

Spotting the diffusion of New Psychoactive Substances over the Internet

Fabio Del Vigna^{1,2}, Marco Avvenuti¹, Clara Bacciu², Paolo Deluca³,
Marinella Petrocchi², Andrea Marchetti², and Maurizio Tesconi²

¹ Dept. of Information Engineering, University of Pisa, Italy

² Institute of Informatics and Telematics (IIT-CNR), Pisa, Italy

³ Institute of Psychiatry, Psychology & Neuroscience, King's College London, UK

marco.avvenuti@unipi.it

{f.delvigna,c.bacciu,m.petrocchi,a.marchetti,m.tesconi}@iit.cnr.it

paolo.deluca@kcl.ac.uk

Abstract. Online availability and diffusion of New Psychoactive Substances (NPS) represents an emerging threat to healthcare systems. In this work, we analyse drugs forums, online shops, and Twitter. By mining the data from these sources, it is possible to understand the dynamics of drug diffusion and its endorsement, as well as timely detect new substances. We propose a set of visual analytics tools to support analysts in tackling NPS spreading and provide a better insight about drugs market and analysis.

Keywords: NPS data mining, drugs forums, NPS online shops, data visualisation and exploration, NPS detection, visual analytics, social media analysis.

1 Introduction

Noticeably, health departments of European countries are facing a raising issue: the online trade of substances that lay in a grey area of legislation, known as New Psychoactive Substances (NPS). European Union (EU) continuously monitors the market to tackle NPS diffusion, forbid NPS trade and sensitise people to the harmful effects of these drugs⁴. Unfortunately, legislation is typically some steps back and newer NPS quickly replace old generation of substances.

Online shops and marketplaces convey NPS through the Internet [20], without any (or with very few) legal consequences. Quite obviously, this attracts drug consumers, which can legally buy these drugs without risk of prosecution. The risks connected to this phenomenon are high: every year, hundreds of consumers get overdoses of these chemical substances and hospitals have difficulties to provide effective countermeasures, given the unknown nature of NPS. Furthermore, products sold over the Internet with the same name may contain different substances, as well as possible changes in drug composition over time [5].

⁴ <http://www.emcdda.europa.eu/start/2016/drug-markets#pane2/4>; All URLs in the paper have been accessed on July 10, 2016.

Social media and specialised forums offer a fertile stage for questionable organisations to promote NPS as a replacement of well known drugs, whose effects have been known for years and whose trading is strictly forbidden. Furthermore, forums are contact points for people willing to experiment with new substances or looking for alternatives to some chemicals, but also a discussion arena for those at the first experiences with drugs, as well as trying to stop with substance misuse or looking for advice regarding doses, assumption and preparation.

The EU-funded project Cassandra⁵ investigates the NPS supply chain, lifecycle, and endorsement, through the analysis of popular social media, drug forums, and online shops. Such analysis is vital to timely detect NPS diffusion: this will support governments and health agencies in confining the progress of substance abuse, prohibiting NPS sales and improving the awareness of citizens towards unhealthy and harmful behaviours.

In this paper, we shed light on the structure and activity of NPS forums and online shops. The main contributions are as follows: i) we give an insight into two popular forums, Bluelight⁶ and Drugsforum⁷, hosting drugs discussions since more than one decade; ii) we map NPS sales (as monitored on online shops) and NPS diffusion and distribution (as monitored on discussion forums); and iii) we provide automatic support to timely NPS detection.

Overall, we show a successful application of Intelligent Data Analysis techniques to complex systems, such as social networks and hierarchical ones. This eases the human exploration and interpretation of the online universe of drugs, with a support for the interactive visualisation of the data analysis results.

The paper is structured as follows. Next section gives related work in the area. Section 3 gives a panoramic view on our data sources. In Section 4, we focus on forums and shops structure and activities, by analysing their data and offering a visualisation of the analysis results. Finally, Section 5 concludes the paper and gives directions for future work.

