2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Community Detection with Edge Augmentation in Criminal Networks

Ashwin Bahulkar*, Boleslaw K. Szymanski*, Nail O. Baycik‡ and Thomas C. Sharkey†
*Computer Science
Rensselaer Polytechnic Institute, 110 8th Street, Troy NY 12180, USA
Email: bahula@rpi.edu, szymab@rpi.edu
†Industrial and Systems Engineering
Rensselaer Polytechnic Institute, 110 8th Street, Troy NY 12180, USA
Email: sharkt@rpi.edu
‡Harry F. Byrd, Jr. School of Business
Shenandoah University, Winchester, VA
Email: nobaycik@gmail.com

***Abstract*—We study community detection in criminal networks and address the problem caused by intentionally hidden edges which hinder the performance of community detection. We make use of link prediction to demonstrate how the community structure of a network can be better identified by augmenting it with edges. We demonstrate the value of this method by showing this method delivers us better quality communities for real life drug trafficking networks. We discuss also the limitations of the approach, and importance of community detection for investigating of criminal networks.**

***Index Terms*—criminal networks, link prediction, community detection, network restoration, hidden networks**

I. Introduction

We study community detection in criminal networks. We find that the performance of such detection in these networks is hindered by edges intentionally hidden by members in the network. This is done to minimize the exposure of the network to law enforcement analysts. We propose a method of network augmentation which improves the performance of community detection in criminal networks. The edges we typically observe in such networks are the ones which can be relatively easily tracked, such as phone call, text messages, or e-mails. Yet, there are possibly several more types of edges among the nodes in the network which can be relatively easy to hide, such as private meetings, calls using single use cellphones, encrypted emails etc. Hence the nodes in the criminal networks might collaborate with each other using undetectable means to carry on their covert activities. We propose a method which tries to identify these hidden edges, adds them to the network, and then performs community detection on the augmented network. We verify the validity of this approach by looking at the changes in the communities observed and checking whether the edges added to the network occur at some future point of time in the network, as it evolves. In order to further test the effectiveness of the approach, we randomly delete edges from the network and try to restore these edges by adding links recommended by link prediction methods.

IEEE/ACM ASONAM 2018, August 28-31, 2018, Barcelona, Spain
978-1-5386-6051-5/18/$31.00 © 2018 IEEE

We test this method over a drug trafficking network and find that there is an improvement in the quality of communities detected, when we compare our findings with partially available ground truth. We also find that a large percentage of the edges recommended by the link prediction occur in the network later in time in its evolution. Finally, we find that link prediction can restore the network to a fair extent when this network is altered by removal of certain percentage of edges intentionally.

The identification of the communities within criminal networks is important in terms of disrupting their operations. Once communities are identified and their roles obtained through investigations, law enforcement can begin to assess the most effective strategies to disrupt the criminal organization. It may be that disruptions to one community of the organization can cause cascading disruptions to other communities and, therefore, law enforcement can benefit from understanding the inter-relationships between the communities and the operations of the organization. For example, Baycik et al. 2018 [2] have examined the cascading impacts to the trafficking network of an organization should law enforcement be able to successfully disrupt the information flow within the network. In order to examine such an impact, law enforcement should know the community of traffickers and how they receive information from the bosses of the criminal organization, both of which can be obtained through the type of analysis presented in this paper.

II. Data

We use the Caviar dataset and the datasets obtained from the Ndrangheta gang investigations [16] to evaluate our method. The Caviar dataset is obtained from investigations and legal documents collected from the Caviar gang investigation which took place in Montreal, Canada over the years 1994-1996. The network of the drug traffickers is created based on the calls made between the gang members. During the Caviar gang investigation, the law enforcement regularly confiscated drug shipments over a period of two years without making arrests until the end of investigation. After each seizure, a network snapshot was created resulting in 11 time snapshots spread over nearly two years. In response to every seizure of drugs the network is continuously evolving from one snapshot to another. Sizes of all network snapshots are listed in Table I. We know the number of calls exchanged between every pair of nodes. We have also limited knowledge about the activities of the different nodes in the network gathered from law enforcement reports during the investigation and from the court records of the subsequent gang prosecution. Importantly, we know who the three lead traffickers in the network are namely N1, N12 and N3 [15]. Moreover, we know that N1 was leading the drug trafficking for hashish, N12 was leading the trafficking of cocaine and N3 was the intermediary between the two with links to non-traffickers as well [15]. In addition to this, we know that two nodes, N76 and N87, dealt with transport and financial investors, respectively. Although we do not know the precise communities in this network, we do believe that each of the prominent nodes mentioned previously lead their own communities.

