Detecting Emerging Products in TOR Network Based on K-Shell Graph Decomposition

Mhd Wesam Al Nabki
Researcher at INCIBE
University of León
mnab@unileon.es

Eduardo Fidalgo
Researcher at INCIBE
University of León
eduardo.fidalgo@unileon.es

Enrique Alegre
Researcher at INCIBE
University of León
ealeg@unileon.es

Victor González-Castro
Researcher at INCIBE
University of León
victor.gonzalez@unileon.es

***Abstract*—In this paper, we present a semi-automatic framework which allows identifying the most popular and also some of the illegal emerging products that are sold in marketplaces located in the Darknet. Using textual information extracted from Darknet domains, we built a Products Correlations Graph (PCG), where the nodes are Darknet products and the edges reflect a simultaneous offering of two products. By applying the k-shell algorithm to decompose the PCG graph, we identified the products contained into the core-shell and we identified the most popular and emerging ones. We applied our emergent detection algorithm to the dataset named Darknet Usage Text Addresses (DUTA), detecting MDMA and Ecstasy as the most relevant and emerging drugs respectively, validating these results against the report of prestigious international drugs organisations. These results make our framework a complementary tool to extract information in illegal markets where transaction logs are not available.**

***Index Terms*—Darknet, TOR, Cybersecurity , Graph Theory, data mining, k-shell**

I. Introduction

The Darknet is the part of the Deep Web whose content is not indexed and a portion of the Web which, until some years ago, can be only accessed through a dedicated browser. One of the most popular Darknet networks is "The Onion Route" (TOR)[1], which provides a high level of privacy and anonymity for domain owners and their visitors. In the recent years, projects like *TOR2WEB* allows the users of the Surface Web to access directly to the content of TOR employing a standard browser, instead of starting an instance in a dedicated one, i.e. TOR browser[2]. When a user of the Surface Web wants to access the content of a TOR address, it should only replace the .onion address extension with a specific one, e.g. *onion.link, onion.cab or onion.tor*.

The high degree of protection offered by TOR has encouraged to the contraband traders to market their products widely and the consumers to shop freely, as the payments are made pseudo-anonymously in Bitcoin, an encrypted digital currency[3]. However, as the purchasing transactions logs are not always available [1], there is still need to monitor the products trends through the offering markets. A conventional control process as could be the manual inspection of a TOR address, represents a significant amount of time in the work schedule of a single person and the need of having specific knowledge of the subject being monitored. All these problems would be solved using a system that can visit the TOR address, collect their data, automatically process their content and provide the requested feedback.

In this paper, we present a semi-automatic framework for emerging products detection on the Darknet marketplaces. The work-flow chart of this framework is depicted in Figure 1

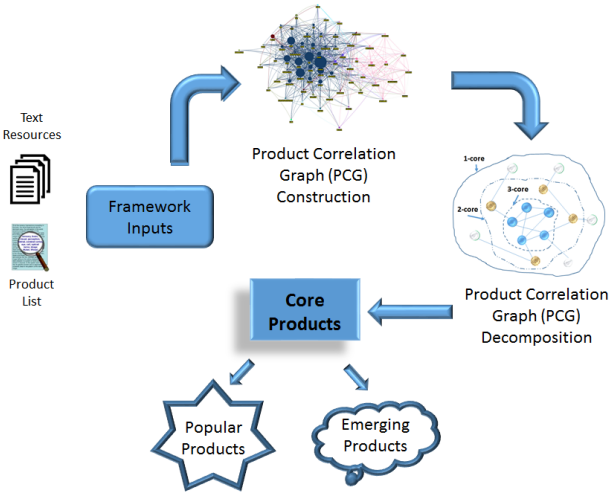

Fig. 1. Stages of the emerging products detection

Our proposal is based on analysing the relationships between the offered products rather than studying the purchasing transactional logs. Using graph theory, we construct a graph containing products correlations where the nodes represent the products and the edges correspond to their existence inside the same marketplace. By adopting the K-shell graph decomposition algorithm [2], we developed a new technique to identify the emerging products in the Darknet domains. We assessed our framework over a set of TOR drugs domains which are extracted from the "Darknet Usage Text Addresses" (DUTA) dataset [3][4]. The results are quantitatively justified by comparing them with the most recent reports published by international drugs organisations, and visually by analysing the products correlations graph. The proposed framework might be helpful in identifying seasonal trends and emerging illegal products in different personal shops and marketplaces located in TOR webs.

