CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	A graph neural network-based role classification in criminal networks CSCI–RTCW
Authors	Dogan, Vedat;Prestwich, Steven D.;O'Sullivan, Barry
Publication date	2024-12-12
Original Citation	Dogan, V., Prestwich, S. and O'Sullivan, B. (2024) 'A graph neural network-based role classification in criminal networks CSCI–RTCW', 2024 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, December 11-13, 2024.
Type of publication	Article (peer-reviewed)
Rights	© 2024, the Authors. This work is made available under the CC BY license (https://creativecommons.org/licenses/by/4.0/)
Download date	2026-11-02 00:59:33
Item downloaded from	https://hdl.handle.net/10468/17404



A Graph Neural Network-based Role Classification in Criminal Networks CSCI–RTCW

Vedat Dogan *[0000−0002−3807−5425], Steven Prestwich[0000−0002−6218−9158], and Barry O'Sullivan[0000−0002−0090−2085]

Insight SFI Research Centre for Data Analytics
School of Computer Science & IT, University College Cork, Ireland
{vedat.dogan, steven.prestwich, barry.osullivan}@insight-centre.org

Abstract. Understanding the roles of individuals in terrorist networks is an important task in counter-terrorism. This paper presents the first application of graph neural networks to this task. We apply our approach to a real-world terrorist network representing three different ideologies and nine specific groups. We demonstrate the challenges associated with this task and present the framework using graph neural networks and their advantages in this context.

Keywords: Terrorist Networks · Counter-terrorism · Role Identification · Graph Neural Networks · Node Classification.

1 Introduction

Terrorism is becoming an international concern that creates a threat to world security. In 2021, the Member States of the European Union (EU) reported 57 completed terrorist attacks and the United Kingdom (UK) reported 62 incidents caused by terrorist groups. The deaths caused by the terrorist attacks have increased 5-fold after the attacks on the United States by Al-Qaeda on 11 September 2001. As terrorism activities are increasing each year worldwide, preventing an attack along with analysing the structure of a terrorist network is becoming vital. In particular, identifying an individual's role in a terrorist network can minimize casualties. For instance, if we can identify the role of a person in a network at an early stage of a terrorist network construction, the information can be used to warn counter-terrorism authorities and understand the possible influence of this person, and prevent a possible future activity. Moreover, the information can be used to reshape the network after the death of an influential person on a network. Another possibility is the obstruction of the online activities of terrorist groups. Terrorists extensively use the internet to spread their messages to radicalise, recruit and facilitate terrorist attacks, as highlighted by Europol in the Terrorism Situation and Trend Report (TE-SAT) in 2023 [25]. Identifying the roles of these individuals could help to prioritize observations of their online activities, to gain insight into potential casualties. Node classification, or the task

* Corresponding Author.

of identifying the role of an individual in the network, is essential in this regard. However, the covert nature of terrorist activities and the lack of available data impedes the performance of traditional node classification methods.

The rapid advancement of machine learning (ML) and artificial intelligence (AI) is reforming various sectors including counter-terrorism. Specifically, Graph Neural Networks (GNNs) are a class of deep learning ML models designed to work with graph-structured data. They have achieved exceptional success in various tasks such as node classification, link prediction, graph construction and community detection. The ability of GNNs to learn from network structures along with node features makes them good candidates for node classification.

In this paper, we propose a framework that takes advantage of the power of GNNs to predict the roles of individuals in terrorist networks. To the best of our knowledge, this is the first attempt to apply GNNs for node classification in terrorism networks. We conducted extensive experiments on real-world terrorist networks and subnetworks considering various kinds of characterization. It leverages the unprecedented ability of GNNs to model structures and relationships in networks, predicting feature importance while identifying the role of individuals. The results demonstrate that the proposed framework with GNNs can handle unique challenges presented by terrorist networks including imbalanced class and using side information while learning the network structure. This framework provides a GNN-based model for node classification in terrorism networks and a benchmark for future research in this area. Moreover, the proposed framework offers important insights into the underlying terrorist networks.