2 Related Work

Recently, academia has started investigating the massive use of social media and online forums to advertise and discuss about psychedelic substances and drugs, and how the preferences of online communities can affect those of consumers. Large forums drew attention, being a primary source of information about NPS and a good sample of consumer tastes [4]. Work in [14] considers the Flashback forum and traces the trend of the discussions, especially in relation with the scheduling of a substance ban. The paper puts in evidence how volumes of discussions drop when a ban is scheduled. In [22], the authors focus on new drugs detection and categorisation by scanning online shops and the dark net. A complete list of the known effects of new drugs, to the publication date, is given in [20,11].

⁵ <http://www.projectcassandra.eu>

⁶ <http://www.bluelight.org>

⁷ <https://drugs-forum.com>

Small subsets of the contents of the Drugsforum and Bluelight forums, which we deeply analyse in the present paper, have been already considered in [21], highlighting how large forums embody a cumulative community knowledge, i.e., a stratified knowledge built over years of forum activities, and showing that drugs effects and dosage are among the most discussed topics.

Other studies explored the abuse of medicines and how these are advertised, e.g., on Twitter, and sold by online pharmacies, with no authorisation [13,6]. Twitter features a rapid spread of contents, especially through small communities of users, which share common interests and tastes. This is the main reason why it has been investigated to mine patterns of drug abuse, also for non-medical purposes, e.g., to improve students performances in study [8,7]. Furthermore, Twitter allows analysts to comprehend rapid disease diffusion and health issues [17], as well as prices and effects of new drugs [16]. Nevertheless, social media play an important role also for contrasting the drugs diffusion [19] and for preventing end users from further consumption [12]. Twitter was also extensively mined to detect geographical diffusion of drug consumers over time [3].

The Web is not the only marketplace where NPS are advertised and sold. Indeed, the TOR network⁸ has drawn much attention from drug consumers and resellers, who search for a channel to buy and sell drugs that guarantees their anonymity. This aspect affects trustworthiness of peers, especially when it is not possible to assess users reputation at all. In [9], the authors investigated the impact of reputation in Silk Road, one of the most popular marketplaces for drugs in the dark net. Data analysis often deals with the quality of the results obtained when searching the web. The work in [18] describes the possibility to improve the recall of queries issued to search engines by exploiting all variants and misspelled words.

With respect to related work, this paper addresses a finer-grained, more detailed picture of NPS data sources and NPS data available on the Internet. As an example, the analysis of forums carried on in [21] was limited in time and quantity. In our work, we overcome this limitation, by analysing more than one decade of data, posted by users all over the world. Overall, we dealt with more than 4 million and a half posts and more than 500,000 users. Furthermore, we integrated more than one source, by monitoring two forums, Twitter, and a number of online shops. The results of our analysis are conveniently conveyed to the reader via a set of interactive visual web interfaces, which are being integrated into a dashboard that will help researchers mine the wealth of gathered data. Ultimately, we are aligned with recent advances in data analysis leading to applications in pattern mining of, e.g., medical records and human anatomies [2,10].

3 Data sources

This section presents the data sources for our analysis. We collected the data by developing ad-hoc software, which scrapes websites and uses APIs to crawl social media.

⁸ <https://www.torproject.org>

3.1 Forums

Bluelight and Drugsforum are two large forums, which host more than a decade of discussion about drugs and addiction. Being particularly rich of information, the two forums provide a historical, worldwide background of drug consumption, comprising that related to NPS. Similar to Google Flu Trends⁹ efforts to detect spreading of diseases, the analysis of the forums' content and structure is significant to understand how psychoactive substances have spread out and to study new infoveillance strategies, to timely detect drugs abuse.

The two forums have a hierarchical structure, which enables proper content categorisation. The root of both forums organises content into sub-forums, which can be nested up to several levels of depth. The forums' structures were subject to different content re-organisations over time.

We carried out a Web scraping activity to create a dump of the entire database of discussions from the two forums, following the links between the forums' sections. During the storage phase, we kept track of the forums' hierarchy and structure, maintaining all the tags and metadata associated to each post and thread. Table 1 summarises the amount of data available from the two forums.