The rest of the paper is organised as follows: First, in Section II we present the related work regarding trend detection. Next, in Section III we introduce the k-shell algorithm

[1]https://www.torproject.org/
[2]https://www.torproject.org/projects/torbrowser.html.en
[3]https://bitcoin.org

[4]In this analysis we have used DUTA version 1.1. In [3] it was used version 1.0, i.e. the initial version

and the techniques related to the association rules learning. The framework methodology is described in Section IV, together with the details about the graph implementation of the problem. After that, in Section V we present our experiments and how the results are validated. Finally, the conclusions and reference to future lines of work are drawn in Section VI.

II. Related Work

Several researchers have investigated the problem of Emerging Trend Detection (ETD) from the perspective of purchases transactions logs. The ETD process falls under two broad categories [4]: On the one hand, the fully-automatic approach takes a text corpus and generates a list of emerging items [5]. On the other hand, the semi-automatic approach depends on a predefined list of articles, which is used by the system to return the detected emerging ones [6], [7], [8]. Glance et al. [5] proposed a temporal analysis for the weblogs, i.e. the topics, the people and content, where they discovered trends based on the application of data mining techniques and Natural Language Processing algorithms. Porter et al. [8] proposed a semi-automatic system for the analysis of technology opportunities which depends on a list of potential keywords set by experts. These keywords were combined into queries and prepared to be an input for the Technology Opportunities Analysis System (TOAS), which extracts the relevant documents with a feature vector that contains information such as words count and date information. Kontostathis et al. [4] used the fully-automatic approach to detect the trends in a timestamped dataset, where they created a matrix and the relationships between documents and terms were modelled. The matrix dimensions were reduced using the Singular Value Decomposition and to finish the process they used a similarity function to cluster the noun phrases which were closely associated. Raeder et al. [9] presented a framework for market basket analysis by modelling the purchase transactions as a product network, and then they clustered the built network using a community detection algorithm.

III. Background

The framework proposed in this paper is based on graph theory and data mining techniques. Therefore, in this Section, we will review the concepts of *degree centrality*, a measure based on the connections that each node has into the graph, and *item support/confidence* that we borrow from the data mining field. To end this section, we present an overview of the k-shell decomposition algorithm.

A. Nodes Measures

The degree centrality of a node in a graph is a well-known concept in graph theory which is defined by the number of links that a specific node has. In our graph, nodes represent items or products. Let $D = d_1, d_2, ..., d_m$ be a set of m products and $T = t_1, t_2, ..., t_n$ a set of n transactions, where each transaction in T contains a subset of products in D. An association rule is defined as $d_i => d_j$, being d_i and d_j elements of D, and this represents that when the product d_i is bought, a customer also bought the product d_j. In the field of data mining, a set containing these two items d_i and d_j it is called *itemset*. The popularity of an item d_i within a set of transactions T is defined as *Support*, and indicates the proportion of transactions in T where the item d_i appears. Eq. 1 shows the support of an *itemset* with one item d_i.

$$Support(d_i) = \frac{Frequency(d_i)}{ItemsetLength} . \quad (1)$$

Another measure that we will use during our experimentation is the *Confidence*, which indicates how often a defined rule between two items is true, e.g. in the rule $d_i => d_j$ how likely item d_j is purchased when item d_i is also purchased (Eq. 2).

$$Confidence(d_i) = \frac{P(d_i \cap d_j)}{Support(d_i)} , \quad (2)$$

being $P(d_i \cap d_j))$ the probability that the products d_i and d_j are purchased together into the same transaction of T.

B. Graph Decomposition Algorithm

The k-shell algorithm is based on k-core algorithm [11]. Let $G = (V, E)$ be a graph where V are the nodes and E are the edges, $|V| = v$ and $|E| = e$. Being the *degree* the number of edges incident in a node, the k-core of G is defined as a sub-graph $H = (C, E|C)$ induced by the subset $C \subseteq V$ where all the nodes have a degree of at least k. In other words, the sub-graph H will have an order k when the condition in Eq.3 is satisfied.

$$\forall v \in C : degree_H(v) \geq k . \quad (3)$$

This means that the k-core of a graph G is calculated recursively by eliminating all the nodes whose degree is less than k, repeating the process until all nodes in the remaining graph have at least degree k (see Figure 2), whereas the k-shell of G is the sub-graph of nodes in the k-core but not in the $(k+1)$-core.