2 Related Work

In this section, we review the existing relevant literature. The discussion is divided into two subsections: Terrorism Networks Analysis and Applications of GNNs.

2.1 Terrorism Networks Analysis

Several works conducted over the years to predict the future activity of a terrorist group or the influence of a specific note or whole network by using a set of computational techniques from data mining, ML and SNA. Artificial neural networks (ANNs) are applied to train a prediction model to handle the terrorism insurgency in [9]. They also provide a risk assessment of improvised explosive device incidents between 2007 - 2013. Identifying the field of potential groups and the key members in it issued by [14]. The results show that visual network models can provide better intuition to learn the hidden structure of the groups. [13] proposes a framework to analyse the group profiles through the events and generate terrorist and criminal group profiles. SNA methods conducted in [4] on Australian-based jihadists across different periods and aimed to reveal hidden connections between separate groups. Recently, a deep neural network has been used to predict future activities in [26] and compared with traditional regression algorithms. They show that deep learning methods are suitable for predicting

the behaviour of terrorist activities. Most recently [1] proposed a GNN-based method to identify terrorist organizations using link prediction. We refer the readers interested in more in-depth literature to [19], which is a systematic survey about computational techniques to counter terrorism.

2.2 Applications of GNNs

Various kinds of domains are explored by GNNs in supervised, semi-supervised and unsupervised settings. Many works have been done in the domain of graph mining [18], chemistry and biology [8], combinatorial optimization for solving TSP [2], routing problem [12], to classify the satisfiability of SAT problem [22]. [21] also provides a theoretical analysis of GNNs on combinatorial problems. It also has been successfully applied to traffic networks [31,32], robotics [20] and economics to model to predict future trends [16]. For more in-depth research please see [10, 28]. Even though GNNs have been explored in numerous domains, there are limited works in the military and counter-terrorism.

In summary, no work has been done on GNN-based node classification for terrorist networks to explore the potential of state-of-the-art AI methods and the limitations of using terrorist network graphs with various characteristics.

3 Preliminaries

We now define the node classification problem. First, we define the problem and then we briefly explain how GNNs work on node classification tasks.

3.1 Problem Statement and Definition

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A}, \mathcal{X}, \mathcal{C})$ be a graph network where $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ is a set of nodes (terrorists), $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is a set of edges between nodes (relationships between terrorists) in $\mathcal{V}$, $\mathcal{A} \in \{0,1\}^{n \times n}$ is an adjacency matrix, $\mathcal{X} \in \mathbb{R}^{n \times d}$ is an attribute matrix assigning attributes to the nodes and $\mathcal{C} \in \{0,1\}^{n \times y}$ is a class matrix there n, d, y are the number of nodes, number of attributes of each node and the number of classes, respectively. The goal of GNNs is to learn effective node representations by combining graph structure information and the node attributes which are used for node classification. After splitting the nodes into train/test sets, for a given adjacency matrix $\mathcal{A}$, an attribute matrix $\mathcal{X}$ and a class matrix $\mathcal{C}$ contains the class information of the nodes in training data, we predict the labels of the nodes in the test set.

3.2 Graph Neural Networks for Node Classification

GNNs are inarguably the hottest topic in graph-based deep learning. They are inspired by convolutional neural networks (CNNs) which use multiple convolutional layers to learn local and global structural patterns. Recently, many approaches proposed and we briefly explain some of them.

The graph convolutional network (GCN) is the standard model proposed in [11]. Another approach with attention mechanisms to learn the relative weights between each connected node is proposed in [27] and called graph attention network (GAT). Other than convolutional graph neural networks, another approach is sampling-based GNNs which focus on computing node representations from subgraphs of an input graph, such as neighbourhoods. For instance, GraphSAGE is proposed in [5], which learns a function that generates embeddings by sampling features from a node neighbourhood. [23] proposed a unified message-passing model called UniMP, which uses a graph transformer network for feature and label embedding as an input for propagation. Numerous GNN algorithms have been developed for node classification in recent years [10, 28, 29].

