

# How has the COVID-19 Pandemic Affected Substance Abuse Dynamics in USA?

## A Capstone Research Report

Flory Huang  
Data Science Institute  
Univ. of California San Diego  
La Jolla, USA  
yuh354@ucsd.edu

Hanbyul Ryu  
Data Science Institute  
Univ. of California San Diego  
La Jolla, USA  
h9ryu@ucsd.edu

Gunther Schwartz  
Data Science Institute  
Univ. of California San Diego  
La Jolla, USA  
gswartz@ucsd.edu

**Abstract**—Drug usage is continuously increasing throughout the world, causing problems such as rising mortality rates. Especially in the time of the pandemic, behavioral changes are inevitable. We proposed that by constructing a knowledge graph from multiple data sources, we can know how COVID-19 has changed the situation for substance use. We observe news announcements, along with reports by the Drug Enforcement Agency (DEA) and social media posts from potential drug users. Utilizing methods such as Natural Language Processing (NLP) and behavior analysis, we aim to ultimately form a pipeline that accepts input from the user to output partial knowledge graphs that pertain to the information given.

**Index Terms**—COVID-19, substance abuse, knowledge graph, Natural Language Processing, Information Integration

### I. INTRODUCTION

The World Drug Report 2020 [1] observes that drug use around the world has been on the rise, in terms of both overall numbers and the proportion of the world’s population that uses drugs. In 2009, the estimated 210 million users represented 4.8 per cent of global population aged 15–64, compared with the estimated 269 million users in 2018, or 5.3 per cent of the population. In North America, the use of synthetic opioids such as fentanyl has fuelled two decades of increases in opioid overdose deaths. In 2018, fentanyls were implicated in two thirds of the 67,367 overdose deaths registered in the United States. At the same time, the drug business is also shifting. For example, the manufacture of methamphetamine was traditionally carried out in small-scale laboratories in the United States to serve the domestic market. But this kind of production seems now to be dwarfed by industrial-size laboratories in Mexico. The methamphetamine seized in the United States over the past few years is increasingly imported, with the trade controlled by Mexican cartels. As recently as Nov. 24, 2020, the Drug Enforcement Agency has seized \$3.5M in U.S. Currency and massive quantities of cocaine and fentanyl in Otay Mesa of the San Diego County<sup>1</sup> across from the Mexico border. These facts demonstrate that the

illicit substances not only pose a health hazard of epidemic proportions, it is also a thriving industry that Federal, State and Local law-enforcement agencies are trying to deter, but new businesses, shifting target population, new manufacturing units and ever-evolving distribution channels are constantly changing the dynamics of the drug ecosystem.

Unfortunately, there is no public-domain information system today that public health researchers can use to monitor, analyze and model illicit drug related activities and their impact. The lack of such a continuous monitoring and analysis is particularly conspicuous during the current COVID-19 period, which has significantly disrupted the normal dynamics of the drug ecosystem. While some trading routes have disappeared, newer suppliers have sprung up to meet the increased demand commensurate with the restrictions on travel and closed-area congregations, the emergent work-from-home culture and increased anxiety levels [2], [3].

The broad goal of the proposed project is to combine heterogeneous information integration, natural language processing and machine learning techniques to construct a *time-evolving knowledge graph* that will allow us to satisfy the analytical needs of public health researchers interested in the impact of the pandemic on the illicit substance marketplace. The knowledge-graph we aim to construct will take the form of an extended property graph model [4], and will be guided by domain-specific research questions. Thanks to Prof. Annick Borquez, Professor of Public Health, and the co-advisor of the proposed project, we have identified three research questions that we propose to answer based on the knowledge graph.

- 1) How are the nature and volume of drug-related conversation in social media and events in the news media changing through the pre/during/post pandemic period?
- 2) Has there been a shift in the manner in which drugs/drug-categories/drug-use-patterns are discussed in these media through the pre/during/post pandemic period?
- 3) How have the discussions around mental health issues including depression, anxiety, suicidal thoughts, and loneliness changed through the pre/during/post pandemic period?

<sup>1</sup><https://www.dea.gov/press-releases/2020/11/24/agents-seize-35-million-us-currency-and-massive-quantities-cocaine>

Since we model the knowledge graph as an instance of the property graph data model [5], the entities (nodes), relationships (edges), node attributes and edge attributes of the knowledge graph will be governed by these three research questions.

#### A. Related work

Four of the works that informed our approach heavily and should be mentioned are (as we call them), the Remine paper [6], the AutoPhrase paper [7], the PREDOSE paper [8], the MentalHealth paper [9]

## II. STUDY DESIGN

#### A. Data Sources

Based on the research questions, we have identified three kinds of data sources that will inform our knowledge graph.

