2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

On Augmented Identifying Codes for Monitoring Drug Trafficking Organizations

Kaustav Basu
NetXT Lab, SCIDSE
Arizona State University
Tempe, USA
kaustav.basu@asu.edu

Arunabha Sen
NetXT Lab, SCIDSE
Arizona State University
Tempe, USA
asen@asu.edu

***Abstract*—A staggering 450,000 people died due to drug consumption in 2015, out of which, a third of the deaths were a direct result of drug overdosing. Illicit manufacturing of Cocaine, Heroin, Cannabis, etc., by Drug Trafficking Organizations (DTOs), all peaked recently, which is a major indication of their worldwide demand. With drug offenses increasing globally, the list of suspect individuals, associated with drug trafficking organizations, has also been growing over the past few decades. As it takes significant amount of technical and human resources to monitor a suspect, an increasing list entails greater resource requirements on the part of law enforcement agencies. Soon, monitoring all the suspects on the list becomes an impossible task. In this paper, we present a novel methodology called Augmented Identifying Codes (AIC), an extension of the mathematical notion of Identifying Codes. We show that our method requires significantly lesser resources, on the part of the law enforcement agencies, when compared to strategies adopting standard network centrality measures, for monitoring of individuals associated with drug trafficking organizations. Finally, we evaluate the efficacy of our approach on real world datasets.**

***Index Terms*—Drug Trafficking Organizations, Augmented Identifying Code, Surveillance**

I. Introduction

About 450,000 people died as a consequence of drug use in 2015, according to the WHO. About 168,000 of these deaths were directly associated with drug overdoses. These *recorded* numbers, in a way, are indicative of the ever growing demand for illicit drug consumption, and various Drug Trafficking Organizations (DTOs) are responding by increasing the production rates of their respective drugs. To counter the activities of DTOs, law enforcement agencies are spending significant sums of money to just monitor individuals associated with DTOs. The US alone spends about $40 billion a year investigating drug offenses [2]. The investigation of drug offenses usually generates a list of suspects. With drug offenses increasing by the passing year, the suspect database also continues to grow. As a result, the amount of technical and human resources required to monitor a suspect in the database also grows. After a certain point of time, monitoring all the suspects in the database may become an impossible task.

Over the past couple of decades, law enforcement agencies have managed to bust numerous DTOs, by deploying various surveillance strategies such as wiretaps, cameras, GPS trackers, etc. [2]. Authorities analyzed transcripts obtained from the interrogation of suspect individuals, to identify key actors and their relationships, to create a social network [2]–[4]. These constructed social networks have been extensively studied by researchers to realize a particular objective. In this paper, we propose a novel Augmented Identifying Code (AIC) approach, an extension of the mathematical concept of Identifying Codes. In a previous study [18], we have shown the effectiveness of Identifying Codes in uniquely monitoring individuals in terrorist networks. Following the Identifying Code strategy [18], law enforcement agencies had the *complete freedom* to deploy agents to ensure that *all the individuals in the network were uniquely monitored*, in a sense, a clean slate approach. The AIC strategy restricts this freedom in the sense that, there are already a certain number of agents deployed to monitor certain individuals, but additional agents are required to ensure *unique monitoring* of *all* individuals in the network. This is in some sense a more *realistic* version, as law enforcement agencies have already deployed agents in the field, to monitor individuals in DTOs. As in the case of Identifying Codes, our approach relies on the following assumption: when an individual in a DTO becomes “active” in drug related activities, his/her friends/associates will have some knowledge of the individual’s plan. Accordingly, even if the individual is not under direct surveillance by the law enforcement authorities (recording phone calls, movement, social interactions with other individuals), but the individual’s friends/associates are, then *the individual involved in drug related activities can be uniquely identified*. More importantly, we show that monitoring individuals using standard centrality metrics - *(i) does not necessarily guarantee unique identification of every individual, and (ii) leads to a wastage of resources on the part of the authorities.* To this end, the key contributions of the paper can be summarized as follows:

- We extend on the concept of Identifying Codes to present a novel problem called Augmented Identifying Codes

ASONAM ’19, August 27-30, 2019, Vancouver, Canada
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-6868-1/19/08/$15.00
http://dx.doi.org/10.1145/3341161.3343530

(AIC), to uniquely monitor suspect individuals in DTOs, and reduce law enforcement requirements.
- We show that utilizing standard social network centrality metrics as a monitoring strategy leads to a wastage in resources, on the part of the authorities.