4 Empirical Studies

In this work, we aim to predict the role of an individual in a terrorism network considering different graph sizes, ideologies and individual background information. We also explore the different classification algorithms' performance and discuss the results in detail.

4.1 Dataset

We used a dataset from the Profiles of Individual Radicalization in the United States (PIRUS) retrieved and maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) [24]. It contains masked individual-level information on the backgrounds, attributes and radicalization process of over 3200 violent and non-violent extremists in the United States covering 1948-2021. The dataset contains 131 attributes such as ideology, management and identification variables, group nature, radicalization details, demographics, socioeconomic status with personal information, and the number of casualties and consequences. It also comes with the relationships between the offenders in PIRUS to perform social network analysis (SNA). These files come from the Social Network of American Radicals (SoNAR) dataset. The dataset is publicly available on [1]. The datasets require cleaning before a network construction and investigation. To make a precise determination, we deleted the individuals from SoNAR who have no information in the PIRUS dataset. The individuals in the PIRUS dataset follow 9 ideologies and have 3 roles in total including *unknown* which is not identified. We also removed them to determine precise experiments. After cleaning all non-connected nodes in the whole dataset, we had 1645 nodes and 4684 edges representing the connections between them. Table 1 summarizes several basic network characteristics of the entire terrorism network and the selected subnetworks. We selected 49 node features that we believe are connected to the individual lethality of a member and vital for the identification of their role in the group. One-hot encoding is applied to the features for all nodes and we obtained 293 encoded features for the whole dataset.

[1] http://www.start.umd.edu/pirus

Table 1. The topological features of the entire network and selected subnetworks. **v** and **e** represent nodes and the edges while **k** and $\mathbf{k}_{max}$ represent average and maximum degree. **density** and **c** stands for network density and the global clustering coefficient.

Networks	v	e	k	$\mathbf{k}_{max}$	density	c
PIRUS Network (all)	1645	4684	3.84	21	0.0023	0.58
Nativist Group	297	617	3.47	21	0.0068	0.45
Islamist Group	510	1315	3.08	11	0.0104	0.51
Anti-Government Group	278	929	4.83	20	0.0174	0.58
ISIL	160	153	1.35	11	0.0084	0.23
Ku Klux Klan	151	200	1.88	7	0.0125	0.31
Sovereign Citizens	80	37	0.87	4	0.0110	0.23
QAnon	116	26	0.22	2	0.0019	0.03
Proud Boys	88	87	1.50	8	0.0172	0.23
al-Qaeda	71	90	2.36	6	0.0338	0.47
Boogaloo Movement	73	175	2.46	10	0.0342	0.27
Jewish Defense League (JDL)	52	91	2.61	10	0.0512	0.26
al-Shabaab	49	135	3.30	8	0.0688	0.44

4.2 Graph Network Construction

We can construct the terrorism network after cleaning the dataset. The subject of the network is the terrorists which can be represented by nodes and edges to illustrate the connections between individuals. A graph can be weighted, unweighted, directed or undirected. In this work, we focus on undirected attributed graphs. Each node contains information about their profiles including *ideology* and *role*. We construct the attributed matrix $\mathcal{X}$ considering these features including SNA metrics such as *betweenness centrality*, *closeness centrality*, etc. to capture network structure details. For each node, we selected 49 attributes representing the profiles of each node. We marked members of the group according to their role in the whole network, which represents their classes. The set of possible roles is $R = \{Leader, Follower, Loose\ Associate\}$ which represents leadership status in the group and therefore the importance of the individuals.