3.2 Online shops

The forums introduced in Section 3.1 are a primary source of information about drugs reviews, feelings, effects and preparation, but little information is available about the drugs market, such as prices and bulk quantities. Thus, we focused our attention also on other data sources, dealing with drugs trading.

Online shops sell both legal and illegal substances. Among others, those that sell NPS have grown in popularity, given the relatively low risks in trading such substances. Many online shops accept payments in pounds, euros and dollars. Also, bitcoins are often accepted. This opens up the possibility to track price trends and, indirectly, to estimate the popularity and quality (or purity) of drugs. Furthermore, many of the marketplaces are advertised and mentioned on forums and social media.

We have started an intense scraping activity on a set of online shops to monitor the market availability of different substances. Online shops can be quite easily found through simple queries to search engines (e.g., "legal highs" and "smart drugs"). We set up a battery of scrapers that collect the information that are present on the shops showcases. Data is collected on a weekly basis, and stored in a relational database, to be easy queryable. Table 2 shows the monitored shops.

3.3 Twitter

Twitter is extensively used by resellers and "pharmacies" to advertise psychoactive substances, and by consumers to discuss their effects and share feelings with

⁹ <https://www.google.org/flutrends/about/>

Forum	First post	Last post	Tot posts	Users
Bluelight	22-10-1999	09-02-2016	3,535,378	347,457
Drugsforum	14-01-2003	26-12-2015	1,174,759	220,071

Table 1. Drug forums: Posts and Users

ID	Website	Substances found
1	http://chem-shop.co.uk	7
2	http://researchchemist.co.uk	45
3	http://researchchemistry.co.uk	56
4	http://sciencesuppliesdirect.com	43
5	http://www.bitcoinhighs.co.uk	4
6	http://www.buylegalrc.eu	17
7	http://www.legalhighlabs.com	33
8	http://www.ukhighs.com	51
9	https://www.buyanychem.eu	78
10	https://www.iceheadshop.co.uk	68

Table 2. Monitored online shops and number of substances they sell

others [6,13]. We have collected about 14 million tweets, over the period March 16, 2015 - February 2, 2016, using the Streaming API¹⁰, which allows applications to gather tweets in real time fashion. We have used a crawler that fetches data relying on a set of ad-hoc keywords. We have also followed a series of Twitter accounts associated to online shops. In the next section, we will detail the monitored keywords, which we chose among known emerging substances.

4 Data analysis and Visualization

This section shows the analysis we have carried out over the data sources described in Section 3. First, we report on a series of analyses over the two drug forums, with the purpose of figuring out their structural features, how their content is organised, and the geographical distribution of their users. Secondly, we mine the forums textual contents, aiming at looking for new substances mentioned in recent discussions. Finally, we provide a picture on the NPS substances sold on online shops, correlating them with mentions on Twitter and the forums.

4.1 Forums: Structural and geographical features

To facilitate the investigation of the forums structural features, we have developed a set of visual interfaces. Figure 1 depicts the screenshot of a zoomable treemap of the two forums. Nested subsections are represented as nested rectangles, the area of which are proportional to the number of posts a subsection contains. Quick visual comparisons of the forums' size and structure may gather

¹⁰ <https://dev.twitter.com/streaming/overview>

meaningful information. For example, compared to Drugsforum, whose structure is quite complex, Bluelight has a shallow organisation. Also, the names of the subsections suggest that the discussion on Drugsforum is mainly focused on drugs and it follows a rigid categorisation, based on the kind of the substance, while the topics on Bluelight are broader and less related specifically to drugs.

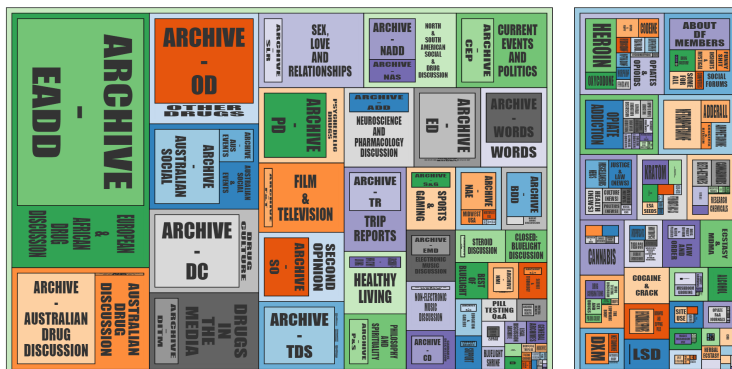


Fig. 1. The structure of Bluelight (left) and Drugsforum (right). Bluelight is about three times bigger.