II. Related Work

In the past few years, significant research on DTO networks have been conducted utilizing Social Network Analysis (SNA). Natarajan in [3], analyzed wiretap data to create an organizational structure and studied the roles of particular individuals. She analyzed over 2000 wiretap conversations and performed SNA of phone contacts, to reveal a large and loosely structured group of 294 individuals in [4]. Bright and Delaney in [5], utilized SNA to study the evolution of a drug trafficking network, based in Australia. They observed changes in centrality scores and the roles performed by particular individuals. Bright *et. al.* in [6] utilized individual attributes coupled with centrality measures to identify key actors in a drug trafficking network. Bright *et. al.* in [7] analyzed judges' sentencing comments to create a network of individuals involved in the distribution of methamphetamine in Australia during the 1990s. Heber in [8], studied the network of drug offenders in Sweden to analyze the types of criminal activity they were involved in. Hughes in [1] analyzed court transcripts and identified key actors to create a social network. They analyzed this social network to explore product diversification in three drug syndicates in Australia.

In addition to drug organizational structure research through SNA, the last few years have seen a significant amount of research on Identifying Codes and its applications to networks. Karpovsky *et. al.* introduced the concept of Identifying Codes in [9] and provided results for Identifying Codes for graphs with specific topologies, such as binary cubes and trees. Using Identifying Codes, Laifenfeld *et. al.* studied joint monitoring and routing in wireless sensor networks in [10], [11]. Charon *et. al.* in [12], [13], studied complexity issues related to computation of minimum Identifying Codes for graphs and showed that in several types of graphs, the problem is NP-hard. Ray *et. al.* in [14] generalized the concept of Identifying Codes, to incorporate robustness properties to deal with faults in sensor networks, and presented a *minimal* algorithm for the computation of Identifying Codes.

SNA studies have usually utilized network centrality measures to identify key actors in the network. In this paper, we show that if these metrics were used to deploy agents for unique monitoring of individuals, then the law enforcement agencies would end up wasting their resources. In our previous efforts, we have utilized Identifying Codes for, (i) monitoring the health of critical infrastructures and civilian structures [15], [19], (ii) monitoring the activities of terrorists [16], [18], (iii) propagation of misinformation on social networks [20] and, (iv) monitoring regions on the surface of the earth [17]. To the best of our knowledge, this is the first study to analyze the social network of a DTO and apply *Augmented Identifying Codes* to uniquely identify individuals in a DTO network.

III. Augmented Identifying Codes

For the ease of understanding, we first define the Identifying Code problem. Next, we present the Augmented Identifying Code (AIC) problem, for undirected and directed graphs.

Definition III.1. Given an undirected graph $G = (V, E)$, the subset $V' \subseteq V$, is defined as an Identifying Code Set (ICS) for the vertex set V, if $\forall v \in V, N[v] \cap V'$ is unique, where, $N[v] = v \cup N(v)$ and $N(v)$ represents the set of nodes adjacent to v in $G = (V, E)$. The Minimum Identifying Code Set (MICS) problem is to find the ICS of smallest cardinality.

By simply replacing $N[v]$ with $N^{out}[v]$, where $N^{out}(v)$ denotes the out-neighborhood of v, one can obtain the Identifying Code definition for directed graphs. The necessary and sufficient condition for Identifying Codes to exist, is that $\forall u \neq v \in V, N[u] \neq N[v]$, for undirected graphs and $(N^{out}(u) \neq N^{out}(v))$, for directed graphs. If the closed neighborhoods of u, v were to be the same, then the graph is said to be containing "twins". In other words, the necessary and sufficient condition for a graph (network) to have an Identifying Code, is that the network be "twin-free".