We also study the two networks obtained from the Ndrangheta, each is obtained from two different investigations. Both networks in this dataset are derived from the phone calls between members of two different drug trafficking based in Italy. One is called Stupor Mundi and the other is known as Chalonero. An interesting feature of both networks is that the precise role of the node is known for most of them. The following roles are identified and assigned to the nodes: a trafficker, a supplier, a buyer, or a courier. We have information of the calls exchanged between every node. For every call we know who the sender and receiver of the call is, and what is the timing of the call. The calls are grouped into multiple time-based snapshots. We obtain two snapshots of six months each for the Chalonero network, and four snapshots of four months each for the Stupor Mundi network. However, like the Caviar dataset, we do not know about the precise communities which exist in the network. An important feature of these networks is that they are structurally more hierarchical than the Caviar network.

TABLE I
AN OVERVIEW OF THE SNAPSHOTS IN THE CAVIAR NETWORK.

Seizure no.	No. nodes	No. edges	Avg degree	Calls per node
1	15	26	2.4	3.9
2	24	35	2.3	5.0
3	33	68	3.4	8.5
4	33	65	2.9	13.8
5	32	50	2.4	5.0
6	27	68	3.4	24.7
7	33	60	3.4	9.6
8	42	75	2.7	7.9
9	34	60	2.6	9.3
10	42	68	2.4	11.4
11	41	72	2.4	8.6

III. RELATED WORK

There are many papers on link predictions, and in the following we discuss those which are most relevant to our paper. In paper [17], the authors attempt to find missing edges to enhance community detection. They first classify the reasons for missing edges, since, unlike in our paper, they consider also edges missing because of imperfect edge detection, errors in recording the edge, and so on. To verify the approach, the authors use a partial network obtained by deleting a certain fraction of edges from the input dataset. The quality of the resulting communities is compared to the ground truth which is assumed to be available. In the subsequent paper, [18], the same authors aim at using community detection to guide the addition of missing edges. In their approach, intra-community edges suggested by link prediction methods are added to the network first, followed by inter-community edges. This would not necessarily work for criminal networks, because the quality of communities detected without edge augmentation is generally poor. The authors verify the approach experimentally using synthetic networks and six very small real-world networks using several link predictors and two community detection algorithms. A more detailed analysis of the relation between community structure and link formation is provided in [11]. Given an array of communities, the density of links inside a community and between any two communities determines the probability of adding a particular link. A further development of this approach is presented by the same lead author in [12]. It uses the local structural information about the network for improved performance. These approaches have been tested on social networks, like the Karate network, for example. The methods which work on social networks may not necessarily work on criminal networks due to the structural differences between these two types of networks. Due to the low quality of communities detected in criminal networks without augmentation, this approach would not be applicable in our case.

A common approach to enhancing the quality of community detection using link prediction techniques is to preprocess the network to reinforce its community structure. In [19], the authors propose a method of such preprocessing in which link prediction assigns weights to the existing edges of a network and then the community detection algorithm is applied to the weighted network. A more complicated solution proposed in [3] applies a link prediction algorithm multiple times on the same input network, thereby creating a family of enhanced networks. Community detection is then performed for each network in this family. The final result is constructed by aggregating community detection results of each individual network. This approach however, was applied only to disjoint community detection. However, the criminal networks on which we focus in this paper are different in nature, they are much more sparse than the purely social networks studied in [3] and often have an underlying community structure which seems to be much more overlapping than traditional social networks, making community detection more challenging. Finally, in [5], the authors cite incompleteness and inaccuracy of network data collection methods as the reason for the communities based on the collected datasets being different from the ground truth. They aim to recover or improve the network community structure by replacing a fraction of low ranking existing links with top ranked predicted links. To this end they use scores provided by the different link prediction techniques. However, since the criminal networks suffer from scarcity of edges, this approach would not be successful for these kinds of networks.

