DIVING INTO THE DARKNET: USING TABLEAU TO SHED SOME LIGHT ON DATA ANALYSIS

Ton Don
Georgia Southern University, College of Engineering and Information Technology
P.O. Box 8150, Statesboro, GA 30640
912-478-4121
td02284@georgiasouthern.edu

Hayden Wimmer*
Georgia Southern University, College of Engineering and Information Technology
P.O. Box 8150, Statesboro, GA 30640
912-478-4121
hwimmer@georgiasouthern.edu

Carl M. Rebman, Jr.
University of San Diego, School of Business Administration
5998 Alcala Park, Coronado 212, San Diego, CA 92110
619-260-4135
carlr@sandiego.edu

Scott Scheidt
Georgia Southern University - Armstrong, Center for Applied Cyber Education
P.O. Box 8150, Statesboro, GA 30640
912-478-4121
Scott.Scheidt@armstrong.edu

*Corresponding Author

ABSTRACT

With the law enforcement closure of Silk Road, an increasing number of crypto-markets surfaced at a rapid rate. Taking an interest in this kind of market, multiple studies were conducted using different techniques and tools to collect the data. This data ranged from the kind of drugs being traded, how long the seller was active and the number of transactions a seller conducted. Visualization has always been an effective process that helps the reader, planner, and decision maker to see and recognize the information much faster than reading a table or sorting through pages of information. Tableau is one of the most modern statistical and visualization software packages available. Combining Tableau visualization capabilities and analyzing prior collected data, this study presents findings that illustrate the effectiveness of data visualization.

Keywords—Darknet, Crypto-market; Visualization

INTRODUCTION

The crypto-market, also known as the Darknet market, was a term researcher used for illegal online drug trading markets. Having evolved from the traditional illegal drug trade, these markets focused on making transactions and identities anonymous. By doing so, sellers could constantly put their products onto display without concern for law enforcement. With evolving technology, buyers had increased the ability to participate. This illegal online activity slowly became a problem for which the authorities had no answer (Aldridge & Décary-Hétu, 2014). Following the exposure of Silk Road in 2013 (Department of Justice, 2014), several researchers took an interest in this new kind

of market (Aldridge & Décary-Hétu, 2014; Alois Afilipoaie, 2015; Anonymous, 2014; Branwen, 2015; Buxton & Bingham, 2015; Christin, 2013). Various tools and techniques were used to extract the data in order to transform it into useful information. This process was complex and in-depth. In order to collect the raw data from the sites, HTTrack was employed by Munksgaard, Demant, and Branwen (2016). The sites took days to download and often times were canceled during the extraction process. Once the sites were downloaded, data science methods were used to clean and extract data. Some steps could be accomplished using tools, while others had to be done manually (Christin, 2013). Because of the vast amount of time and effort that was put into collecting and analyzing the data, there were 2 sets of data discovered.

In this work, we adopted the data set from Dr. Aldridge's and Dr. David Décary-Hétu's previous research. Many attributes were not clearly described or documented; therefore, the majority of the obfuscated attributes were omitted, leaving only the meaningful ones. With the revised and reconstructed data set, Tableau, a visualization statistical package, is utilized to present and visualize the Darknet data. Next, we employ some common predictive methods from data science to determine the predictability of the data. The remainder of this paper is structured as follows: section 2 provides a literature review, section 3 details our methods and results, and section 4 concludes this work.

LITERATURE REVIEW

The Darknet

With the explosion of internet usage, the Darknet drug markets emerged and flourished based on newfound channels to deliver illegal substances to consumers. Buxton and Bingham (2015) attempted to discover how Darknet markets operated and how law enforcement interacted with them. With the tech-savvy generation, Darknet markets became an incentive for drug vendors to participate in reaching a new consumer base. Law enforcement strategies and policies were incapable of confronting the Darknet and hidden services (Buxton & Bingham, 2015).