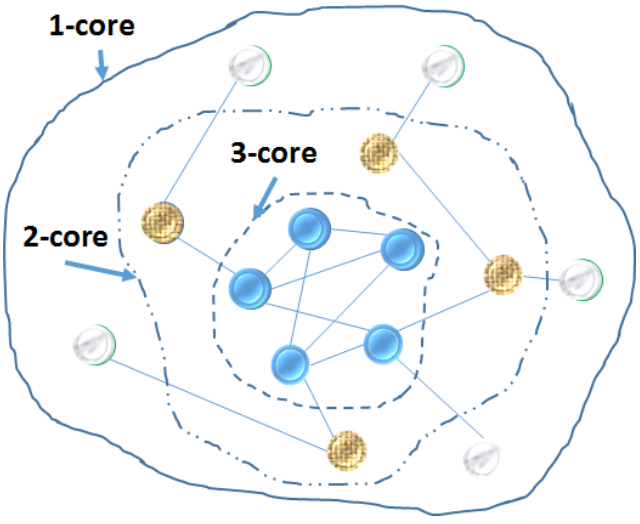

Fig. 2. Example of a K-core with k=1, k=2 and k=3 [14]

IV. Methodology

In this section, we review the four parts of the proposed framework. First, we explain what are the two inputs of the framework: (1) textual resources, i.e. a dataset, and (2) a list of products related to the field of study. Next, using a graph representation, we explain how a Products Corrections Graph (PCG) is created. After that, we apply the k-shell algorithm, decomposing the graph into levels based on its nodes connectivities. Finally, we run the algorithm proposed to recover a set of emerging products.

A. The Framework Inputs

As it was introduced, the system has two inputs, which are inserted into the system as lists. The first one is a predefined list that is set by experts and contains the relevant products names in the field that we are investigating. In case we were mining emergent sports, the list should include sports names like tennis, soccer or baseball. The second input contains textual resources, that in our case are dumps of marketplaces websites, where the system is going to perform the analysis. Then, by intersecting these two lists, we yield a list of the web resources ID's with their matching products names set.

B. Construction of the Products Correlations Graph

Once the framework inputs have been received and pre-processed, we model our problem using a graph that we named Products Correlations Graph (PCG). The PCG is a weighted undirected graph containing nodes and edges. On the one hand, each node represents a product, and it has three different parameters associated: (i) the frequency of its presence in all the textual resources analysed, (ii) the degree and (iii) the support which is measured as it was explained in Section III-A. On the other hand, an edge is added in the PCG between two products when these two products exist in the same textual resource. In other words, an edge between a product d_1 and another different one d_2 is created if they have been offered together in the same marketplace at least once. These edges are weighted by their frequency, i.e. the weight of an edge between the products d_1 and d_2 reflects the number of times that they have been offered together into the same textual resource.

To control the weight of the model, we introduce two thresholds: the first one, we named it α, removes the edges whose weights are less than the average weight of all the edges. The second threshold, β controls the presence of the nodes based on their *supports*, i.e. if the support of a node is less than the average support, the node is removed.

In the literature related with the detection of trending products, a node usually refers to a product, and the edge represents the presence of those products in the same purchasing transaction [9]. However, this approach is not feasible in our case, since we are mining the Darknet markets where the customers' transactions logs are not available. Therefore, we designed this framework to overcome this difficulty and to detect the emerging products based on the offering products records instead.

C. Decomposition of the Products Correlations Graph

After building the PCG, we apply the k-shell algorithm commented in Section III-B to decompose the PCG into shells according to the nodes' level of connectivity.

D. Algorithm for Detecting Emerging Products

Several studies have proved the efficient use of the k-shell algorithm in detecting the relevant nodes within a graph that reside in the deepest shell [11], [12], [13]. Based on this assumption, we propose an Emerging Product Detection Algorithm, which is visually summarised in Figure 3. At this stage, and from the previous PCG decomposition, we took as an input the products present in the core-shell, and we separated them into two groups: popular products and candidates for emerging products. To measure the popularity of the nodes, i.e. products, and to separate the products in the two above mentioned groups, we introduced a new automatic threshold, and we named it γ, which is computed as the average weight of the nodes contained in the core-shell. If a product weight is higher than γ we consider it as popular; otherwise, we categorise it as a candidate to be an emerging product. Finally, after this filtering stage, we sorted the products categorised as candidates according to the sum of their weighted edges with the popular products, and finally, we nominated as emerging products the ones whose weight value was higher than the calculated average. The rest of the nodes are discarded and catalogued as not emerging products.