We also look at some of the work done on the datasets we experiment with in this paper. There has been significant work undertaken to understand the structure of criminal networks. Analyses of the Caviar datasets are presented in [15] and [16]. These two papers focus on retrieving the relative positions of the nodes in the network by using methods based on centrality and use the real position of the nodes revealed in prosecution of the criminals to verify the quality of the proposed method. The two most relevant papers on the Ndrangheta dataset are [4] and [20]. They propose the use of a spectral embedding approach to get to know the relative positions of all the nodes in the network. Another paper [7] uses community detection for another criminal dataset and focuses on identifying important gang members from communities of networks. However, in contrast to these works, we study improving the quality of community detection in criminal networks using link prediction methods [1], [5]. We further extend the usage of these methods to restore networks with partially observable edges.

IV. Methodology

Our method is to augment the criminal networks with edges which probably should exist in the network, had they not been purposefully hidden. We use three standard link prediction metrics to decide which edges to augment the network with. In the following sections, we first introduce the link predictors which we use to select with which edges to augment the criminal network under analysis. We then discuss the community detection algorithms we use to detect communities for analysis of the structure and operation of the criminal network under investigation.

A. Link Predictors

In the literature, several link prediction methods have been suggested in literature, ranging from random walk based methods to machine learning based methods, as mentioned in [10], [13] and [14]. Due to the unique challenges that arise in criminal network community detection discussed above, here we use methods which are fundamental in nature and account for both the network structure and the calls frequency information which is available to us. The reference [9] discusses link prediction methods for criminal networks. Link prediction using attribute similarity between nodes has been demonstrated to work well on certain criminal networks. Since our dataset lacks node attribute information, many of the published methods that rely on sophisticated approaches cannot be applied. Often used in these approaches is learning from previous snapshots of the network, which in the case of criminal networks, may not be useful because of the rapidly evolving nature of such networks. In essence, we use three predictors to assign weights to all created and then sort the edges according to these weights. The predictors used are as follows:

- Number of Common Neighbors, which counts the number of common neighbors between a pair of nodes. We refer to this predictor as $NoCmn$.
- Number of Common Neighbors Weighted by calls which sums the weights of common neighbors by the total number of calls that the two nodes at the endpoints of the created edge make with the common neighbor. We denote this predictor by $WtNoCmn$.
- Number of Common Neighbors Weighted by Hierarchy, which adds edges that connect nodes which occupy a similar position in the hierarchy. We estimate a position of a node in a hierarchy by the volume of communication with the supervisor of this node. Two nodes are assumed to belong to the same level in the hierarchy in the network if they have a similar volume of communication with their common neighbors, who might be their supervisors or colleagues. We weigh each common neighbor of the two nodes of the predicted edge by the inverse of the difference in the communication of the two nodes with this neighbor. A higher difference indicates a difference in the position, and thus a lower propensity to communicate and therefore lower rank of the predicted edge. We refer to this indicator as $WtHier$.

B. Network Augmentation and Community Detection

To augment the network in order to improve community detection, we set a predefined number of edges, n, to be added to the network. In the experiments conducted in this paper, we set the value of n at 25% of the total number of existing edges in the network. These are the edges which do not exist in the network at that particular point of time. We select the top n edges from the list of predicted edges, sorted by weights assigned to these edges by one of the three link predictors defined above. To evaluate the performance of this link augmentation, we check how many of these edges occur in the future snapshots. Even though we evaluate the change in community detection after we perform network augmentation, we still evaluate network augmentation separately, in terms of how real the augmented edges are. We expect that at least some of the edges selected for network augmentation should come into existence, after they have been predicted to exist. So, to evaluate the performance, we perform network augmentation at each snapshot of the network and calculate how many edges out of the ones we predicted occur in the succeeding snapshots.

Community Detection: We perform community detection on the network using the well-known Louvain [6] and SpeakEasy [8] community detection algorithms. The Louvain algorithm is a widely used community detection algorithm which follows a bottom-up approach for identifying communities by optimizing the local modularity of communities. The drawback of the Louvain method is that the communities identified can be unstable, resulting from local modularity optimization. Moreover, the limited connectivity between the communities in the criminal networks we use causes further instability in the communities detected in the Caviar dataset using the Louvain algorithm. The edge augmentation by link prediction approach adds edges which tend to be intra-community rather than inter-community, resulting in more stable communities with higher modularity.