Silk Road, one of the biggest online illegal trading services in the Darknet, was shut down in October 2013. Nicholas Christin and Kyle Soska conducted a study on the Darknet. Using web crawling mechanisms, the authors collected data from many websites. Next, they parsed all text from downloaded sites. All information was imported into a database. While the authors understood that there were limitations to their method, they anticipated that the information and their method could be improved in future works (Soska & Christin, 2015)

A list of drugs on the Silk Road crypto-market was downloaded in September of 2013 by Aldridge and Décary-Hétu (2014). With the high price-quantity sales, drug sales were the main source of revenue for Silk Road. There was an increase of 600% from mid-2012 to September of 2013. Authors found that six drug categories comprised of 90% of the market. Out of all the listings, they collected the sellers' lifespan, the price of the drug, rating of sellers, and transaction count. After calculation and analysis, they determined that the resellers were the ones who paid for the large quantity of drugs or listing of a higher price. Not only that, most of the revenue from the crypto-market was generated by the higher price listing (Aldridge & Décary-Hétu, 2014)

Criminal activities can now be executed through innovative ways with the Darknet market one of the fastest growing criminal opportunities. Rhumorbarbe, Staehli, Broséus, Rossy, and Esseiva (2016) takes all aspects of the Darknet market into consideration, from technologies to physical,

and then chemical. Using the collected data from an independent researcher named Gwern, authors give readers a better overview of the Darknet market. Moreover, the team of researchers sought permission from an Attorney General to buy the drugs for educational purposes. A difference was found in the chemical description of the drugs, which were sold by the seller. Additionally, authors state that the online drug market is also a different way to distribute drugs on a larger scale (Rhumorbarbe et al., 2016).

The crypto-market is a web-based market where legal and illegal services are being traded from sellers to buyers. Aldridge and Décary-Hétu (2016) took an interest in the market and wrote several articles based on the data collected from this market, specifically Silk Road. Using various articles with the same interests in the crypto-market, predictions about how the crypto-market will grow in the future were formed. Considerations emerged including the efficacy of law enforcement on crypto-markets. For example, should the online drug market be the priority or not? Even though, law enforcement showed that they have the upper hand but with the ongoing activities of the market, they might not have much impact (Aldridge & Décary-Hétu, 2016).

Aldridge and Décary-Hétu (2016) briefly walk readers through the history of the crypto-market, with a focus Silk Road. After that, the authors would go through an overview of how the market was operated. For this section, the author would show which basic tools buyers needed to start surfing the crypto-markets; all the way through how the transaction was finalized (Aldridge & Décary-Hétu, 2016).

Christin (2013) used a crawling mechanism to automate the data collection process. Software called HTTrack was used to download all information related to the visited site, including pictures and related link-structures. The targets of the crawling process were item, user, and category web pages. Instead of spending hours manually downloading all the needed information, the researcher would just let the machine run over a period time then come back to check the results. Statistical analysis was performed on the crawled data to extract patterns. First, the types of drugs sold on the market were analyzed. Second, researchers examined the countries evolved in the trading process, both sending and receiving countries. Next, they investigated revenue generated by the website. Finally, they made recommendations how to stop the illegal trading (Christin, 2013).

Research, performed in 2015, about Darknet data was published with a method for the collected data set. In order to validate the method and the data collected, a group of researchers replicated the process. They used the same mechanisms and compared the results. Surprisingly, the data collected by Gwern and his group had a large margin of error compared to the original study. With a question in mind, Gwern concluded his study with a set of suggestions for a future researcher on how to validate their results (Munksgaard et al., 2016).

In November of 2014, once again, the Federal Bureau of Investigation and several European agencies, also known as Operation Onymous, worked together to seize and shut down multiple crypto-markets; Pandora, Blue Sky, Hydra, Cloud Nine, and Silk Road 2.0 (Department of Justice, 2014). After the operation was completed, the Global Drug Policy Observatory also reported that there was an undercover agent working as one of the administrators which contributed to the success of the operation (Alois Afilipoaie, 2015). In the end, the anonymous system, which was designed to protect the crypto-market, was used to destroy it.

Data Visualization

In 2013, Stephen Few, a researcher with more than 20 years of experience as an innovator, consultant, and educator in the field of business intelligence, wrote an article for human

perception using data visualization. In this paper, Few mentioned one of the most influential psychological studies about visualization for human perception Gestalt principles: the law of proximity and the law of similarity (Few, 2013). As shown in figure 1, using the law of proximity, the human brain would see that there were three columns of circles forming a group instead of seeing six separate columns. Figure 2 was used to demonstrate the law of similarity. At first glance, the brain would tell us which objects would be in the same group because of the assigned colors.