Figure 2 shows the worldwide distribution of the Drugsforum users. The information has been extracted from the users' profiles (when available). Looking at the figure, we understand that drugs discussions on forums is a wide phenomenon, quite naturally leading to a widespread word of mouth. The colours in the figure are proportional to the density of users. Noticeably, the most involved areas are North America, Australia, UK, and Scandinavia.

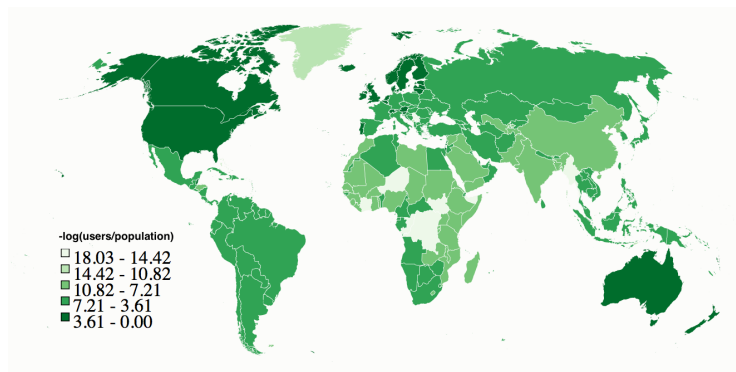


Fig. 2. Geographical distribution of Drugsforum users.

We have also investigated some topological aspects of the forum, like the number of posts per user and the number of posts per thread, on both forums (Figure 4). With the *powerlaw* Python package [1], we have compared the four real data distributions with the exponential, power law, truncated power law and lognormal distributions. The tool measured the *xmin*, no more than 4 for all the cases. Furthermore, with a *p*-value less than 10^{-8} for all the distributions, the power law distribution results in a better fit than the exponential one, as expected [15]. With regard to the lognormal and truncated power law distributions, the lognormal distribution fits slightly better than the power law one, while the truncated power law distribution fits better than the lognormal one. We can conclude that the (truncated) power law distribution assumption holds, as shown in figs. 3a to 3d. These results highlight that there is a small amount of users responsible for most of the activity, on both forums.

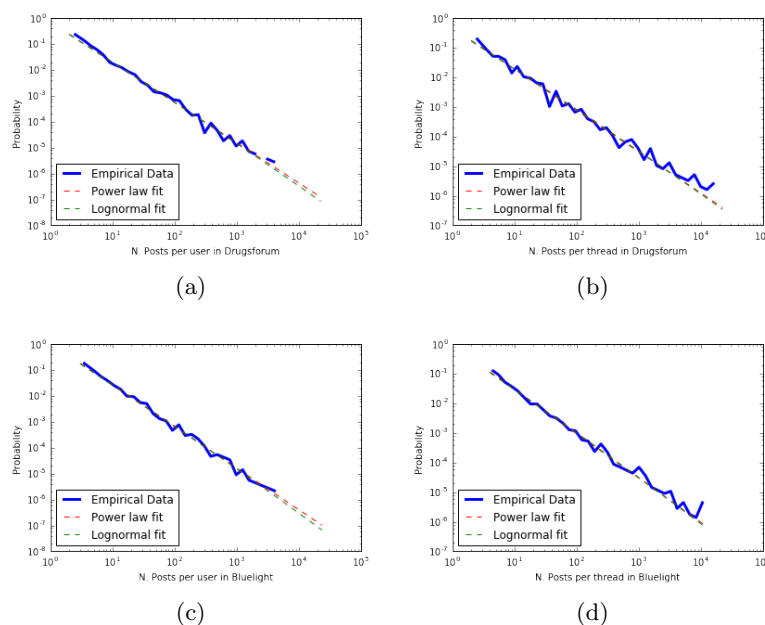


Fig. 3. Probability Density Function of the real data and the different distributions