The SpeakEasy community detection identifies communities using Label Propagation in a top-down as well as a bottom up manner, resulting in stable communities. However, the communities identified with SpeakEasy too are not as we expect them to be. First, a larger number of small communities are identified and nodes with large degrees and central locations are not assigned to any community and are identified as a community by themselves. For our analysis, it is important for us to know the community membership of these high degree nodes. With the Caviar network, we perform community detection starting with the 6th snapshot of the network, since it is known that the nodes in the network collaborate well till about the 6th snapshot. From the 6th snapshot onwards, the network begins to modularize a little and specialized, well separated communities can be seen in the 7th snapshot. The four groups we expect to form are the following: a group which focuses on hashish smuggling, a group which specializes in cocaine smuggling, a group which covers the transportation and a group that deals with the financial operations and works with both: the cocaine and hashish groups. Even though these groups are clearly distinguishable from each other from the 7th snapshot onwards, the separation is much fuzzier for the 6th snapshot and even more so for earlier snapshots. For these snapshots the nodes can be clustered into only two communities, one with the hashish leader and the other with the cocaine leader. We perform community detection on the original and the augmented networks and observe the differences in the communities. While we do not know about the community membership of each node, we do know that each of the principal nodes, as mentioned in Section II describing the data, are each expected to form individual community of its own. We also expect communities to be stable and the results to be repeatable. In Section IV, we discuss how network augmentation helps us get stable communities which are closer to the community structure we expect based on the information gained later during prosecution of the gang.

C. Network Restoration

We want to find if our link augmentation techniques can be used to restore community structures in the criminal networks in which many edges are intentionally hidden. We test the performance of our approach on restoring links. This mimics the situations where some links between nodes are either intentionally hidden or not recorded due to the errors in interaction monitoring. Thus, such links are no longer visible to the community detection algorithm and our focus is on examining the accuracy of link predictions in reconstructing these hidden links. To a certain extent, this provides another set of ground truth analyses; however, we will randomly remove edges while criminals may be more strategic in hiding these links. Link restoration is different from network augmentation, as proposed in [5]. With link restoration, we try to restore these edges by choosing the top ranked edges based on the weights assigned to them by the link predictors mentioned above. The link predictors are then recalculated on the network which has some edges randomly removed. We then evaluate how many edges of the ones we removed we are able to restore.

V. Results

We first look at how accurate is the selection of added edges made by the link predictors we use. We then look at a qualitative assessment of community structure detected in the Caviar and Ndrangheta datasets. Finally we look at how many edges we are able to restore from altered criminal networks.

A. Network Augmentation

We predict how many edges created by our link predictor are actually realized in the succeeding snapshots. We report the percentages for accuracy. It measures how many of the predicted edges actually form in the three succeeding snapshots. Table II lists the results for link prediction. We observe that the results are the best for weighted number of common neighbors, followed by number of common neighbors and finally followed by the hierarchy method. We also observe that the prediction results get gradually worse in the later snapshots, indicating increasing stability of the community structure in the network. While these results are not close to what link prediction can achieve in several other social networks, the results are quite significant for a criminal network, where several edges will never be realized, since their formation will be intentionally avoided. The implication of this result is that, the network augmentation method based on the link predictors used can give us reasonably realistic edges. We perform link prediction also on both networks in Ndrangheta dataset but find that the results are poor, which indicates that these network have a strict, hierarchical nature not amenable to our approach.

TABLE II
Accuracy of Link Augmentation for the Caviar dataset for the three link predictors

Seizure no.	$NoCmn$	$WtNoCmn$	$WtHier$
1	66	83	16
2	13	25	25
3	6	13	0
4	25	56	6
5	9	42	41
6	29	32	25
7	0	21	0
8	27	0	0
9	7	0	6
10	6	6	0

B. Community Detection

We first look at the communities obtained with the Louvain community detection algorithm, which is where we see a significant improvement with our approach. We initially perform community detection on the snapshots two to five. We observe that there is a large community with the traffickers, N1, N12 and N3, and there are also three smaller communities at the periphery of the larger community. The communities at the periphery contain nodes which we know are not directly involved in trafficking. In these snapshots, we are not able to split the main trafficker community. In fact we find that the trafficker community does not really split till the 6th snapshot. From this snapshot number, a few more distinct communities begin to form. In this section, we will be presenting community detection results only for the sixth snapshot.