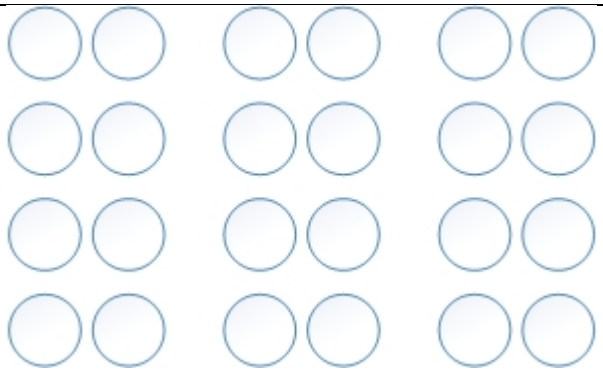


Figure 1: Law of Proximity (Few, 2013)

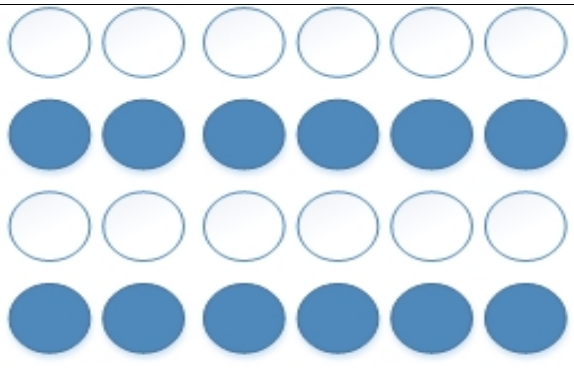


Figure2: Law of Similarity (Few, 2013)

Demonstrated in figure 3, visualization could be expressed as a message. As Andy Kirk described, a message was the channel of communication. The messenger embedded his/her ideas, discoveries, or results into a visual aid. This visual aid could be anything; a chart, an infographic, etc. Using the visual aid created by the messenger, the receiver should be able to process all the information through this message (Kirk, 2012). As the designer of this message, it depended on the field of study or the data which would be presented, the appropriate visual aid would be chosen.

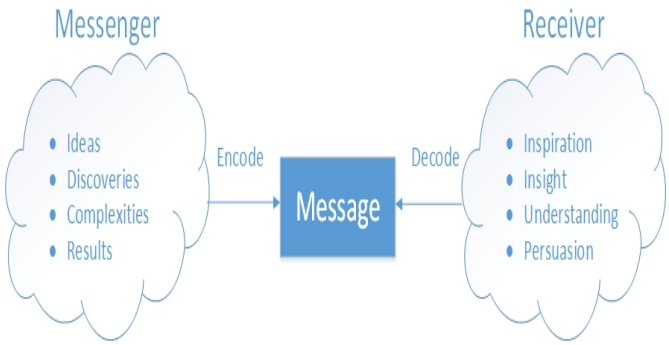


Figure 3: Relationship between messenger and receiver (Kirk, 2012)

Maureen Stone, the founder of Stone Soup Consulting, published a book named "A Field Guide to Digital Color" in 2003. On top of that, as of the beginning of 2017, she has more than 30 published papers and 12 patents on digital color using interface technology and computer graphics. In (Stone, 2006), Maureen Stone explained how to choose the correct color to present your findings. A wrong decision with the color representation could affect the ability to deliver the message. A correct combination, however, will help the viewer understand the roles and the relationships of the elements (Stone, 2006).

Many Eyes and Tableau are emerging as the most popular online data analytic and sharing tools. The researchers wanted to understand the reason why these tools are so popular. Researchers traced and figured out how much Many Eyes and Tableau Public were being used for an extended period. Four dimensions were taken into consideration: (1) types of users which leveraged the system, (2) how users interacting with the published content, (3) how users analyzed a single data set, and (4) how they integrated data sources. Researchers also mentioned that there is room for improvement for web-based visualization analytics systems (Morton, Balazinska, Grossman, Kosara, & Mackinlay, 2014).