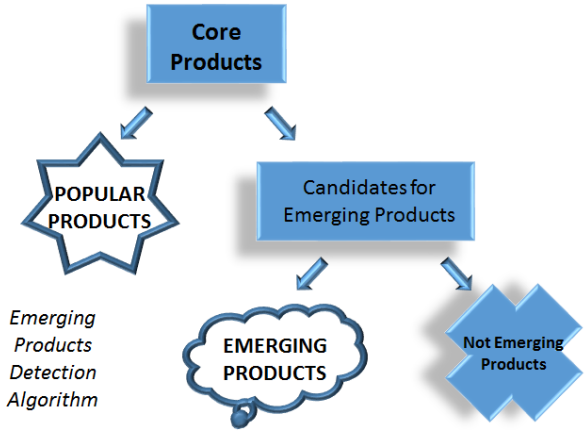

Fig. 3. Visual representation of the Emerging Products Detection Algorithm

V. Experimental Results

A. Experimental Setting

We conducted our experiments on an Intel Core i7 PC with 32GB of RAM. For the graph construction, we used *Python3* with the *NetworkX* library[5]. Concerning the chart visualisation, we applied the *vis.js* library[6]. In this work, we identified the emerging products related to Drugs. In order to do that, the textual resources samples were extracted from a subset of the *Darknet Usage Text Addresses* (DUTA) dataset [3][7]. To the best of our knowledge, DUTA is the most

[5]https://networkx.github.io/
[6]http://visjs.org/
[7]For this analysis, we used DUTA version 1.1

up-to-date publicly available dataset, since it was collected between May and July of 2016 and contains more than $6.8k$ samples of TOR domains dumped as HTML and classified into 26 classes. In our experiments, we have used from DUTA dataset the drugs products which fall into two categories: Drugs (Illegal or legal) and Marketplace (Black). The selected subset has 302 unique Darknet domains addresses, however, after preprocessing and removing the duplicated ones, only 197 different domains related to drugs remained. The list of drugs names was extracted from the drugs.com[8] and druginfo[9] websites. Then, we filtered them manually to remove the polysemic words i.e. the words that have several meanings like "brown" which is a street name for the heroin, and finally, we yielded a list of 862 different drugs.

B. Analysis of Darknet Products (Drugs)

Before start the analysis of emergent products, we introduce some results and key findings derived from our study. First, we ordered the PCG nodes based on their weights, and we found that the most popular products were MDMA, LSD and Cannabis, offered by 24%, 18% and 17% of the Darknet markets, respectively. Figure 4 depicts the list of the top-20 most popular drugs offered during the crawling period when DUTA was created. The previous findings are compatible and represents the evolution of the results of the Global Drug Survey [15], which has indicated that those three products are the most commonly bought in the Darknet during their analysis period (i.e. November-2015 to January-2016).

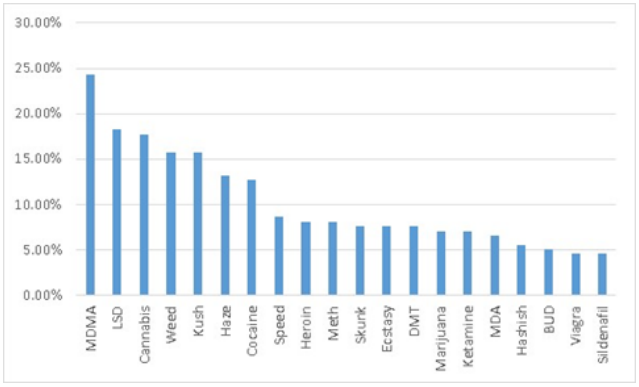

Fig. 4. Top-20 popular drugs in DUTA.

Moreover, we also studied the association rules between the products by ordering the edges based on their weights. The top-20 most popular associations between products are shown in Figure 5. The LSD and the MDMA are the most popular pair of products, followed by the Weed and Haze, since they are present with a frequency of 12.7% and 8.7% in the markets respectively.

Based on the results of the previous experiments, it can be concluded that the proposed method helps to identify the most relevant products and the pairs of products that are mainly offered together. The Table I present the confidence analysis over 10 rules between the most important products discovered in the previous analysis. Each rule is represented as a (A→B) relation and it can be interpreted as how frequent the product B is offered together with the product A.

TABLE I
DRUGS PRODUCTS CONFIDENCE ANALYSIS

Rules	Confidence
(Speed → MDMA)	76%
(LSD → MDMA)	69%
(Haze → Weed)	65%
(Heroin → MDMA)	75%
(Cocaine → LSD)	44%
(MDMA → Heroin)	25%
(Marijuana → Kush)	71%
(Cocaine → MDMA)	64%
(LSD → Meth)	31%
(Kush → Haze)	48%

At this state and following the framework presented in Figure 1, we pre-process the information from a Drug product list, 862 products, and the categories of DUTA previously described, i.e. both legal and illegal drugs and black marketplaces. After this stage, we extracted 395 different drugs, i.e. nodes, with 38347 mutual offering relations between them. Afterwards, both thresholds previously described, i.e. $\alpha = 1$ for edges and $\beta = 0.153$ for nodes, were automatically calculated as explained in Section IV-B). The resulting PCG, after being filtered by the previous thresholds, had 66 nodes and 797 edges with a density of 0.371. By applying the k-shell algorithm, the PCG was decomposed into 12 shells where the core-shell had 27 relevant products with different popularity.

