampling, where we tried to assess which digital data-oriented research methods and tools (e.g., data analysis software) are being used by social scientists. However, the survey results were not included in the research study due to the small sample size.

The findings which therefore derived from the one-on-one interviews (n = 22) reinforced the importance of interdisciplinary collaboration and highlighted the significance of

"interpreters," i.e., individuals able to communicate with and connect various areas of science, education, and academic institutions' role in enhancing interdisciplinary collaborations of sciences. Additional concerns of participants emerged in terms of research methodology applied in the digital world (i.e., data validity, credibility and research ethics). Finally, participants identified open science and the transparency of research as the key to the future development of social sciences.

2 TEXT MINING-BASED DARKNET RESEARCH

Amid the global opioid crisis, the volume of drug trade via online anonymous marketplaces designed to facilitate these transactions, called *darknet markets* or *cryptomarkets* has risen to an all-time high. Although darknet markets accounted for only a tiny slice of the global illicit drug market even in their most prosperous years, figures show that their share has steadily increased in the second half of the 2010s (The United Nations Office on Drugs and Crime, 2020).

Data show that, in the short term, the restrictions caused by the COVID-19 pandemic have even boosted this increase in the number of darknet drug purchases (European Monitoring Centre for Drugs and Drug Addiction & European Police Office, 2020; Hawdon és mtsai., 2022). Although drug dealers who operate through face-to-face meetings bought less from darknet markets under restrictions, the number of individual consumer darknet market orders increased significantly during this period (European Monitoring Centre for Drugs and Drug Addiction & European Police Office, 2020). However, recent reports suggest that in the long term, the pandemic may have contributed to a decrease in the volume of the darknet drug trade (European Monitoring Centre for Drugs and Drug Addiction., 2022) since delivery difficulties due to the lockdowns have destroyed the reliable link between vendors and customers (Bergeron és mtsai., 2022).

Directly exploring the darknet drug trade is challenging due to the difficulty of reaching its anonymous participants, making it difficult to apply traditional methods such as survey questionnaires (Karden & Strizek, 2022). However, scraping textual data from darknet markets provides an opportunity for the direct observation of vendor reputation data, which Jardine (2019) suggested should be used as an element of darknet threat metrics.

The data analysed in the second sub-project of this PhD research was scraped from the Dark0de Reborn darknet market between June 10 and June 27, 2021. The Dark0de Reborn darknet market, whose predecessor was a hacker forum that operated until 2015, opened in May 2020 and closed in February 2022. While it existed, this darknet market was a dominant player among illegal online drug markets based on daily minimum sales (The United Nations Office on Drugs and Crime, 2023). Thus, although Dark0de Reborn darknet market was only a slice of the darknet drug markets, the data scraped from it provided an opportunity to directly

examine community factors behind the operation of darknet drug markets on one of its flagship platforms.

2.1 Qualitative content analysis

Based on the results of the research project presented in the first chapter, before implementing the text analytics procedure on the data scraped from the darknet, we also performed qualitative analysis to increase the reliability of the results. The qualitative content analysis stage had two objectives. On the one hand, it aimed to explore the characteristics of vendors' services and reputation by analysing the descriptions of the products sold on the darknet market and the customer reviews of the bought products. On the other hand, it also aimed to gather preliminary knowledge and to contribute to the construction of a *domain knowledge* on the research field, which was essential to implement the next research stage applying automatic text analytics.

The qualitative content analysis was conducted on a sample (n = 100) randomly selected from 6,357 product descriptions and a sample (n = 500) randomly selected from 34,619 customer reviews. Both samples are from products found in the drug category of the darknet market Dark0de Reborn. On the supply side, vendors tended to provide basic information on the drugs, a mention of their high quality, the speed and stealth of delivery, their availability for responding to messages, the effects of the drugs, and sometimes even instructions for use. Regarding the demand side, customers usually praised the quality of the product, mentioned the speed and stealth-secure packaging of delivery as essentials, and expressed only a small number of issues.

These results support the applicability of Norbert Elias' social figuration theory in which the interdependencies of the actors are fuelled by trust. This theoretical frame sheds light on the social value of the community of darknet markets. Furthermore, the findings formulate a robust hypothesis for future research about the previously undervalued role of delivery providers.