METHODS AND RESULTS

From the sample data set used in the method section, we imported the data into Tableau then generated figure 4. Figure 4 illustrates the top rank among all the sellers, from 1 to 10 with 1 as the highest and 10 as the lowest. Out of this group, we can also see the average rating on the top right of figure 1, from 5 to 3.4 with 5 is the highest rating a seller can get from a user and 3.4 is the lowest in this group. Not only that, we use red as the demonstration of rating, with the darker the color means the higher the rating, and the lighter means lower. Furthermore, we also take the life span of these sellers into consideration. Since we already used the color for average rating, we decide to use the size of the circle to express the total life span. The larger the circle, the longer or larger the total life span of the sellers, and vice versa.

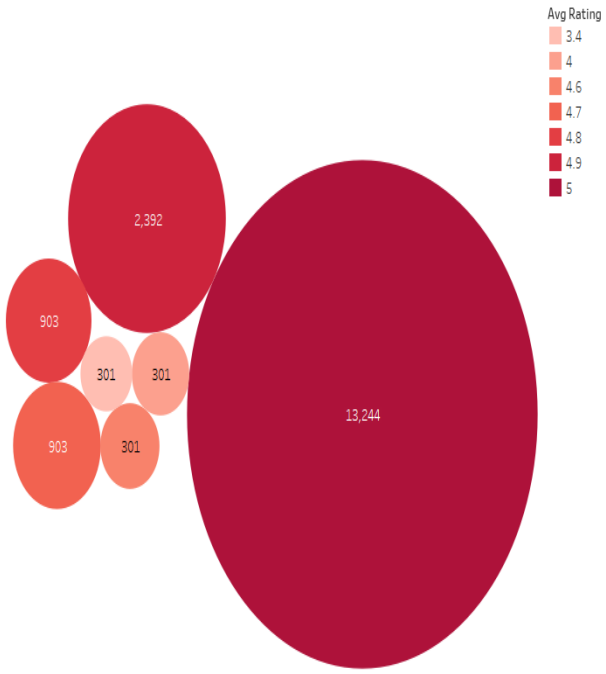


Figure 4: Top 10 sellers based on ranking (Aldridge & Décary-Hétu, 2014)

Using the second data set from Dr. Aldridge and Dr. Décary-Hétu (Aldridge & Décary-Hétu, 2014), we imported the data into Tableau. Using this data, we decided to create two maps. The first map, figure 5, is the countries from where all the trades are coming. The second map, figure

6, would have the countries to where the trades are delivered. In order to create figure 5 and 6, we display all the recorded categories. Not stopping at that, we also use the color scheme to show the origins and the destinations of the products. Each color is linked to a specific drug, by doing this, readers can clearly see the regions or countries which the drugs were coming from.