Following the pipeline detailed in Figure 3, we calculated the popular and emerging products in DUTA dataset. Since we do not have a ground truth for the emerging drugs, the results were evaluated qualitatively with reports published by international organisations that have similar statistical studies in the approximately same period to DUTA. Table II shows the candidates for emerging products found in the core-shell. In the shell 19, we had 27 products we ordered by the sum of the weight of their edges, and we selected 9 emerging products from that list since are the one with a weight higher than the average one. Our results indicate that the Ecstasy, which is a famous member of the MDMA drugs family, is the most emerging drug during the crawling period covered by DUTA and it is followed by the Ketamine and the Dimethyltryptamine (DMT).

We validated our findings through the revision of a three

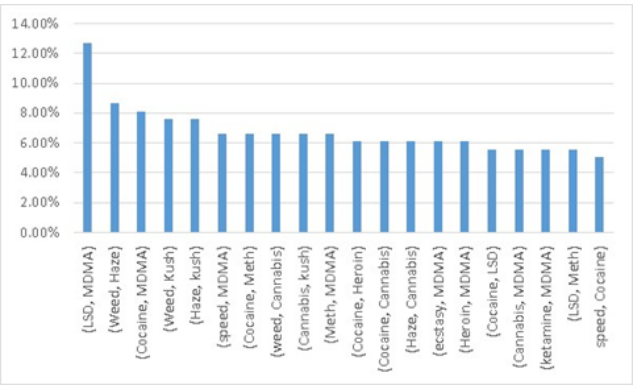

Fig. 5. Top-20 popular association rules of drugs in DUTA.

[8]https://www.drugs.com/alpha/a1.html
[9]https://druginfo.nlm.nih.gov/drugportal/drug/names