It may be noted that the MICS problems assumed that *no individuals (or nodes)* in the network were being initially monitored. We now propose an enhanced version of the MICS problem called the Augmented Identifying Code (AIC) problem, which accounts for the presence of law enforcement authorities currently monitoring certain individuals in the network (denoted by Initial Set). We formally define the Augmented Identifying Code problem for the two types of graphs as follows:

Definition III.2. Given an undirected graph $G = (V, E)$ and a subset $V_1 \subseteq V$ (Initial Set), determine the smallest subset $V_2 \subseteq V$, such that, $\forall v \in V$, $N[v] \cap V_3$ is unique, where $N[v] = v \cup N(v)$, $N(v)$ denotes the neighborhood of v and $V_3 = V_1 \cup V_2$.

Definition III.3. Given a directed graph $G = (V, E)$ and a subset $V_1 \subseteq V$ (Initial Set), determine the smallest subset $V_2 \subseteq V$, such that, $\forall v \in V$, $N^{out}[v] \cap V_3$ is unique, where $N^{out}[v] = v \cup N^{out}(v)$, $N^{out}(v)$ denotes the out-neighborhood of v and $V_3 = V_1 \cup V_2$.

For undirected and directed graphs, the general objective of the AIC problem is to determine the smallest set of individuals to monitor, V_2, *in addition to the individuals currently being monitored*, V_1, in order to ensure that each node in V can be uniquely monitored. Finally, it may be noted that the Initial Set does not necessarily guarantee unique identification of nodes.

Due to lack of space, we are not able to present how the MICS can be computed as a novel variation of the traditional Graph Coloring problem, called the Graph Coloring with Seepage (GCS) problem. This variation has been presented in detail in [15]–[20].

IV. Problem Solution

In this section, we provide a solution technique for the undirected DTO networks, utilizing an Integer Linear Program (ILPs). The ILP for directed DTO networks is almost identical to that of the undirected DTO.

A. Augmented Identifying Code

Instance: $G = (V, E)$, an undirected graph and a set $V_1 \subseteq V$, the set of nodes currently being monitored.
Problem: Find the smallest subset $V_2 \subseteq V$, such that $N[v] \cap V_3$ is unique for each node $v \in V$ (each node receives a unique color, either atomic or composite, through seepage), and $V_3 = V_1 \cup V_2$.
We use the notation $N[v_i]$ to denote the closed neighborhood of v_i, $\forall v_i \in V$. Corresponding to each $v_i \in V$, we use an indicator variable x_i,

$$x_i = \begin{cases} 1, & \text{if a color is injected at node } v_i, \\ 0, & \text{otherwise} \end{cases}$$

Objective Function: Minimize $\sum_{v_i \in V} x_i$

Coloring Constraint: $\sum_{v_i \in N[v_j]} x_i \geq 1, \forall v_j \in V$

Unique Coloring Constraint:
$\sum_{v_i \in \{N[v_j] \bigoplus N[v_k]\}} x_i \geq 1, \forall v_j \neq v_k, \in V$

Initial Set Constraint: $x_i = 1, \forall v_i \in V_1$

V. Experimental Results

To highlight the effectiveness of our algorithms in the reduction of resources without compromising the ability of unique identification of a suspect, we executed the ILPs on various real world drug network datasets [21]. It may be recalled that for a network to have an Identifying Code, it must be “twin-free”. In other words, the necessary and sufficient condition for a network to have an MICS is that the network be “twin-free”. For “twin-free” networks $G = (V, E)$, one trivial Identifying Code set solution is the node set V, although, V may not be the Identifying Code set of minimum cardinality. This implies that, if monitors were placed on every node in the network, then all the nodes in the network would receive a unique identification. However, our algorithms show that unique identification for all the nodes in the network can be obtained by deploying agents to monitor a subset $V' \subseteq V$. Since no additional benefits are realized by deploying additional agents, there is absolutely no reason to deploy a larger number of monitors. In Table I, we highlight on the reduction in resource requirements brought about by our methods. As monitoring suspect individuals in drug networks can be a costly affair on the part of law enforcement authorities [2], a significant reduction in resources will be of great interest to the respective authorities. Furthermore, we also highlight the amount of resource wasted when standard network centrality metrics are utilized for monitoring individuals in a network.