We start with results obtained by augmenting the network in the 6th snapshot of the network using the $WtHier$ link predictor. Communities obtained with the $NoCmn$ link predictor are fairly equivalent, however, worse because a couple of nodes do not belong to the expected communities. Communities obtained with the $WtNoCmn$ link predictor are in comparison less stable, possibly because of the augmentation adds several edges between communities, thus leading to detection of unstable communities.

Since we do not know the membership of each and every node, we cannot provide ground truth comparisons for every node but can examine the communities qualitatively. However, we do know that several communities should exist within the network, and that N1, N12, N76 and N3 should be leading communities of their own, as mentioned in [15] and [16]. So, we expect there to be at least four prominent, stable communities. We perform community detection with different values of the resolution parameter r, with values equal to 1.0 and 1.5, a lower value indicating a higher resolution. We find that the community detection results are quite unstable at the higher resolution level. For example, the communities led by N76, N87 and N3 merge alternatively giving significantly different results each time we run community detection.

We look at the different possible outcomes of community detection on the original network as seen in Figure 1. The communities led by N76 and N1 cannot be well differentiated. Moreover, as seen in Figure 2, we see that the communities led by N1, N3 and N76 cannot be differentiated, while in Figure 3, the communities led by N3 and N76 cannot be differentiated. At low resolution levels, two easily recognizable communities exist, with N1 leading the one, which we call B, and N12 at the center of the second, to which we refer to as community A, as shown in Figure 4. We do not expect this community to split further. On the other hand, community B includes nodes N1, N3, N76 and N87 which play leadership roles in the network, so we expect each of these nodes to form a community of its own. After augmenting the network, even at higher resolution, we are able to obtain four distinct communities, A, B, C and D, as shown in Figure 5. Community A remains unchanged; it is still led by node N12, in the sense that N12 is the node with the highest degree. Community B is led by N1, so we can infer that this community smuggles hashish. Community C includes node N76, so we can conclude that this community is involved in the transportation. Finally, community D includes nodes N3 and N87. This we infer is an intermediary community between communities A and B, with links to the node involved with

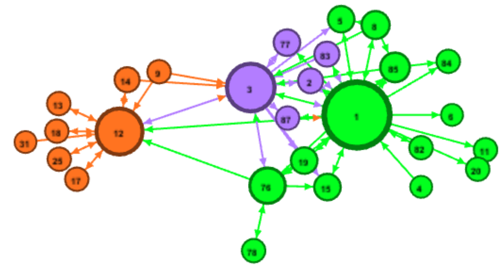

Fig. 1. Communities discovered on the 6th snapshot of the Caviar network with the Louvain algorithm at the r = 1.0, communities led by N76 and N1 are undifferentiated.The different colors represent different communities as detected by the algorithms which however do not agree well with the roles of the nodes which belong to these communities.

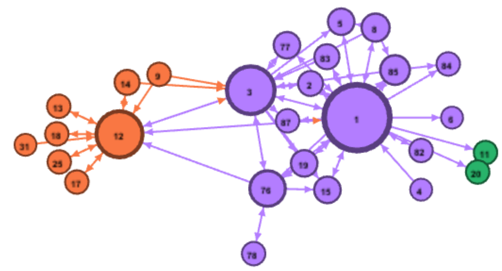

Fig. 2. Another set of communities discovered on the 6th snapshot of the Caviar network with the Louvain algorithm at the r = 1.0, communities led by N1, N76 and N3 are undifferentiated. The different colors represent different communities as detected by the algorithm. The colors denote the same communities as in the previous figure but now the membership of these communities drastically change, indicating that the community detection algorithm is unsuccessful in finding stable communities.

financial investors, N87. So, by augmenting the network, we are able to get closer to a community structure that most likely exists in this network. We are able to obtain distinct communities, each of which plays a different role in the network, as known from the leader of each community.

We also perform community detection using the SpeakEasy community detection algorithm [8]. While the stability of communities is a bit better with SpeakEasy, we still observe that often the high degree nodes are not assigned to any large community and each is considered to be a community on its own. However, knowing the membership of the high degree nodes, which are N1, N3, N12, and N76, is important to us, since they enable us to deduce the function of every group of nodes, as mentioned in [15] and [16]. By augmenting the networks, we observe that the high degree nodes are assigned to the appropriate communities.