4.3 Experimental Setup

In this study, we conduct several experiments to classify nodes according to their role property. We first evaluate performance considering the whole network. Then we divide the network into subnetworks considering the individual ideology, i.e. *islamist*, *nativist* and *anti-government*. Then we evaluate the experiments on these subnetworks to explore the performance of classification algorithms in these subnetworks. The evaluations in these classes are performed to observe the effect of ideological network structure on the identification of important individuals. We also conduct experiments to measure the classification performance on different groups of ideologies. The network which follows Islamic ideology contains ISIS, AlQaeda, AlShabaab groups, and the network follows Nativist ideology contains Ku Klux Klan, Proud Boys, JDL and Anti-Government ideology contains Sovereign Citizens, Boogaloo Movement and QAnon.

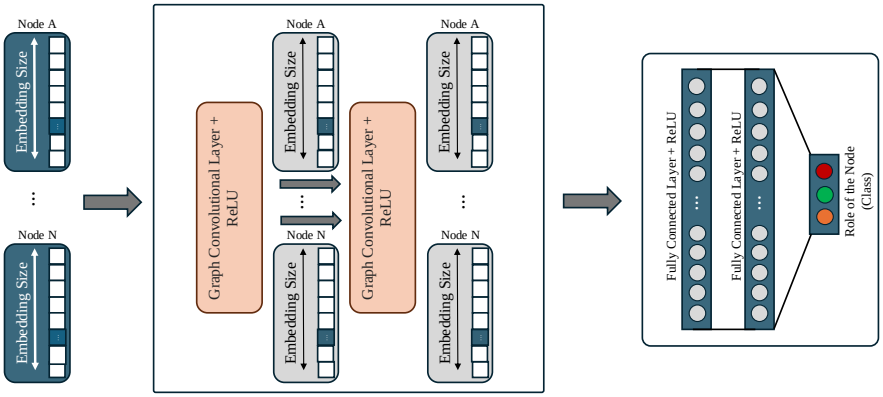


Fig. 1. Schematic representation of GNN architecture used to identify the role of an individual on a criminal network.

We use 3 GNNs: GCN, GraphSAGE and UniMP which use transformer layers as discussed in Section 3.2. We use 2 convolutional layers for each GNN algorithm with 16 hidden layers. We shared the general schematic representation of GNN algorithm architecture in Figure 1. We use the weighted cross-entropy loss function for dealing with imbalanced class problems on group networks. For the UniMP model, we set the *heads* parameter to 1, which is the default. We also use a graph-agnostic classifier which ignores the topology structure: a multi-layer perception (MLP) with 2 linear layers and 16 hidden layers, to observe the effect of topological structure on classification performance. We choose the weighted cross-entropy loss function to deal with the imbalanced class problem on group network datasets. We use 70% of the data for training and 30% for testing. We apply the SMOTE method for the whole Pirus network and ideological group subnetworks and use the weighted cross-entropy loss function for small group networks. All experiments are repeated 10 times and shuffle split cross-validation is applied to the dataset for each run. In our experiments, we use *weighted average* precision, recall and F1-score to evaluate classification quality, to better reflect the performance of minority classes. The average scores and standard deviation of whole runs are shared. We use the Python programming language on a single core of 1.4 GHz Quad Core i5, 8Gb 2133 MHz LPDDR3 RAM for the experiments.

4.4 Classification Quality on Real World Terrorism Graphs

We divide the original network into three subnetworks to explore the ideological terrorism network's effect on classifying individuals' roles in their groups. Each subnetwork contains 49 attributes for each node identifying personal information including SNA metrics. We first conduct our experiment on the whole PIRUS data including all ideologies, then we run the experiments on subnetworks.

Table 2. Performance comparison of node classification task for whole network and subnetworks with different philosophies.