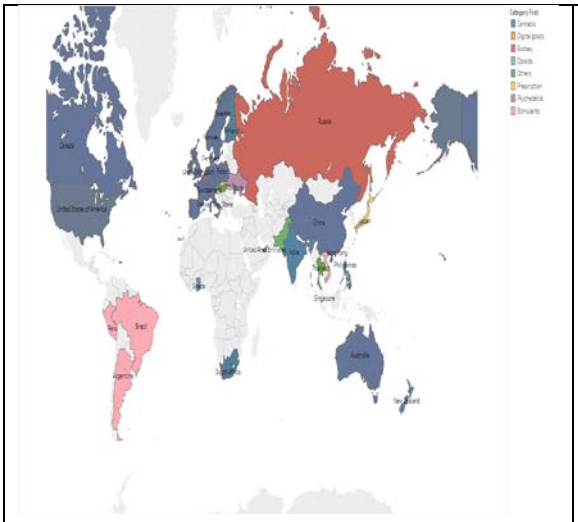


Figure 5: Sellers' countries of origin

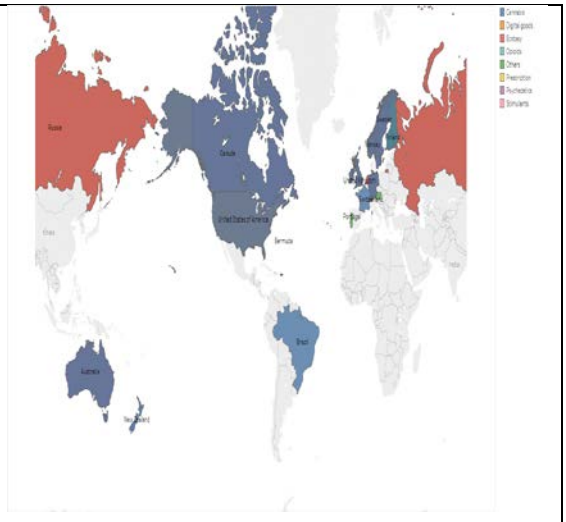


Figure 6: Transactions' destinations

Analyzing the Darknet with Data Science

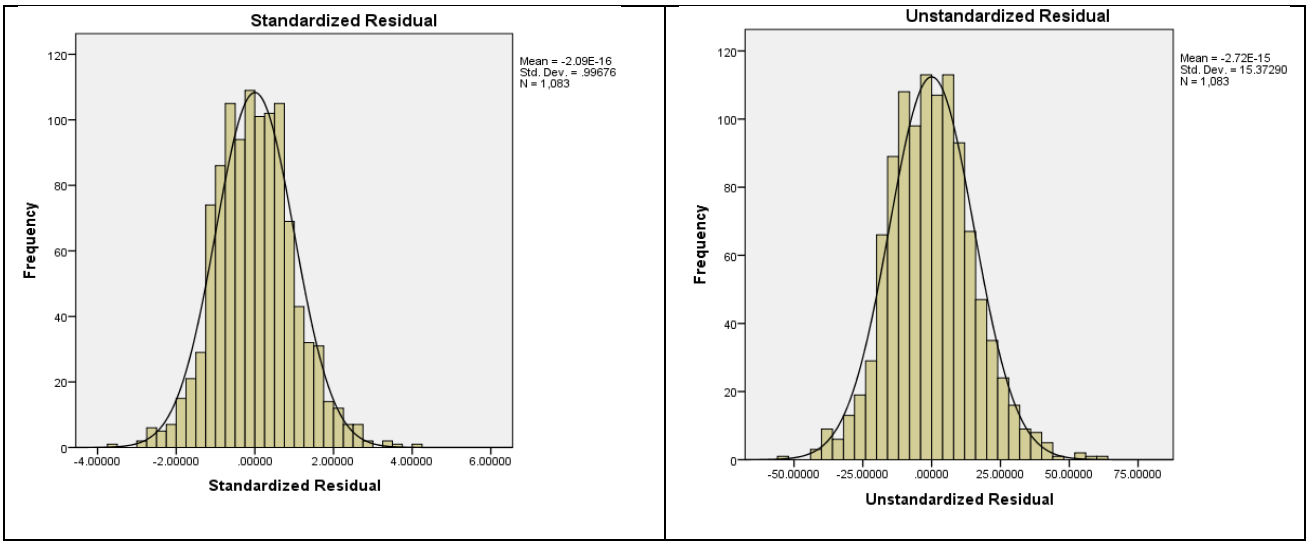
According to EMC, linear regression is a statistical technique used to express a relationship between variables and a continuous outcome value. If linear regression is chosen as a method, it is safe to assume the relationship between the input and the output is linear. Even when the prediction is restricted to a line, there are possibilities which the input can be transformed and modified so that the outcome will be linear (EMC Education Services, 2015).

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{p-1} X_{p-1} + \varepsilon$$

Formula 1: Linear Regression (EMC Education Services, 2015)

With the data collected in Formula 1, we used a linear regression model for predicting the relationship between the attributes within the dataset. For this study, we used to rank as our dependent variable. Since there were some missing data in the rank column, we decided to replace all the null value with the mean, 45.9. For the model summary, we found out that adjusted R-square value was 66%. This R-square result means that 66% of our data set fit into the linear regression model.