It is worth noting that, even if Bluelight has about 0.6 times the number of users Drugsforum has (see Table 1), the number of active users (i.e., that have written at least one post) is almost the same for both. As for the distribution of posts per thread, shown in Figure 5, Bluelight features a large number of threads having 1,000 posts. This is due to a limit on the maximum number of posts for certain threads: when exceeding the threshold, the moderators start a new thread for the discussion.

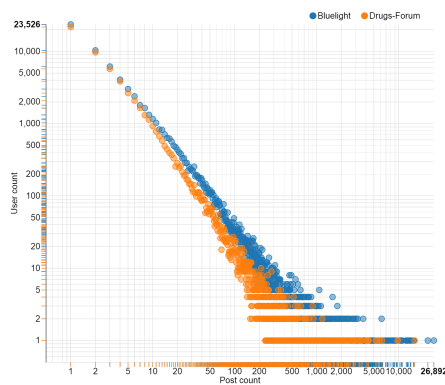


Fig. 4. Posts per user.

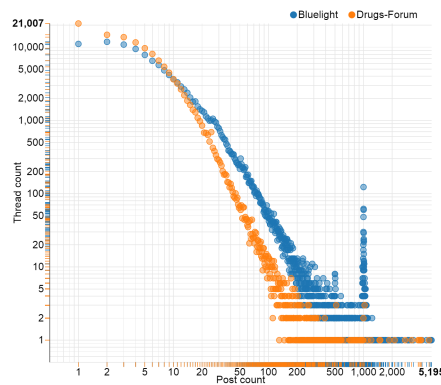


Fig. 5. Posts per thread.

4.2 Content analysis

A text analysis that is really useful in our scenario is the measurement of volumes of discussion over time, given a term. This investigation helps determining whether some drugs raise in popularity and in which section of the forum this happens, possibly obtaining some clues about the nature of the substance (being it a NPS or not).

Figure 6 shows the frequency of the term “mephedrone” over time, normalised to the whole volume of discussion, for Drugsforum (top) and Bluelight (bottom). Even if not identical, the shapes of the spike are similar, meaning that the substance has gained popularity within both the communities approximately at the same time.

Figure 7 shows a higher level of detail: each line represents a subsection of the forum. As shown in the top-left part of the screenshot, we can choose which forum to analyse. A darker colour indicates a higher frequency of the term, for the corresponding time frame. The search for “mephedrone” in Drugsforum shows a high volume of discussion in the first half of 2010 in a series of subsections, particularly in the one called “Beta-Ketones”. This indicates the category of the substance.

As shown in the example of Figure 8, computing the terms that co-occur with a given one gives interesting insights. Indeed, the generated wordclouds may provide knowledge on substances that are similar, with similar effects and market trends. In the figure, each word occupies an area that is proportional to its frequency. The wordclouds can be generated for both Twitter and the two forums.

Really endorsed drugs are presented and discussed in forums. To timely detect NPS, we have investigated neologisms and terminology on both the forums, to discover new names. As an example, in Figure 9, we plot the Drugsforum terms that appeared only after 2010. The result clearly indicates a lot of new drugs, appeared on the market from 2010 to 2015. It is possible to notice the name

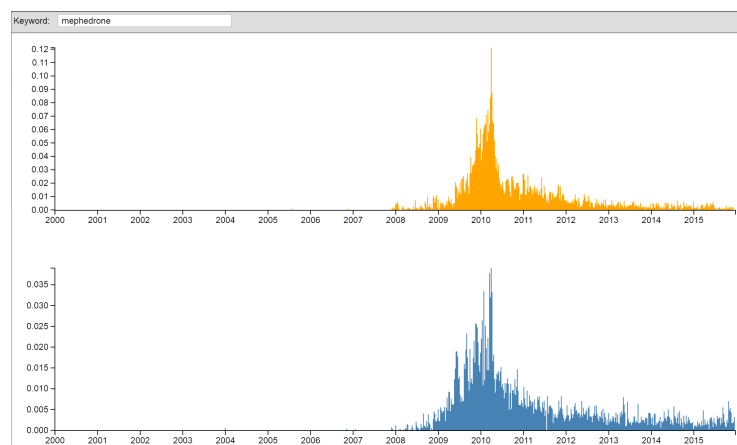


Fig. 6. Frequency of “mephedrone” over time, normalised to the whole volume of discussion, for Drugsforum (top) and Bluelight (bottom).