We also perform community detection on the Ndrangheta network, and find that the communities are already well separated, thus leaving no scope for improvement. Figure 6 demonstrates communities detected in this network. We additionally mark nodes in this figure according to the roles they play in the network. There are four communities in the network, three of which consist largely of traffickers, support

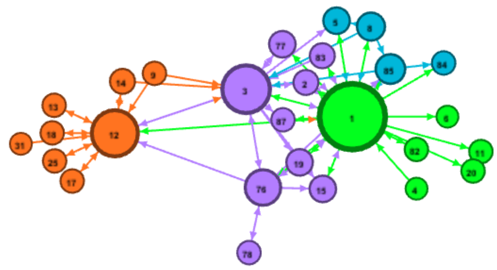

Fig. 3. Another set of communities discovered on the 6th snapshot of the Caviar network with the Louvain algorithm at the r = 1.0, communities led by N3 and N76 communities are not differentiated. The colors denote the same communities as in the previous figures but now the membership of these communities have changed even more, indicating that the community detection algorithm is unsuccessful in finding stable communities.

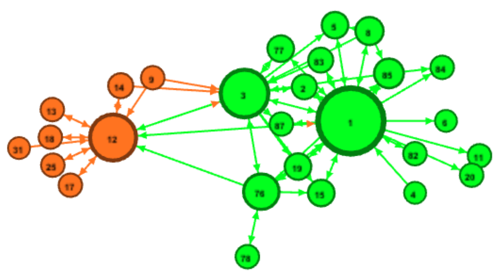

Fig. 4. Communities discovered on the 6th snapshot of the Caviar network with the Louvain algorithm at the r = 1.5, two well differentiated communities can be seen. At this resolution level, the two communities are stable.

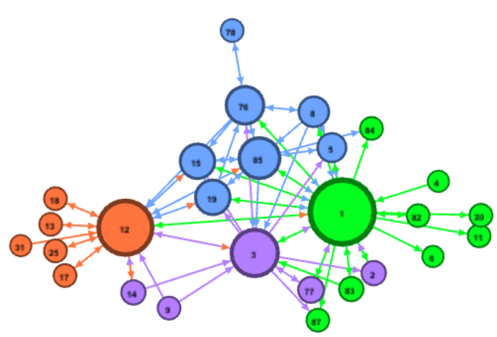

Fig. 5. Well differentiated communities of N1, N3, N76 and N12 after edge augmentation, on snapshot 6 of the Caviar network, with r = 1.0 using the Louvain algorithm.

persons and suppliers, while one of the communities consists only of suppliers and retailers. Similar results for Stupor Mundi gang are shown in Figure 7. We can see for both Italian gang networks a fully evolved structure, indicated by a stable, non-evolving structure, where members strictly obey the modular and hierarchical structure, which is unlike what we observe in the Caviar network.

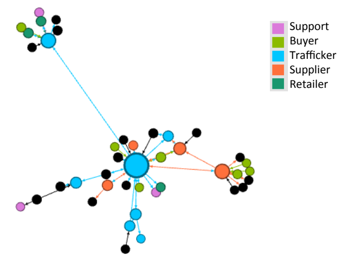

Fig. 6. Strict, hierarchical network of the second snapshot of the Chalonero network, nodes are colored by the roles they have. The communities are clearly differentiated. Each community is led by a trafficker. The roles are available as part of the dataset. Roles are unknown for nodes colored black.

Fig. 7. Strict, hierarchical network of the third snapshot of the Stupor Mundi operation, nodes are colored by the roles they have. The communities are clearly differentiated. Each community is led by a trafficker. The roles are available as part of the dataset. Roles are unknown for nodes colored black.