An unstandardized and a standardized coefficient test was also conducted. The results indicate that there is a relationship between the rank of the sellers who shipped their products worldwide and the number of transactions that they completed, with p-value is less than 0.001. Not only that, the ranking of the sellers also related to the number of products they sold because the significant value for bulk sellers is p = .049 which is less than p = .05. Residuals were also tested for normality of the data set. The results are illustrated in Figures 7 and 8.



Figures 7 and 8: Residual tests

Using the same data set, a CART decision tree was utilized to find the pattern between the attributes. In order for this test to work with our dependent variable being rank, it was necessary to transform the ranking of the data set. Sellers were assigned "highest" for those whose rank ranged between 76 and 100. Likewise, sellers were assigned "high" for those whose rank ranged from 51 to 75. "Low" is for sellers whose rank is 50 to 26, and the "lowest" starts from 25 to 0. The highest, high, low, and lowest rankings were selected as the dependent variable After the CART decision tree test, the dataset was broken down with seven leaf nodes. The first split is where the transactions were less than 260.5 and greater than or equal to 260.5. Figure 9 details the decision tree produced by the CART algorithm.

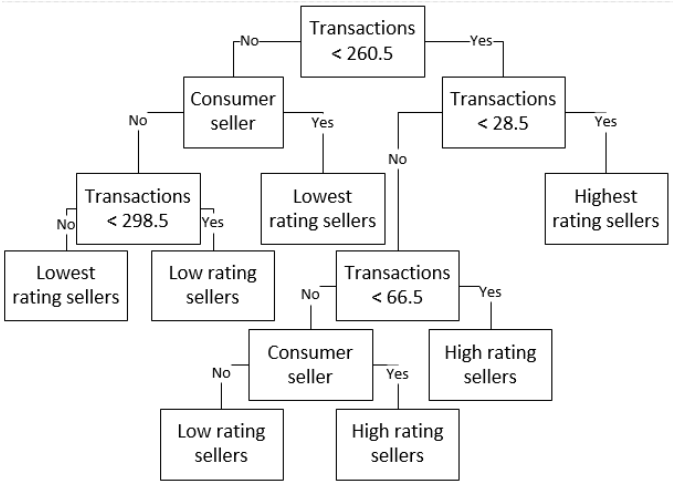


Figure 9: CART Decision Tree

For the group of sellers who has less than 260.5 transactions, the tree breaks down to sellers who have less than 28.5 transactions and sellers who have equal to or more than 28.5. The group of sellers who has less than 28.5 transactions has the highest rating. As for sellers who made more than or equal to 28.5 transactions, the decision tree splits between sellers who completed equal or greater than 66.5 transactions, and sellers who had less than 66.5 transactions will be categorized

as high. As for sellers who had 66.5 transactions or more, they are broken down into high if they are consumer sellers and low if they are not consumer sellers.

As for the group of sellers who recorded with more than or equal to 260.5 transactions, they split between sellers who used their own product and sellers who do not. The ones who do are categorized as lowest rank. This time, the sellers who have less than 298.5 transactions are in low ranking, and the ones who completed equal to or more than 298.5 transactions are in the lowest ranking. There are 1083 instances in this model. While using the CART decision tree, it was found there were 62.8 percent (680 cases) that could correctly classify the ranking of the sellers. However, the model incorrectly specified the other 37.2 percent or 403 cases. The logistic regression test with 10-fold stratified cross-validation in place produced similar results, with 61.5 percent correctly classified and 38.5 incorrectly classified instances.

CONCLUSION

Unlike other studies where their focuses are in criminal justice or terrorism (Alexander, 2011; Eberle, Graves, & Holder, 2010; Roberts, 2012; Van Dongen, 2011), the purpose of our research is to review visualization and demonstrate the application of Tableau to the visualization of Darknet data as well as apply basic data science to Darknet data. We did not seek to validate the information or the data validity from the original study; however, we aim to demonstrate a new method to improve understanding of Darknet data. Many studies were conducted about crypto-markets. Unfortunately, these studies did not employ software such as Tableau to display the results nor did they apply data science and predictive analytics. Throughout this research, we learn (1) how crypto-markets are operating, (2) how the data related to the markets is collected, (3) how visualization can be used, (4) how we can use Tableau to process and display the result and (5) an application of data science for predictions in Darknet markets. The authors hope that this study will help future researchers further advance their results and help readers better understand their findings. As for future research, it is suggested one study R and implementing with Tableau. Since Tableau integrates with R, this will provide even more customization, visualization, and analyzation on the data. If implemented correctly, we believe that this could turn Tableau into a very powerful data analysis and visualization tool for future studies.