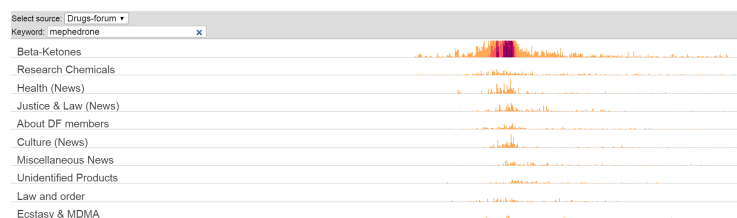


Fig. 7. Horizon charts showing the frequency of a given term over time, for each sub-section of the chosen forum.

of some new drugs and medicines, such as α -PVP, Diclazepam, Pentedrone, Naphyrone.

4.3 NPS trading

As a final set of analyses, we have explored the hyperlinks on the forums. Then, we have compared them with a comprehensive list of NPS online shops and with the links in the posts of monitored Twitter accounts. Not surprisingly, they do not overlap, meaning that forums discussions do not link shops. This is mainly due to the specific policies of the forums. We have also tested which are the NPS sold on the shops and also mentioned on forums, finding that almost every substance is mentioned. It is not possible to estimate the trade volume of NPS from online shops, but we can try to infer some information about popularity by observing the discussions in forums. Checking the words frequency, we concluded that the very same substances are also advertised through Twitter. Table 3 reports an excerpt of some substances, with a measure of the discussion activity about them on Twitter, on forums and on online shops. In the table, the numbers

surrogates, which leaves few margin for intervention by healthcare institutions and governments. This paper has put in evidence some unique features of online NPS forums and shops. Monitoring such websites and elaborating the available data made it possible to explore a large quantity of information, also across platforms, allowing analysts to perform comparisons among them. We also gave a measurement of the relevance of NPS diffusion and advertisement, as well as user engagement. Furthermore, we showed how trading and discussions are correlated, through terms used by both online shops, social media, and forums, despite the prohibition, which hold on forums, to post explicit links to shops. Noticeably, co-occurrences analysis and temporal analysis of neologisms are a valid support for NPS detection.

Currently, the analyses are led by the data scientist, which is assisted by the developed software. The analyses are applicable both to offline datasets and online streaming sources. We aim at fully automating some of the work, e.g., the detection of the psychic and physical effects of NPS on the individual, based on comments by the online users. Finally, we plan to extend the analysis to the dark web marketplaces.

6 Acknowledgements

This publication arises from the project CASSANDRA (Computer Assisted Solutions for Studying the Availability aNd DistRibution of novel psychoActive substances), which has received funding from the European Union under the ISEC programme: Prevention of and fight against crime [JUST2013 / ISEC / DRUGS / AG / 6414].