C. Network Restoration

In this section, we report results for the network restoration performed using the link predictors defined above. We restore the edges by selecting a certain number of top ranked edges. We report results of the experiment in which we alter the network by removing 10% of all network edges, and then attempt to restore them by augmenting twice this number of edges; that is we add 20% of the best edges as predicted by one of the three edge prediction methods. We report the recall of the results, which is, the percentage of deleted edges we are able to identify. Table III lists these predictions. We computed the recall in 10,000 runs, each with different 10% edges removed. We observe that with deletion of certain of edges, it is very difficult to repair the network with the link predictors we use. So first we report results for runs in which we are able to retrieve at least one edge using any of the three predictors. This happens in about 30% of the 10,000 runs we made. Table III lists the results for only these runs on the Caviar network, Table IV lists the results for all runs, for which the results for all the configurations are understandably poorer, since they include a large number of zero values for recall. Similarly, Tables V and VI lists the results for network restoration on the Chalonero network and VII and VIII list the results for network restoration on the Stupor Mundi network. With these results, we are able to demonstrate that link prediction can be of significant use when trying to predict missing edges in these networks. The results also show that when edges essential for communities are hidden, our current method cannot restore them well and therefore, criminals may be able to smartly hide certain links in order to make it difficult for law enforcement to identify them. Clearly, more sophisticated edge augmentation methods are needed perhaps based on the role of nodes.

TABLE III
AVERAGE RECALL OF NETWORK RESTORATION FOR THE CAVIAR NETWORK WITH THE THREE LINK PREDICTORS, USING ONLY THE RUNS WITH RECALL GREATER THAN 0.

Seizure no.	$NoCmn$	$WtNoCmn$	$WtHier$
1	76.5	76.9	76.9
2	57.1	56.3	56.7
3	23.1	34.2	30.2
4	30.0	29.8	28.2
5	33.0	41.0	39.6
6	23.7	36.5	31.4
7	19.2	35.9	34.8
8	24.0	31.9	29.2
9	33.4	36.6	35.2
10	24.3	30.8	26.1
11	27.3	26.5	24.4

TABLE IV
AVERAGE RECALL OF NETWORK RESTORATION FOR THE CAVIAR NETWORK FOR THE THREE LINK PREDICTORS FOR ALL THE 10,000 RUNS

Seizure no.	$NoCmn$	$WtNoCmn$	$WtHier$
1	1.1	0.7	0.5
2	4.5	5.5	6.2
3	7.6	13.2	8.1
4	13	7.2	6.1
5	13.9	5.8	3.4
6	8.0	21.5	10.1
7	1.9	8.1	5.9
8	5.7	11.2	11.9
9	13.9	11.4	10.8
10	5.9	6.3	4.5
11	4.6	4.5	9.5

TABLE V
AVERAGE RECALL OF NETWORK RESTORATION FOR THE CHALONERO NETWORK WITH THE THREE LINK PREDICTORS, USING ONLY THE RUNS WITH RECALL GREATER THAN 0.

Seizure no.	$NoCmn$	$WtNoCmn$	$WtHier$
1	56.4	56.1	0.5
2	36.7	36.8	6.2
3	41.3	40.8	8.1
4	43.7	43.3	6.1

VI. Conclusion and Discussion

We propose a link prediction based method which improves community detection in criminal networks. We demonstrate that the method improves the quality of community detection in a real life criminal network. We also show that the link prediction based methods are able to repair the networks in which certain percentage of edges is randomly removed. However, as demonstrated in the network restoration subsection, if the links essential for community detection are hidden, our current network augmentation methods may not always be able to restore such links. We suspect that such essential edges link leaders to their subordinates. Hence, in the future work we will investigate methods of edge augmentation based on the role and position of the nodes in criminal network hierarchy.

TABLE VI
AVERAGE RECALL OF NETWORK RESTORATION FOR CHALONERO NETWORK WITH THE THREE LINK PREDICTORS FOR ALL THE 10,000 RUNS

Seizure no.	$NoCmn$	$WtNoCmn$	$WtHier$
1	48.2	48.2	0.0
2	33.1	33.8	0.0
3	25.2	25	0.0
4	25.2	25.6	6.1

TABLE VII
AVERAGE RECALL OF NETWORK RESTORATION FOR THE STUPOR MUNDI NETWORK WITH THE THREE LINK PREDICTORS, USING ONLY THE RUNS WITH RECALL GREATER THAN 0.

Seizure no.	$NoCmn$	$WtNoCmn$	$WtHier$
1	49.7	49.5	45.1
2	41.8	42.1	40.3

TABLE VIII
AVERAGE RECALL OF NETWORK RESTORATION FOR STUPOR MUNDI NETWORK WITH THE THREE LINK PREDICTORS FOR ALL THE 10,000 RUNS

Seizure no.	$NoCmn$	$WtNoCmn$	$WtHier$
1	27.1	26.5	23.9
2	22.6	23.2	22.5

Acknowledgment

This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number 2017-ST-061-CINA01, by the Office of Naval Research (ONR) Grant No. N00014-15-1-2640, and by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053 (the ARL Network Science CTA). The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security or the Army Research Laboratory.