2.2 Topic modelling

In the second phase of the text mining-based darknet research, we used topic modelling to examine the customer reviews scraped from the Dark0de Reborn darknet market. This research

stage was aimed at exploring the risk mitigation efforts of the platform's participants and thus measuring the threats posed by darknet markets. Results from the previous qualitative research phase highlighted the social value of the communities formed by the actors around darknet markets and argued that previous research underestimated the role of delivery service providers. In contrast to the qualitative approach, which allowed for the analysis of a relatively small subset of customer reviews, this study aimed to analyse the entire sample using topic modelling. In addition to increasing the study sample, the machine learning-based methodology of this research phase also provided a new perspective on previously identified community factors influencing the darknet drug trade.

We performed Latent Dirichlet Allocation topic modelling on customer reviews of drug products (n=25,107) scraped from the darknet market Dark0de Reborn in 2021. We obtained a model resulting in 4 topics (coherence score=0.57): (1) feedback on satisfaction with the transaction; (2) report on order not received; (3) information on the quality of the product; and (4) feedback on vendor reliability.

These topics identified in the customer reviews suggest that the community of the selected darknet market implemented a safer form of drug supply, reducing risks at the payment and delivery stages and the potential harms of drug consumption. However, the pitfalls of this form of community-initiated safer supply support the need for universally available and professional harm reduction and drug checking services. These findings and our methodological remarks on applying text mining can enhance future research to further examine risk and harm reduction efforts across darknet markets.

3   SURVEY-BASED DARKNET RESEARCH

The reputation dynamics in the vendor-customer relationships on darknet markets have been very similar to surface web markets (Espinosa, 2019; Janetos & Tilly, 2017; Przepiorka és mtsai., 2017). However, while some factors of horizontal trust are equally important in legal and illegal trade (for example, reliability of vendors, payment options, or quality of goods) (Spagnoletti és mtsai., 2022), others are not as prominent in darknet purchasing (e.g. legal and data protection guidelines, guarantees of buyers' rights, public relations, and registration options) and those which are more beneficial in illegal trade (e.g. anonymity, secret operation) than in legal online shopping.

Although the factors potentially affecting customers' risk perception can be identified based on previous research, cyber criminology has not produced empirical research ranking the importance of the specific factors. This study was designed to fill this gap by developing a tool that measures the importance of the various trust factors. Based on previous research findings on legal and illegal online purchase transactions, we developed a survey module consisting of 21 trust factors, for which we formulated three main hypotheses:

> *Hypothesis 1 (H1): All trust factors are considered necessary in legal and illegal online purchasing situations.*

> *Hypothesis 2 (H2): The factors related to anonymity and encryption are considered more important in illegal online purchasing situations compared to legal online purchasing.*

> *Hypothesis 3 (H3): If it is not possible to purchase the drug in a legitimate commercial way, but it is possible to purchase it illegally both offline and online, users would prefer to order the substance on an online platform.*

To test out the measurement tool, a large-scale survey with projective situational questions was conducted among university students in Hungary. The sample (n = 5481) was compiled to include potential darknet market customers, respondents with above-average computer skills needed to access the darknet, and taking into account that university students are a group of society particularly exposed to drug consumption.

The end product of this research is a trust matrix ranking the factors affecting illicit drug purchases on darknet markets. Among the factors, the survey's target group ranked reliable and

undamaged delivery of goods and the reliability of vendors as the most important. The measurement tool developed in this research will facilitate further criminological research on vendor reputation. Its findings also point to the need for further research on delivery providers and predict that influencing the delivery-related risk perception of potential customers could effectively reduce demand.

4   NEW SCIENTIFIC RESULTS

This doctoral research has gathered new scientific knowledge in two areas: the methodology of scientific investigation of the digital society and, based on the results of that preliminary research, the operation and policing of darknet markets.

4.1   New scientific findings on social research methods

4.1.1   The potential of multidisciplinary collaboration

In today's digital society, the application of data science methods brings new opportunities that can be exploited by conducting research in multidisciplinary collaborations. Thus, when studying crime and the social and institutional responses to it (i.e. research in the field of criminology and law enforcement studies), it is suggested to create multidisciplinary research teams that combine theoretical knowledge and traditional social science research methodology with data science expertise. Multidisciplinary collaborations not only enable innovative research methods to be applied but can also provide the technological expertise and infrastructure to carry out applied research, resulting end products (e.g. technological tools) that can be directly applied in practice. A significant proportion of today's research and innovation calls for the creation of such end products, based on scientific results but with practical use cases.

4.1.2   Advantages of mixed-methods research concepts

The reliability and validity of data science methods can be enhanced by designing multi-method or mixed-methods research concepts. For example, knowledge gathered in the given research field by qualitative methods can facilitate the application of a data science method. In the case of this doctoral thesis, the domain knowledge constructed in the qualitative stage laid the foundation for the text analytical study. Although the construction of the domain knowledge (i.e. the exploration of the platform under study), is presumably also happens in the case of a stand-alone application of a data science method, a systematic, documented and transparently presented (qualitative) preliminary study can increase the reliability of the research.