Weighted Average Scores					
Ideologies	**Metrics**	**MLP**	**GCN**	**GraphSAGE**	**UniMP**
PIRUS Network (All)	*Precision*	0.6265±0.02	0.6741±0.03	0.6840±0.02	0.6937±0.02
	Recall	0.5799±0.03	0.5806±0.02	0.5896±0.03	0.6452±0.02
	F1-Score	0.5669±0.03	0.5854±0.02	0.5891±0.03	0.6708±0.01
Nativist Group	*Precision*	0.5273±0.03	0.5941±0.04	0.6274±0.03	0.6367±0.03
	Recall	0.5083±0.03	0.4799±0.04	0.5613±0.03	0.6363±0.03
	F1-Score	0.5167±0.03	0.5116±0.04	0.5775±0.03	0.6338±0.03
Islamist Group	*Precision*	0.5558±0.03	0.6468±0.04	0.6801±0.04	0.6638±0.02
	Recall	0.5387±0.03	0.5667±0.05	0.5933±0.05	0.6403±0.04
	F1-Score	0.5455±0.03	0.5611±0.05	0.5633±0.04	0.6385±0.03
Anti-government Group	*Precision*	0.6403±0.04	0.6461±0.06	0.6346±0.03	0.6633±0.04
	Recall	0.6330±0.05	0.6520±0.08	0.6634±0.05	0.6732±0.05
	F1-Score	0.6315±0.04	0.6367±0.07	0.6716±0.05	0.6524±0.04

Performances on Whole Network We aim to clarify how classification performance is affected when we use the whole network. To investigate the whole network, we need to deal with class imbalance between different roles in the groups. The imbalance of roles in a group comes naturally as it is expected that the number of *leaders* is much less than the number of *loose associates* or *followers*. To classify this imbalanced data, we first split our data training and test sets. Then we applied the popular Synthetic Minority Over-sampling Technique (SMOTE) to training data presented in [3]. To avoid augmenting the test set we left it as real data. We conduct experiments with 3 GNN-based classification methods and 1 artificial neural network to compare GNN results.

Observations Table 2 shows the average node classification performance of the network including the standard deviation of all ideologies in the dataset. We compare the classification results performance using precision, recall and F1-score. Compared with the other GNN-based methods, we see that UniMP has the best F1-score. Typically, GCN and GraphSAGE, which use neighbourhood embeddings, do not perform as well as UniMP which uses a message-passing model with graph transformer-based layers for feature and label embeddings. In terms of the precision metric UniMP reached the best score, however, its recall and F1-score scores are weaker than the others'. Note that we share the weighted average scores, which means the measures take into account the support while calculating the average. There are few studies that mention class imbalanced problems with GNNs, such as [17]. Classifying the leader in a criminal network is crucial in uncertain situations, as mentioned in Section 1. For presenting the class-based performances, we show the average class accuracies for each class in Table 3. We observe that the GraphSAGE model reached 68% accuracy, which is the best for classifying the leader for the whole network.

Table 3. Performance comparison of node classification task per class for all networks.

Accuracies per Class						
Networks	**Category**	**MLP**	**GCN**	**GraphSAGE**	**UniMP**	**Majority Voting Ensemble**
PIRUS Network (All)	*Loose Associate*	0.2279	0.3372	0.2721	0.1651	0.3255
	Follower	0.6273	0.3928	0.4713	0.6766	0.7734
	Leader	0.5524	0.6444	0.6872	0.5628	0.5882
Nativist Group	*Loose Associate*	0.0636	0.2182	0.0545	0.0414	0.0909
	Follower	0.6938	0.4682	0.5636	0.7312	0.7586
	Leader	0.5297	0.5484	0.6438	0.5484	0.5781
Islamist Group	*Loose Associate*	0.2100	0.2700	0.2700	0.1700	0.1769
	Follower	0.7648	0.4132	0.5121	0.7505	0.7534
	Leader	0.2389	0.4500	0.5222	0.2778	0.3333
Anti-Government Group	*Loose Associate*	0.2857	0.3143	0.3143	0.2571	0.2428
	Follower	0.7310	0.4831	0.5479	0.7085	0.8148
	Leader	0.5000	0.5471	0.6471	0.5206	0.7647

Performances on Ideological Groups We aim to clarify and examine the classification results with different ideological terrorist networks. The goal is to analyse the performance and classification performance according to structured networks with their specific philosophy. We used the same network to create the subnetworks. Then we choose the top three ideologies according to their size: Nativists, Islamists and Anti-government groups. We split data nodes train and test sets for each group and applied the SMOTE method to deal with class imbalance on training data.