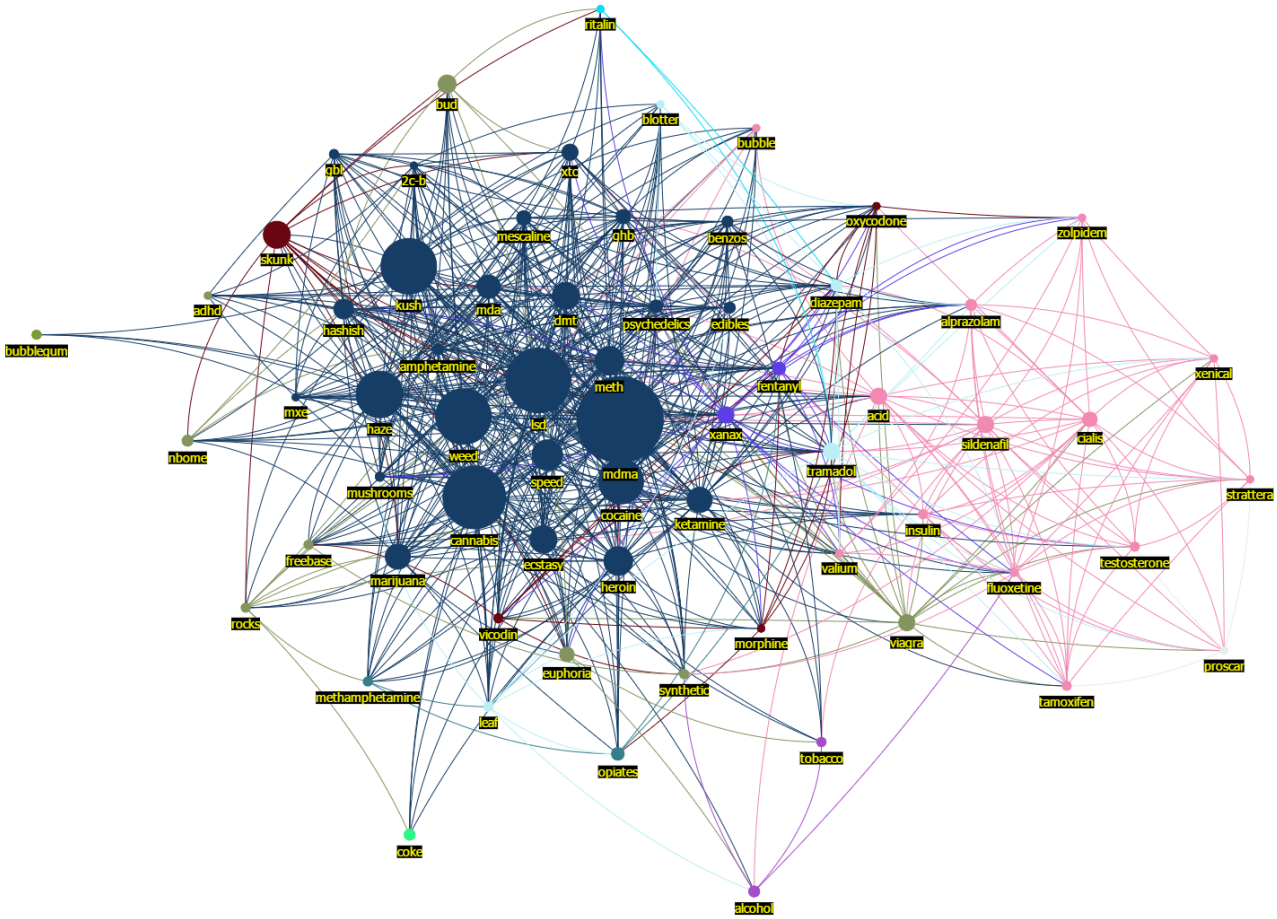

Fig. 6. Drugs PCG graph

different public studies and report. The first one is a study published by the European Monitoring Center for Drugs and Drug Addiction (EMCDDA) [19], which has reported an increase in the use of Ecstasy in adults during 2016. The second one, is a report from the BBC News titled *The growing popularity, and potency, of Ecstasy and MDMA* [18], which emphasised on the growing usage of the Ecstasy drugs. To end with, the study of Global Drug Survey [15] also discussed the raising of the DMT and the Ketamine, the second and third emerging drugs detected through our algorithm.

TABLE II
THE EMERGING DRUGS PRODUCTS IN THE CORE-SHELL FOR A SUBSET OF DUTA DATASET

Shell ID	Core-Shell Products	Candidates for Emergent	Selected Emerging Products
19	Marijuana, DMT, Edibles, Ecstasy, XTC, Cocaine, MDMA, Speed, Heroin, Weed, Mxe, 2C-B,,Haze, Mescaline, GHB, MDA, Benzos, Cannabis, LSD Mushrooms, GBL, Meth, Hashish, Amphetamine, Ketamine, Psychedelics, Kush	Ecstasy, Ketamine, DMT, MDA, Marijuana, Hashish, Mescaline, GHB, Amphetamine,,Psychedelics, Mushrooms, 2C-B, Edibles, GBL, Mxe, Benzos, XTC	Ecstasy,Ketamine, DMT, MDA, Marijuana, Hashish, Mescaline, GHB, XTC

In addition to the quantitative analysis, we present a visualisation of the PCG since it might help in understanding the underlying relationships between the products in the graph. In the graph (better viewed in colour), we coloured the graph's nodes based one their shells, while the nodes' sizes depend on their frequencies in DUTA. The Figure 6 shows the complete PCG graph[10]. This Figure shows clearly that the most relevant products are located in the centre since they have the highest connections with the other nodes, while the lower connectivity nodes are in the graph borders. However, having a small node in the centre indicates that this node tends to grow since it has appeared connected with the popular ones within the same shell.

For further clarification, we took a sub-graph of the PCG that contains the core-shell nodes only as shown in Figure 7, and we highlighted the "Ecstasy", which occupied the first position among the emerging products. This Figure shows 10 favourite products, i.e. popular, and 17 candidates ones. By applying the emergent detection algorithm we found that Ecstasy is offered with all the products contained in the popular list, and the total sum of its edges with popular products was 68, while the *Benzos* drug is provided only with 7 popular and the sum of its weighted edges was 15. Therefore, the Benzos was not proposed as an emerging product. However, the Global Drug Survey [15] survey has reported that the 2C-B and the DMT are both emerging, and

[10]The interactive visualisation is available upon email request to the correspondence author