REFERENCES

Aldridge, J., & Décary-Hétu, D. (2014). Not an'Ebay for Drugs': The Cryptomarket'Silk Road'as a Paradigm Shifting Criminal Innovation. *Available at SSRN 2436643*.

Aldridge, J., & Décary-Hétu, D. (2016). Cryptomarkets and the future of illicit drug markets. *European Monitoring Centre of Drugs and Drug Addiction (EMCDDA)*.

Alexander, D. C. (2011). Student Projects Involving the Analysis of Web Sites of Extremist and Extremist-Affiliated Groups in the United States. *Journal of Applied Security Research, 6*(2), 184-195.

Alois Afilipoaie, P. S. (2015). Operation Onymous: International law enforcement agencies target the Dark Net in November 2014. *Global Drug Policy Observatory*.

Anonymous. (2014). DarkNet Stats. from https://dnstats.net/

Branwen, G. (2015). 2014 in DNMs: by the numbers. from https://www.reddit.com/r/DarkNetMarkets/comments/2r58vs/2014_in_dnms_by_the_numbers/

Buxton, J., & Bingham, T. (2015). The rise and challenge of dark net drug markets. *Policy Brief, 7*.

Christin, N. (2013). *Traveling the Silk Road: A measurement analysis of a large anonymous online marketplace.* Paper presented at the Proceedings of the 22nd international conference on World Wide Web.

Department of Justice. (2014). Dozens of Online 'Dark Markets' Seized Pursuant to Forfeiture Complaint Filed in Manhattan Federal Court in Conjunction with the Arrest of the Operator of Silk Road 2.0. from https://www.fbi.gov/contact-us/field-offices/newyork/news/press-releases/dozens-of-online-dark-markets-seized-pursuant-to-forfeiture-complaint-filed-in-manhattan-federal-court-in-conjunction-with-the-arrest-of-the-operator-of-silk-road-2.0

Eberle, W., Graves, J., & Holder, L. (2010). Insider threat detection using a graph-based approach. *Journal of Applied Security Research, 6*(1), 32-81.

EMC Education Services. (2015). *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*: John Wiley & Sons.

Few, S. (2013). Data visualization for human perception. *The Encyclopedia of Human-Computer Interaction, 2nd Ed.*

Kirk, A. (2012). *Data Visualization: a successful design process*: Packt Publishing Ltd.

Morton, K., Balazinska, M., Grossman, D., Kosara, R., & Mackinlay, J. (2014). Public Data and Visualizations: How are Many Eyes and Tableau Public Used for Collaborative Analytics? *ACM SIGMOD Record, 43*(2), 17-22.

Munksgaard, R., Demant, J., & Branwen, G. (2016). A replication and methodological critique of the study "Evaluating drug trafficking on the Tor Network". *International Journal of Drug Policy*.

Rhumorbarbe, D., Staehli, L., Broséus, J., Rossy, Q., & Esseiva, P. (2016). Buying drugs on a Darknet market: A better deal? Studying the online illicit drug market through the analysis of digital, physical and chemical data. *Forensic science international, 267*, 173-182.

Roberts, C. (2012). The US Federal Protective Service: A troubled agency—The need for improved contract guard training and oversight. *Journal of Applied Security Research, 7*(4), 478-488.

Soska, K., & Christin, N. (2015). *Measuring the longitudinal evolution of the online anonymous marketplace ecosystem.* Paper presented at the 24th USENIX Security Symposium (USENIX Security 15).

Stone, M. (2006). Choosing colors for data visualization. *Business Intelligence Network, 2*.

Van Dongen, T. W. (2011). Break it down: An alternative approach to measuring effectiveness in counterterrorism. *Journal of Applied Security Research, 6*(3), 357-371.