References

[1] A. Bahulkar, B. K. Szymanski, O. Lizardo, Y. Dong, Y. Yang and N. V. Chawla, “Analysis of Link Formation, Persistence and Dissolution in NetSense Data, 6th Workshop on Social Networks Analysis and Applications (SNAA),” in Proc. 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, CA, August 18-21, pp. 1197-1204, 2016.

[2] N. O. Baycik, T. C. Sharkey, and C. E. Rainwater, “Interdicting layered physical and information flow networks,” *IISE Trans.*, vol. 50, no. 4, pp. 316–331, Apr. 2018.

[3] M. Burgess, E. Adar, M. Cafarella, "Link-prediction enhanced consensus clustering for complex networks," PloS ONE, 11(5):e0153384, 2015.
[4] F. Calderoni, D. B. Skillicorn and Q. Zheng, Quan, "Inductive Discovery Of Criminal Group Structure Using Spectral Embedding," Information and Security: An International Journal. 31. 49-66. 10.11610/isij.3102, 2014.
[5] M. Chen, A. Bahulkar, K. Kuzmin and B. K. Szymanski, "Improving Network Community Structure with Link Prediction Ranking," Proc. 7th Workshop on Complex Networks (CompleNet), Dijon, France, Mar. 23-27, Studies in Artificial Intelligence, 644, Springer, Zurich, Switzerland, pp. 145-158, 2016.
[6] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks, Phys. Rev. E, vol. 70, pp. 16, Dec. 2004.
[7] E. Ferrara and M. De, P. C. Salvatore and F. Giacomo, "Detecting criminal organizations in mobile phone networks. Expert Systems with Applications. 41. 10.1016/j.eswa.2014.03.024, 2014
[8] C. Gaiteri, M. Chen, B.K. Szymanski, at al., "Identifying robust communities and multi-community nodes by combining top-down and bottom-up approaches to clustering," Scientific Reports, 5(16361), 2015.
[9] B., Giulia, F. Calderoni, N. Parolini, M. Verani and C. Piccardi, "Link Prediction in Criminal Networks: A Tool for Criminal Intelligence Analysis," PloS one 11 4, e0154244, 2016.
[10] Liben-Nowell, D., Kleinberg, J., The link prediction problem for social networks, Proceedings of the twelfth international conference on Information and knowledge management, ACM, pp. 556-559 (2003)
[11] Z. Liu, J. L. He, K. Kapoor, J. Srivastava, "Correlations between community structure and link formation in complex networks. PLoS ONE 8, 72,908, 2013.
[12] Z. Liu, W. Dong, Y. Fu, "Local degree blocking model for missing link prediction in complex networks. Chaos: An Interdisciplinary Journal of Nonlinear Science, 25(1):01315, 2015.
[13] Lu, L., Zhou, T.: Link Prediction in Complex Networks: A Survey Physica A: Statistical Mechanics and its Applications Volume 390, Issue 6, P1150-1170 (2011)
[14] Menon, A., Elkan, C.,: Link Prediction via Matrix Factorization, ECML PKDD 2011, Part II, LNAI 6912, pp. 437452, (2011)
[15] C. Morselli and K. Petit "Law-Enforcement Disruption of a Drug Importation Network", Global Crime, vol. 8, Issue 2, 2007.
[16] C. Morselli and C. Giguere, "Legitimate strengths in criminal networks", Crime, Law and Social Change, pp. 185–200, 2006.
[17] B. Yan, S. Gregory, "Finding missing edges and communities in incomplete networks," J. Phys. A 44, 495,102, 2011.
[18] B. Yan, S. Gregory, "Finding missing edges in networks based on their community structure," Phys. Rev. E 85(5), 056,112, 2012.
[19] B. Yan, S. Gregory, "Detecting community structure in networks using edge prediction methods," J. Stat. Mech: Theory Exp. 2012(09), P09,008, 2012.
[20] Q. Zheng and D. B. Skillicorn, "Analysis of criminal social networks with typed and directed edges," 2015 IEEE International Conference on Intelligence and Security Informatics (ISI), 1-610.1109/ISI.2015.7165930, 2015.