A. Datasets

The datasets used for evaluating the efficacy of our approach were obtained from [21]. In each dataset, the nodes represents individuals and the edges represent the relationship between these individuals. Natarajan in [3] analyzed wiretaps and created a social network of Cocaine dealers in New York City. The second dataset, Cocaine Smuggling, contains four social networks unearthed by four investigative operations carried out by law enforcement agencies, involved in smuggling Cocaine from Colombia to Spain. Operation Mambo identified 31 suspect individuals based in Colombia, Operation Juanes identified 51 suspect individuals based in Mexico and Operations Jake and Acero identified 38 and 25 suspect individuals respectively, based in Madrid. The third dataset, Drug Net is a social network of 294 drug users in Hartford. Natarajan [4] uncovered a Heroin trafficking organization based in New York City consisting of 38 suspect individuals, which is the fourth dataset. The fifth dataset is the Montreal Street Gang dataset, reconstructed from the drug-distribution operations in Montreal North and contains 35 organizations.

B. Augmented Identifying Codes

For simulation purposes, we assumed that the Initial Set of nodes being currently monitored are the $k-$nodes with the highest degree, betweenness and eigen vector centrality scores. For each drug network, we present the number of nodes in the network, the MICS cardinality (utilizing the Identifying Code ILP defined in [15]–[20]) of the network and the cardinality of the AIC set, if $k-$nodes were currently monitored (where k is varied from 25%, 50%, 75% and 100% of the total nodes in the network), for each centrality metric. It may be noted that some nodes in Initial Set may have equal centrality scores with the nodes not present in the Initial Set. For example, consider the following centrality score sequence of 6 nodes, ordered descendingly: 0.88, 0.88, 0.88, 0.85, 0.77, 0.77. The top 25% of these nodes would comprise of nodes 1 and 2 which have 0.88 as their centrality scores. It can be seen that node 3 also has a score of 0.88. The AIC solution cardinality obtained by initially monitoring nodes 1 and 2 may be different from the AIC solution cardinalities obtained by monitoring nodes 1 and 3, and nodes 2 and 3 respectively. As a result, all such combinations should be considered and the AIC solution cardinality should be averaged over the number of combinations considered. It should be noted that for the $k = 100\%$ scenario, we have assumed that authorities can deploy *at most* n number of agents, for the unique monitoring of n individuals in the network. The Operation Juanes drug network has 50 nodes in the network and the MICS cardinality of such a network is 22. For $k = 25\%$ of degree centrality, we considered the top 25% nodes with the highest degree centrality score which turns out to be 12. Furthermore, an additional 15 nodes have to be monitored to ensure that every node in the network has a unique identification, resulting in an AIC cardinality of 27. As all 12 nodes in the Initial Set had unique degree centrality scores when compared to the nodes not in the Initial Set, only 15 additional individuals have to