- 1) **Terminology Sources.** These sources include *ontologies* (e.g., RxNorm) that contains the scientific, generic and street names of drugs, classifications of drug families, "prescription terms" of drugs and drug strengths, as well as *dictionaries* for names of locations and various Government Agencies related to drug administration, control, law enforcement and epidemics.
- 2) **Social Media.** Social media data captures public perceptions but tends to be noisy due free-flowing conversation that may be irrelevant. We have selected Reddit ([www.reddit.com](http://www.reddit.com)) and its different topical forums for general conversation around drugs. The queries and subreddits we used to collect reddit data: *Queries:* "xanax", "heroin", "weed", "meth", "DEA", "FBI", "THC", "COVID", "Coronavirus", "Corona", "opiates", "hydros", "hydrocodone", "opiate", "opioid", "best", "fentanyl", "heroin", "adderal", "amps", "speed", "amphetamine", "stims", "stimulant", "hydro", "oxy", "hydrocodone", "oxycodone", "subs", "suboxone", "xannies" *Subreddits:* drugs, researchchemicals, opiates, stims, cocaine,

#### B. Assumptions

To address the research challenges within the scope of our Capstone Project, we will make a number of simplifying assumptions during knowledge modeling and information extraction from our data sources.

- (a) Drug-related information often have multiple levels of granularity. For example, a drug may be specified by name, name+dose, name+form, and so forth. We standardized these entities to the drug type.
- (b) We treated different drug combinations occurring in sentences as their own entities without considering any complex semantic connections within the sentence. However, we included the names of the individual drugs as properties of these compound entity nodes.
- (c) For each text source, we considered only sentences containing two or more significant entity candidates and assume these will constitute most of our informative sentences.
- (d) We used results from existing literature to identify mental health related comments from social media data. We assumed that news articles do not have mental health related information.

#### C. Methods

We used the ontology list as our dictionary to extract the drug terms as our entity. We performed similar matching on our social media data and found all sentences that contain at least one drug term. The Drug term is one type of our entity in the knowledge graph. Another type of entity is the noun chunks that appear in the same post as the drug term.

To generate relationships between noun phrases and drug terms, especially, long-range associations between terms. We used a state-of-the-art relationship extraction method Blank-Matching [10]. Unfortunately, the eight default relationships did not work well on our social media data. So we tried to define our own relationships *Drug-Consumption*, *Drug-Comment*, *Drug-Consequence* that fit in with the social media drug discussion content. We generated some training data for these three relations and use them to tune the Blank-Matching [10] package. However, because of the messiness of social media content and limited time in this project, our training data size is limited which did not lead to an effective result. So we did not include the output relationship in our result.

To answer our third research question regarding mental health issues, we used *behavior analysis* on the textual content to identify the emotional state of the sentence. We used a similar approach as prior study [9]. An emotion analysis package *Limbic* [11] and *manually defined lexicons* (e.g. suicidal tendencies, isolation, domestic stress) were used to classify emotion from sentences. The emotion scores are also marked as an attribute of the drug entity.

## III. RESULT

- (a) The volume of drug-related conversations in Reddit increased significantly after the COVID-19 pandemic. From PRE-COVID (2019/03-2020/03) to MID-COVID (2020/03-2021/03): The total posts/comments number in drug-related SubReddit increased from 5047 to 16834. [Fig 1]
- (b) The most discussed drugs have also changed during the COVID-19 pandemic. *Smoke* is the most popular term before COVID (14.9 percent of posts or comments contained it). During COVID-19, the occurrence of *Smoke* decreased to 8.1 percentage ( $z=-14.512, p=1.014e-47$ ). Mention of *Heroin* increased from 7 percentage to 10.7 percentage ( $z= 7.900, p=2.798e-15$ ). Mention of *Alcohol* increased from 2 percentage to 3 percentage ( $z= 3.749, p=1.779e-4$ ) [Fig 2 3]
- (c) Emotion expressed in online drug discussion also changed in COVID-19 pandemic. Within eight basic emotion category, *Anger*, *Disgust*, *Anticipation*, *Trust*, *Joy*, *Surprise*, *Fear*, *Sadness*, *Trust* and *Disgust* have changed significant. The score for *Trust* have increased 0.012 (mean=0.252,

$t=2.634$ ,  $p=0.008$ ), and score for disgust have decreased 0.009 (mean=0.141,  $t=-2.349$ ,  $p=0.018$ )

Fig. 1.