4.1.3 The importance of interoperability of research methods and transparency of research processes

Finally, the replication of research using data science methods and their innovative adaptation in other research areas can be facilitated by ensuring the interoperability of the methodology and the transparency of the research process. Thus, in addition to a detailed description of the research process, it is necessary to focus on the limitations of a given study, taking into account the various data validity issues derived from conducting research in an online environment. Transparency is also required because of the relevant research ethics considerations, since researchers are usually cannot provide participants with informed consent forms when collecting transactional data. Thus, describing the research process and the decisions taken about the use of the data can reduce the risk of harm caused by the research.

4.2 New scientific findings on the operation and policing of darknet markets

4.2.1 The community nature of darknet markets

The results confirmed that there are interdependencies between the different actors participating in the process of the darknet drug trade, which interdependencies are based on various trust factors. Placing the phenomenon in a social theoretical framework has revealed additional new scientific findings not or only tangentially discussed in previous research (see the following sections).

4.2.2 The key role of delivery in darknet drug trade

Previous research has not included delivery providers on the list of actors involved in the social figuration of drug trafficking via the darknet. In contrast, the multi-method text mining-based research and the survey research implemented among potential buyers have shown that delivery providers play an important role in the reliable operation of darknet markets. Delivery emerged as the riskiest element of the process and draws the attention of practitioners and legislators alike to focus prevention and law enforcement actions on this phase. However, the discussion of the results also pointed out that increasing the risks for real darknet market users would not necessarily reduce harm, as darknet markets may be able to mitigate social harm by reducing risks.

4.2.3 Efforts by darknet market communities to ensure a safer supply

The actors involved in the darknet drug trade are making efforts to reduce the risks, not only the risks of arrest and financial victimisation, but also the potential harms of the consumption. Identifying the relevant risks can help to improve prevention by providing ammunition for risk awareness campaigns. Finally, the nature of how the community operates can also guide the design of darknet drug policy and law enforcement interventions to reduce the social and health harms caused by drug consumption.

5 LIST OF SINGLE AND CO-AUTHORED PUBLICATIONS ON THE TOPIC

Kiss, T., Dornfeld, L., & Szigeti, Á. (2021). *Vásárlói bizalom az illegális szerek online piacán* (Kiss, T. Szerk.). Ludovika Egyetemi Kiadó.

Kiss, T., & Szigeti, Á. (2023). Ranking Trust Factors Affecting Risk Perception in Illicit Drug Purchase on the Darknet: A Large-Scale Survey Study in Hungary. *European Journal on Criminal Policy and Research*. https://doi.org/10.1007/s10610-023-09545-x

Parti, K., & Szigeti, Á. (2021a). Innováció a szociológiában: A társadalomtudomány és az adattudomány metszetében elhelyezkedő, innovatív kutatási módszerekre irányuló kutatói attitűdök vizsgálata. *Socio.hu Társadalomtudományi Szemle*, *11*(1), 147–171. https://doi.org/10.18030/socio.hu.2021.1.147

Parti, K., & Szigeti, A. (2021b). The Future of Interdisciplinary Research in the Digital Era: Obstacles and Perspectives of Collaboration in Social and Data Sciences—An Empirical Study. *Cogent Social Sciences*, *7*(1), 1970880. https://doi.org/10.1080/23311886.2021.1970880

Szigeti, Á. (2021). Matthew J. Salganik: Bitről bitre. Társadalomkutatás a digitális korban. *Socio.hu Társadalomtudományi Szemle*, *11*(1), 172–176. https://doi.org/10.18030/socio.hu.2021.1.172

Szigeti, Á. (2022a). Szövegbányászat a dark neten: Rendészettudományi alkalmazások. *Belügyi Szemle*, *70*(4), 757–767. https://doi.org/10.38146/BSZ.2022.4.7

Szigeti, Á. (2022b). Távolról is olvasható? A szövegbányászat perspektívája a szociológiai kutatásban. *Szociológiai szemle*, *32*(2), 91–100. https://doi.org/10.51624/SzocSzemle.2022.2.5

Szigeti, Á., Frank, R., & Kiss, T. (2023). Trust Factors in the Social Figuration of Online Drug Trafficking: A Qualitative Content Analysis on a Darknet Market. *Journal of Contemporary Criminal Justice*, 10439862231159996. https://doi.org/10.1177/10439862231159996

Szigeti, Á. (2023). Út egy érték- és evidenciaalapú, demokratikusabb drogpolitika felé: Recenzió Alison Ritter Drug Policy című művéről. *Magyar Rendészet*, *23*(2), 221–226. https://doi.org/10.32577/mr.2023.2.13

Szigeti, Á., Frank, R., & Kiss. T. (Bírálat alatt). Contribution to the harm assessment of darknet markets: topic modelling drug reviews on Dark0de Reborn.