TABLE I

AUGMENTED MICS CARDINALITIES FOR DRUG NETWORKS CORRESPONDING TO VARIOUS CENTRALITY MEASURES

Network	Num. Nodes	MICS	Centrality Measure											
			Degree				Betweenness				Eigen Vector			
			$k = 25\%$	$k = 50\%$	$k = 75\%$	$k = 100\%$	$k = 25\%$	$k = 50\%$	$k = 75\%$	$k = 100\%$	$k = 25\%$	$k = 50\%$	$k = 75\%$	$k = 100\%$
Op. Juanes	50	22	12 + 15	25 + 9.5	38 + 8.77	45 + 0	12 + 15	25 + 9.5	38 + 8.77	45 + 0	12 + 15	25 + 9.5	38 + 8.77	45 + 0
Op. Acero	25	13	6 + 9	12 + 6	18 + 5	23 + 0	6 + 9	12 + 6	18 + 5	23 + 0	6 + 9	12 + 4	18 + 5	23 + 0
Op. Mambo	30	16	7 + 11	15 + 7.5	22 + 6.33	29 + 0	7 + 11	15 + 7.5	22 + 6.33	29 + 0	7 + 11	15 + 7.5	22 + 6.33	29 + 0
H. Dealing	37	15	9 + 10	18 + 8.6	27 + 4.5	36 + 0	9 + 10	18 + 8.6	27 + 4.5	36 + 0	9 + 10	18 + 8.6	27 + 4.5	36 + 0
M. Gangs	35	16	9 + 9	18 + 8.8	27 + 6.28	35 + 0	9 + 9	18 + 8.8	27 + 6.28	35 + 0	9 + 9	18 + 8.8	27 + 6.28	35 + 0
C. Dealers	28	23	7 + 20	14 + 14	21 + 7	25 + 0	7 + 20	14 + 14	21 + 7	25 + 0	7 + 20	14 + 14	21 + 7	25 + 0
Op. Jake	38	29	9 + 22	20 + 15.16	29 + 7.2	36 + 0	9 + 22	20 + 15.16	29 + 7.2	36 + 0	9 + 22	20 + 15.16	29 + 7.2	36 + 0
Drugnet	281	207	70 + 162.48	140 + 131.76	210 + 71	281 + 0	70 + 162.47	140 + 131.76	210 + 71	281 + 0	70 + 162.47	140 + 131.76	210 + 71	281 + 0

be monitored for unique identification of all the nodes in the network. For $k = 50\%$ we observe that some of the 25 nodes in the Initial Set had the same degree centrality with nodes not in the Initial Set. As described previously, all such combinations were generated and the average cardinality of the additional set of monitors required turns out to be 9.5, resulting in the average AIC cardinality of 34.5. Similarly, the cadinality of the additional set of monitors for $k = 75\%$ turns out to be 8.77, resulting in the average AIC cardinality of 46.77. For $k = 100\%$, we observe that the law enforcement agencies will need to monitor 45 individuals with the highest centrality scores, to ensure that each node in the network has unique identification.The results for other centrality metrics of other drug networks are shown in Table I.

It can be observed that the AIC cardinality is greater than the MICS cardinality for all the drug networks. Ideally, in the Operation Juanes network, for $k = 25\%, 50\%, 75\%$ and 100%, *if the nodes in the Initial Set were a subset of the MICS*, then the authorities would only need to monitor an additional 10, 0, 0 and 0 individuals respectively, to get unique identifications for all the nodes in the network. The fact that the authorities require an additional 15, 9.5 and 8.77 individuals (on average), indicate that there was some sub-optimal initial monitoring of individuals (Initial Set), essentially leading to a wastage in resources. Only for the $k = 100\%$ scenario (when they have complete freedom of deploying agents) do the agencies not require any additional resources. Such a scenario corresponds to the MICS strategy, where we have shown that far lesser resources are required for unique monitoring. Finally, these results are counter-intuitive in the sense that the Initial Set consists of the top k important individuals. In other words, if the law enforcement authorities were to monitor *only* important individuals, then they would require additional resources, for unique monitoring of the *entire* network of individuals. Our approach provides an optimal balance between monitoring important as well as unimportant individuals.

VI. CONCLUSION

In this paper, we presented Augmented Identifying Codes, which results in a significant reduction in resource requirements on the part of the authorities. More importantly, we conducted extensive experimentation to show that utilizing standard SNA centrality metrics for monitoring DTOs, leads to wastage in resources, on the part of the authorities. Our experimentation also demonstrates how our approach reduces resource requirements on the part of the authorities.