References

1. Alstott, J., Bullmore, E., Plenz, D.: powerlaw: a Python package for analysis of heavy-tailed distributions. *PloS one* 9(1), e85777 (2014)
2. de Bono, B., Grenon, P., Helvensteijn, M., Kok, J., Kokash, N.: Apinatomy: Towards multiscale views of human anatomy. In: *Advances in Intelligent Data Analysis XIII*, pp. 72–83. Springer (2014)
3. Buntain, C., Golbeck, J.: This is your Twitter on drugs: Any questions? In: *24th World Wide Web Conference - Companion Volume*. pp. 777–782. ACM (2015)
4. Davey, Z., Schifano, F., Corazza, O., Deluca, P.: e-Psychonauts: Conducting research in online drug forum communities. *Mental Health* 21(4), 386–394 (2012)
5. Davies, S., et al.: Purchasing legal highs on the Internet - is there consistency in what you get? *QJM* 103(7), 489–493 (2010)
6. Freifeld, C.C., Brownstein, J.S., Menone, C.M., Bao, W., Filice, R., Kass-Hout, T., Dasgupta, N.: Digital drug safety surveillance: monitoring pharmaceutical products in Twitter. *Drug Safety* 37(5), 343–350 (2014)
7. Hanson, C.L., Burton, S.H., Giraud-Carrier, C., West, J.H., Barnes, M.D., Hansen, B.: Tweaking and tweeting: exploring Twitter for non medical use of a psychostimulant drug (Adderall) among college students. *Journal of Medical Internet Research* 15(4), e62 (2013)

8. Hanson, C.L., et al.: An exploration of social circles and prescription drug abuse through Twitter. *Medical Internet Research* 15(9) (2013)
9. Hardy, R.A., Norgaard, J.R.: Reputation in the Internet black market: an empirical and theoretical analysis of the Deep Web. *Journal of Institutional Economics* FirstView Article, 1–25 (2015)
10. Hielscher, T., Spiliopoulou, M., Völzke, H., Kühn, J.P.: Mining longitudinal epidemiological data to understand a reversible disorder. In: *Advances in Intelligent Data Analysis XIII*, pp. 120–130. Springer (2014)
11. Hillebrand, J., Olszewski, D., Sedefov, R.: Legal highs on the Internet. *Substance Use & Misuse* 45(3), 330–340 (2010)
12. Inciardi, J.A., Surratt, H.L., Cicero, T.J., Rosenblum, A., Ahwah, C., Bailey, J.E., Dart, R.C., Burke, J.J.: Prescription drugs purchased through the Internet: who are the end users? *Drug and Alcohol Dependence* 110(1), 21–29 (2010)
13. Katsuki, T., Mackey, T.K., Cuomo, R.: Establishing a link between prescription drug abuse and illicit online pharmacies: Analysis of Twitter data. *Journal of Medical Internet Research* 17(12) (2015)
14. Ledberg, A.: The interest in eight new psychoactive substances before and after scheduling. *Drug and Alcohol Dependence* 152, 73 – 78 (2015)
15. Muchnik, L., Pei, S., Parra, L.C., Reis, S.D., Andrade Jr, J.S., Havlin, S., Makse, H.A.: Origins of power-law degree distribution in the heterogeneity of human activity in social networks. *Scientific Reports* 3 (2013)
16. OConnor, K., Pimpalkhute, P., Nikfarjam, A., Ginn, R., Smith, K.L., Gonzalez, G.: Pharmacovigilance on Twitter? Mining tweets for adverse drug reactions. In: *AMIA Annual Symposium*. p. 924. American Medical Informatics Association (2014)
17. Paul, M.J., Dredze, M.: You are what you tweet: Analyzing Twitter for public health. *ICWSM 20*, 265–272 (2011)
18. Pimpalkhute, P., Patki, A., Nikfarjam, A., Gonzalez, G.: Phonetic spelling filter for keyword selection in drug mention mining from social media. *AMIA Summits on Translational Science* p. 90 (2014)
19. R. Scott, K., Nelson, L., Meisel, Z., Perrone, J.: Opportunities for exploring and reducing prescription drug abuse through social media. *Journal of Addictive Diseases* 34(2-3), 178–184 (2015)
20. Schmidt, M.M., Sharma, A., Schifano, F., Feinmann, C.: Legal highs on the net: Evaluation of UK-based websites, products and product information. *Forensic Science International* 206(1), 92–97 (2011)
21. Soussan, C., Kjellgren, A.: Harm reduction and knowledge exchange—a qualitative analysis of drug-related Internet discussion forums. *Harm Reduction Journal* 11(1), 1–9 (2014)
22. Zawilska, J.B., et al.: Next generation of novel psychoactive substances on the horizon – a complex problem to face. *Drug and Alcohol Dependence* 157, 1 – 17 (2015)