changes in drug-related subreddit posts&comments volumn from 2019-03 to 2021-03

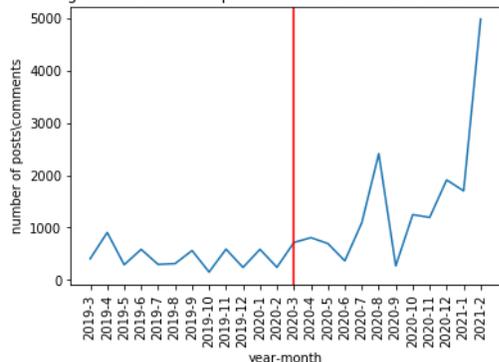


Fig. 2. Top mentioned drugs before COVID-19

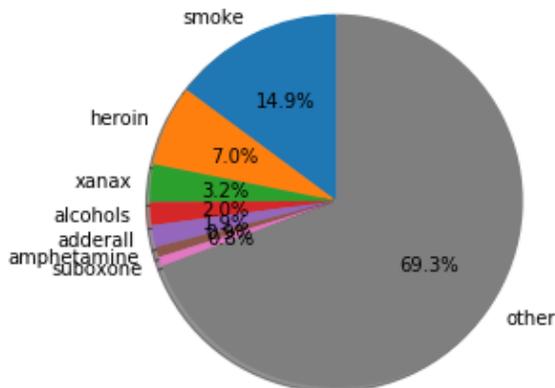
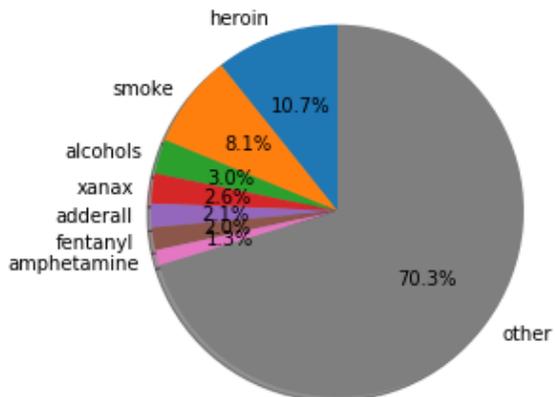


Fig. 3. Top mentioned drugs during COVID-19



#### IV. CONCLUSION

The results stated above were indicative of our initial assumption that the COVID-19 pandemic will lead to a significant change in online drug discussion. The number of drug-related conversations increased significantly in social media and the most talked-about drug categories also have notable changes. In addition, the emotions that were related to such discussion also went through statistically significant changes from the pre-pandemic period to the current day. These changes possibly reflect the changes in people's drug-use patterns as well.

The COVID-19 pandemic led to people staying home more often, usually in an isolated situation. With more time spent alone and at home, people may rely more on drugs or decide to try new drugs as well. The resulting consequences tend to be relatively harder to notice as they usually manifest in mental health issues or non-tangible symptoms. Social media can help reveal some of these changes which makes it an important information source.

#### ACKNOWLEDGMENT

We acknowledge the help of our advisor Prof. Amarnath Gupta and co-advisor Prof. Annick Borquez for their guidance. We also acknowledge the help of Dr. Subhasis Dasgupta who manages the Advanced Query Processing Lab at the San Diego Supercomputer Center.

#### REFERENCES

- [1] U. N. O. of Drugs and Crime, *World Drug Report 2020*. United Nations Publication, Sales No. E.20.XI.6), Dec 2020.
- [2] S. Slavova, P. Rock, H. M. Bush, D. Quesinberry, and S. L. Walsh, "Signal of increased opioid overdose during covid-19 from emergency medical services data," *Drug and alcohol dependence*, vol. 214, p. 108176, 2020.
- [3] S. Chiappini, A. Guirguis, A. John, J. M. Corkery, and F. Schifano, "Covid-19: The hidden impact on mental health and drug addiction," *Frontiers in psychiatry*, vol. 11, p. 767, 2020.
- [4] M. Junghanns, M. Kießling, N. Teichmann, K. Gómez, A. Petermann, and E. Rahm, "Declarative and distributed graph analytics with gradoop," *Proceedings of the VLDB Endowment*, vol. 11, no. 12, pp. 2006–2009, 2018.
- [5] A. Green, M. Junghanns, M. Kießling, T. Lindaaker, S. Plantikow, and P. Selmer, "opencypher: New directions in property graph querying," in *Proceedings of the 21st International Conference on Extending Database Technology, EDBT 2018, Vienna, Austria, March 26-29, 2018*, M. H. Böhlen, R. Pichler, N. May, E. Rahm, S. Wu, and K. Hose, Eds. OpenProceedings.org, 2018, pp. 520–523.
- [6] Q. Zhu, X. Ren, J. Shang, Y. Zhang, F. F. Xu, and J. Han, "Open information extraction with global structure constraints," in *Companion Proceedings of the The Web Conference 2018*, 2018, pp. 57–58.
- [7] J. Shang, J. Liu, M. Jiang, X. Ren, C. R. Voss, and J. Han, "Automated phrase mining from massive text corpora," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 10, pp. 1825–1837, 2018.
- [8] D. Cameron, G. A. Smith, R. Daniulaityte, A. P. Sheth, D. Dave, L. Chen, G. Anand, R. Carlson, K. Z. Watkins, and R. Falck, "Predose: a semantic web platform for drug abuse epidemiology using social media," *Journal of Biomedical Informatics*, vol. 46, no. 6, pp. 985–997, 2013.
- [9] D. M. Low, L. Rumker, T. Talkar, J. Torous, G. Cecchi, and S. S. Ghosh, "Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during covid-19: Observational study," *Journal of medical Internet research*, vol. 22, no. 10, p. e22635, 2020.

- [10] L. B. Soares, N. FitzGerald, J. Ling, and T. Kwiatkowski, "Matching the blanks: Distributional similarity for relation learning," *arXiv preprint arXiv:1906.03158*, 2019.
- [11] "Limbic: Python package for emotion analysis from text," <https://glhuilli.github.io/limbic-package.html>, accessed:2021-03-07.