REFERENCES

[1] C. E. Hughes, D. A. Bright, and J. Chalmers. “Social network analysis of Australian poly-drug trafficking networks: How do drug traffickers manage multiple illicit drugs?” Social Networks 51 (2017): 135-147.

[2] “How Drug Dealers Get Caught” [Online]. Available: https://www.robertkinglawfirm.com/blog/2013/november/how-drug-dealers-get-caught/

[3] M. Natarajan, “Understanding the structure of a drug trafficking organization: a conv. analysis.” Crime Prev. Studies 11 (2000): 273-298.

[4] M. Natarajan, “Understanding the structure of a large heroin distribution network: A quantitative analysis of qualitative data.” Journal of Quantitative Criminology 22.2 (2006): 171-192.

[5] D. A. Bright, and J. J. Delaney. “Evolution of a drug trafficking network: Mapping changes in network structure and function across time.” Global Crime 14.2-3 (2013): 238-260.

[6] D. A. Bright, et al. “The use of actor-level attributes and centrality measures to identify key actors: A case study of an Australian drug trafficking network.” J. of Contemp. Criminal Justice (2015): 262-278.

[7] D. A. Bright, C. E. Hughes, and J. Chalmers. “Illuminating dark networks: A social network analysis of an Australian drug trafficking syndicate.” Crime, law and social change 57.2 (2012): 151-176.

[8] A. Heber, “The networks of drug offenders.” Trends in Organized Crime 12.1 (2009): 1-20.

[9] M. G. Karpovsky, K. Chakrabarty, and L. B. Levitin. “On a new class of codes for identifying vertices in graphs.” IEEE Transactions on Information Theory 44.2 (1998): 599-611.

[10] M. Laifenfeld, and A. Trachtenberg. “Identifying codes and covering problems.” IEEE Trans. on Information Theory 54.9 (2008): 3929-3950.

[11] M. Laifenfeld et al. “Joint monitoring and routing in wireless sensor networks using robust identifying codes.” Mobile Networks and Applications 14.4 (2009): 415-432.

[12] I. Charon, O. Hudry, and A. Lobstein. “Identifying and locating-dominating codes: NP-completeness results for directed graphs.” IEEE Transactions on Information Theory 48.8 (2002): 2192-2200.

[13] I. Charon, O. Hudry, and A. Lobstein. “Minimizing the size of an identifying or locating-dominating code in a graph is NP-hard.” Theoretical Computer Science 290.3 (2003): 2109-2120.

[14] S. Ray et al. “Robust location detection in emergency sensor networks.” IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No. 03CH37428). Vol. 2. IEEE, 2003.

[15] K. Basu et al. “Health Monitoring of Critical Power System Equipments Using Identifying Codes.” International Conference on Critical Information Infrastructures Security. Springer, Cham, 2018.

[16] A. Sen. et al. “Terrorist Network Monitoring with Identifying Code.” International Conference on SBP-BRiMS. Springer, Cham, 2018.

[17] A. Sen et al. ”On upper and lower bounds of identifying code set for soccer ball graph with application to satellite deployment.” In Proceedings of the 20th ICDCN, pp. 307-316. ACM, 2019.

[18] K. Basu et al. “A Novel Graph Analytic Approach to Monitor Terrorist Networks.” IEEE SocialCom, 2018.

[19] K. Basu et al. ”Sensor Networks for Structural Health Monitoring of Critical Infrastructures Using Identifying Codes.” In 2019 15th IEEE (DRCN), pp. 43-50., 2019.

[20] K. Basu. “Identification of the Source(s) of Misinformation Propagation Utilizing Identifying Codes”. In Companion Proceedings of The 2019 World Wide Web Conference (WWW ’19), Ling Liu and Ryen White (Eds.). ACM, New York, NY, USA, 7-11.

[21] “Covert Networks” [Online]. Available: https://sites.google.com/site/ucinetsoftware/datasets/covert-networks