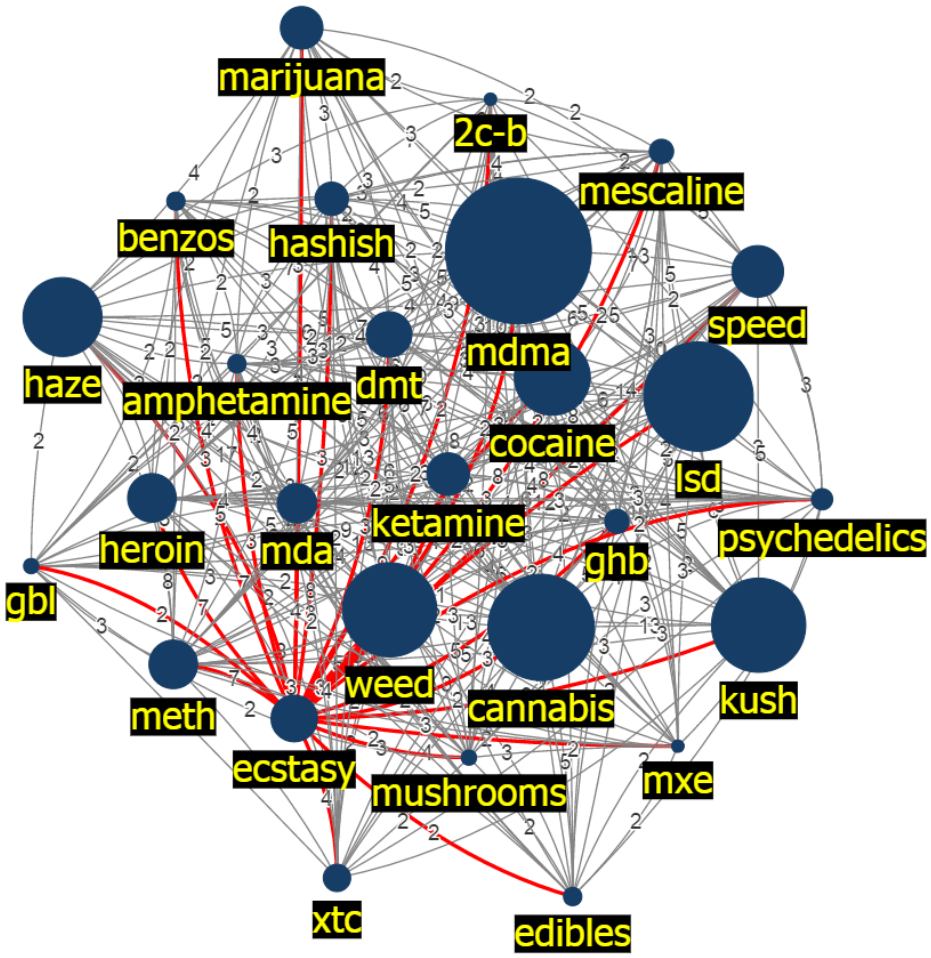

Fig. 7. Sub graph of the PCG that contains only shell 19

our framework has detected the DMT only but ignoring the 2C-B. The reason behind that is related to the difference in the studied sample. In ours, we have the drug 2C-B in the core-shell as a candidate for emerging drugs. But, in this case, it was not separated to the final emerging drugs list because it has only 6 connections with the popular product, a value lower than the threshold proposed. In the case of DMT, the weight 10 is higher than the threshold used to propose the final emerging product list, i.e. the average weight of their edges, so it is nominated as a formal emerging drug.

VI. Conclusions and Future Work

In this paper, we presented a semi-automatic framework for the problem of emerging products detection in the Darknet using graph theory and the k-shell decomposition. After pre-processing a product list and the text resources extracted from Darknet domains, we constructed a Products Correlations Graph (PCG) where the nodes correspond to the Darknet products, and the edges reflect a simultaneous offering of two products. In other words, a new edge is created between two nodes when they have been offered in the same Darknet domain at least once. Then, by running the k-shell algorithm, we decomposed the PCG graph into levels of importance where the core-shell contains the most relevant products. Finally, we applied our emergent detection algorithm into the products contained into the core-shell, and we obtained what the most popular products and a list of emerging products are. Furthermore, from the PCG, we could conclude association rules between the products that are hard to infer when the markets transaction logs are not available.

We conducted our experiments over a subset of DUTA, whose text resources contains the categories of Drugs (both legal and illegal) and Black Marketplaces. The results of our experimentation are validated quantitatively by matching them with reports of prestigious international organisations who are concerned with the drugs spread.

Our results are encouraging to investigate the factors that affect emergent products and assess the proposed approach over more classes of DUTA, like the Violence-Weapons. In future work, we are looking forward to fully automating the framework by using the Natural Language Processing (NLP) techniques in extracting the products names list. To handle the problem of the words that have multiple meanings, we propose the use of some vectorization techniques, e.g. word2vec [20].

Acknowledgement

This research was funded by the framework agreement between the University of León and INCIBE (Spanish Na-

tional Cybersecurity Institute) under addendum 22. We want to thanks to Francisco J. Rodríguez and Antonio Sepúlveda, from INCIBE, for their help and valuable comments.