6 REFERENCES

Bergeron, A., Décary-Hétu, D., & Ouellet, M. (2022). Conflict and Victimization in Online Drug Markets. *Victims & Offenders*, *17*(3), 350–371. https://doi.org/10.1080/15564886.2021.1943090

boyd, D., & Crawford, K. (2012). Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Digital (2020). *3.8 billion people use social media. Digital 2020 Reports.* https://wearesocial.com/digital-2020

Espinosa, R. (2019). Scamming and the reputation of drug dealers on Darknet Markets. *International Journal of Industrial Organization*, *67*, 102523. https://doi.org/10.1016/j.ijindorg.2019.102523

European Monitoring Centre for Drugs and Drug Addiction. (2022). *European drug report 2022: Trends and developments.* Publications Office. https://data.europa.eu/doi/10.2810/75644

European Monitoring Centre for Drugs and Drug Addiction. & European Police Office. (2020). *EU drug markets: Impact of COVID 19.* Publications Office. https://data.europa.eu/doi/10.2810/19284

Hawdon, J., Parti, K., & Dearden, T. (2022). Changes in Online Illegal Drug Buying during COVID-19: Assessing Effects due to a Changing Market or Changes in Strain using a Longitudinal Sample Design. *American Journal of Criminal Justice : AJCJ*, *47*(4), 712–734. https://doi.org/10.1007/s12103-022-09698-1

International Telecommunication Union. (2019). *Measuring digital development. Facts and figures 2019.* https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2019.pdf

Janetos, N., & Tilly, J. (2017). *Reputation Dynamics in a Market for Illicit Drugs.* arXiv. https://doi.org/10.48550/ARXIV.1703.01937

Jardine, E. (2019). The trouble with (supply-side) counts: The potential and limitations of counting sites, vendors or products as a metric for threat trends on the Dark Web. *Intelligence and National Security*, *34*(1), 95–111. https://doi.org/10.1080/02684527.2018.1528752

Karden, A., & Strizek, J. (2022). The potential for using web surveys to investigate drug sales through cryptomarkets on the darknet. In *Monitoring drug use in the digital age: Studies in web surveys*. Publications Office of the European Union. https://www.emcdda. europa.eu/publications/insights/web-surveys/potential-web-surveys-investigate-drug- sales-cryptomarkets-darknet_en

Karpf, D. (2012). Social Science Research Methods in Internet Time. *Information, Communication & Society*, *15*(5), 639–661. https://doi.org/10.1080/1369118X.2012.665468

Metzler, K., Kim, D. A., Allum, N., & Denman, A. (2016). *Who Is Doing Computational Social Science? Trends in Big Data Research A SAGE White Paper*. SAGE Publishing. https://doi.org/10.4135/wp160926

Németh, R., & Barna, I. (2019). Digitális társadalomkutatások az ELTE-n Beszámoló A társadalom kutatásának új útjai című workshop digitális társadalomkutatással foglalkozó szekcióiról. *Szociológiai Szemle*, *29*(4), 115–121. https://doi.org/10.51624/SzocSzemle.2019.4.5

Przepiorka, W., Norbutas, L., & Corten, R. (2017). Order without Law: Reputation Promotes Cooperation in a Cryptomarket for Illegal Drugs. *European Sociological Review*, *33*(6), 752–764. https://doi.org/10.1093/esr/jcx072

Salganik, M. J. (2019). *Bit by bit: Social research in the digital age* (First paperback printing). Princeton University Press.

Spagnoletti, P., Ceci, F., & Bygstad, B. (2022). Online Black-Markets: An Investigation of a Digital Infrastructure in the Dark. *Information Systems Frontiers*, *24*(6), 1811–1826. https://doi.org/10.1007/s10796-021-10187-9

The United Nations Office on Drugs and Crime. (2020). *In Focus: Trafficking over the Darknet—World Drug Report 2020.* https://www.unodc.org/documents/Focus/WDR20_Booklet_4_Darknet_web.pdf

The United Nations Office on Drugs and Crime. (2023). *Use of the Dark Web and Social Media for Drug Supply – World Drug Report 2023.* https://www.unodc.org/res/WDR-2023/WDR23_B3_CH7_darkweb.pdf