REFERENCES

[1] Buxton, Julia and Bingham, T. (2015). The Rise and Challenge of Dark Net Drug Markets. Global Drugs Policy Observatory Policy Brief. https://doi.org/2054-1910

[2] Carmi, S., Havlin, S., Kirkpatrick, S., Shavitt, Y., and Shir, E. (2007). A model of Internet topology using k-shell decomposition. Proceedings of the National Academy of Sciences of the United States of America, 104(27), 11150–4. https://doi.org/10.1073/pnas.0701175104

[3] M. AL NABKI, E. Fidalgo, E. Alegre and I. de Paz, "Classifying Illegal Activities on Tor Network Based on Web Textual Contents", European Chapter of the Association for Computational Linguistics, 2017.

[4] Kontostathis, A., Galitsky, L. M., Pottenger, W. M., Roy, S., and Phelps, D. J. (2004). A survey of emerging trend detection in textual data mining. Survey of Text Mining, 185–224. https://doi.org/10.1007/978-1-4757-4305-0_9

[5] Glance, N., Hurst, M., and Tomokiyo, T. (2004). BlogPulse: Automated trend discovery for weblogs. WWW 2004 Workshop on the Weblogging Ecosystem ACM New York, 2004, 1–8

[6] Davidson, G. S., Hendrickson, B., Johnson, D. K., Meyers, C. E., and Wylie, B. N. (1998). Knowledge mining with vxInsight: Discovery through interaction. Journal of Intelligent Information Systems, 11(3), 259–285. https://doi.org/10.1023/A:1008690008856

[7] Swan, R., and Allan, J. (2000). Automatic generation of overview timelines. Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 49–56.

[8] Porter, Alan L., and Michael J. Detampel. "Technology opportunities analysis" Technological Forecasting and Social Change 49.3 (1995): 237-255.

[9] Raeder, T., and Chawla, N. V. (2011). Market basket analysis with networks. Social Network Analysis and Mining, 1(2), 97–113. https://doi.org/10.1007/s13278-010-0003-7

[10] Pastor-Satorras, R. (Romualdo) and Vespignani, A. (2004). Evolution and structure of the Internet : a statistical physics approach. Cambridge University Press.

[11] Miorandi, D., De Pellegrini, F., and De Pellegrini, F. K. (2010). K-Shell Decomposition for Dynamic Complex Networks-Shell Decomposition for Dynamic Complex Net- works. WiOpt'10: Modeling and Optimization in Mobile, 499–507.

[12] Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., and Makse, H. A. (2010). Identification of influential spreaders in complex networks. Nature Physics, 6(11), 888–893. https://doi.org/10.1038/nphys1746

[13] Liu, Y., Tang, M., Zhou, T., and Do, Y. (2015). Improving the accuracy of the k-shell method by removing redundant links-from a perspective of spreading dynamics. arXiv Preprint arXiv:1505.07354, 5(August), 1–11. https://doi.org/10.1038/srep13172

[14] Alvarez-Hamelin, J. I., Dall'Asta, L., Barrat, A., and Vespignani, A. (2005). k-core decomposition: a tool for the visualization of large scale networks.

[15] Globaldrugsurvey.com. (2017). The Global Drug Survey 2016 findings — Global Drug Survey. [online] Available at: https://www.globaldrugsurvey.com/past-findings/the-global-drug-survey-2016-findings/ [Accessed 28 Feb. 2017].

[16] Berry, M. J., and Linoff, G. (2004). Data Mining Techniques: For Marketing, Sales, and Customer Support. Wiley.

[17] Annie, L., Mc, C., and Kumar, A. D. (2012). Market Basket Analysis for a Supermarket based on Frequent Itemset Mining. International Journal of Computer Science Issues(IJCSI), 9(5), 257–264.

[18] BBC News. (2017). The growing popularity, and potency, of ecstasy and MDMA - BBC News. [online] Available at: http://www.bbc.com/news/uk-37156380 [Accessed 28 Feb. 2017].

[19] Emcdda.europa.eu. (2017). EMCDDA home page: www.emcdda.europa.eu. [online] Available at: http://www.emcdda.europa.eu/system/files/publications/2637/TDAT16001ENN.pdf [Accessed 28 Feb. 2017].

[20] Goldberg, Y., and Levy, O. (2014). word2vec Explained: Deriving Mikolov et al.'s Negative-Sampling Word-Embedding Method. arXiv Preprint arXiv:1402.3722, (2), 1–5. https://doi.org/10.1162/jmlr.2003.3.4-5.951