Observations Table 2 shows the node classification performance with different subnetworks created from the main network, according to their philosophy. The UniMP reached the best performance compared with other methods. We can observe that in the Nativist Groups experiment, the best F1-score is obtained by the UniMP algorithm which uses the message-passing models with transformer layers. We believe the reason for the success of GNNs comes from the structure of the Nativist network. In Table 1 the maximum degree of the Nativist Group is 21 and the average degree is 2.4705 with 0.0068 density. This means that there are nodes that exist connected with many nodes, providing more neighbourhood information, but the low density shows that many nodes are not connected. Therefore GNN algorithms such as UniMP can identify successfully the roles of individuals by using these node features and their neighbourhoods. UniMP also reached the best F1-score for the Anti-government group, which has a maximum degree of 20, close to that of the Nativist Group. We can see from Table 1 that the average density of the network is 4.8345 which is higher than the Nativist Group. MLP algorithm reached better compared with GNNs, with the best result for the Islamist group which is relatively close to the GNNs. We believe the reason for the weak performance of GNNs is the low average and maximum density in the network. Even though the node and edge numbers are high, the densities are the lowest which means there are many isolated groups in the network. This may lead to the wrong information for connected nodes and the learnt structure

Table 4. Performance comparison of node classification task for subnetwork with different groups.

Weighted Average Scores					
Groups	**Metrics**	**MLP**	**GCN**	**GraphSAGE**	**UniMP**
ISIL	*Precision*	0.6891±0.06	0.6329±0.04	0.7143±0.04	0.6806±0.06
	Recall	0.7016±0.05	0.6094±0.06	0.7216±0.05	0.7047±0.06
	F1-Score	0.6880±0.05	0.6174±0.05	0.7126±0.04	0.6898±0.06
Ku Klux Klan	*Precision*	0.6442±0.06	0.5526±0.08	0.6484±0.07	0.5977±0.06
	Recall	0.6459±0.06	0.5508±0.08	0.6593±0.07	0.6098±0.05
	F1-Score	0.6390±0.06	0.5432±0.08	0.6440±0.07	0.5876±0.05
Sovereign Citizens	*Precision*	0.6732±0.08	0.6410±0.09	0.6845±0.09	0.7136±0.06
	Recall	0.7031±0.08	0.6781±0.08	0.7156±0.07	0.7413±0.06
	F1-Score	0.6733±0.07	0.6535±0.07	0.6868±0.07	0.7182±0.05
QAnon	*Precision*	0.7095±0.05	0.6854±0.04	0.7089±0.05	0.6836±0.06
	Recall	0.7383±0.04	0.7000±0.04	0.7383±0.04	0.7064±0.06
	F1-Score	0.7195±0.04	0.6857±0.03	0.7154±0.05	0.6885±0.05
Proud Boys	*Precision*	0.6226±0.08	0.5950±0.07	0.5842±0.07	0.6056±0.10
	Recall	0.6222±0.07	0.5694±0.08	0.5944±0.06	0.6306±0.09
	F1-Score	0.6108±0.07	0.5668±0.07	0.5757±0.04	0.5882±0.08
AlQaeda	*Precision*	0.5949±0.07	0.5393±0.05	0.5756±0.07	0.5237±0.04
	Recall	0.6655±0.06	0.5690±0.10	0.6552±0.07	0.6241±0.09
	F1-Score	0.6203±0.05	0.5461±0.06	0.6026±0.05	0.5666±0.06
Boogaloo Movement	*Precision*	0.6623±0.08	0.5271±0.12	0.6126±0.11	0.6048±0.08
	Recall	0.6933±0.06	0.4067±0.11	0.6800±0.04	0.6200±0.11
	F1-Score	0.6637±0.06	0.4388±0.11	0.6273±0.06	0.5992±0.10
JDL	*Precision*	0.5072±0.14	0.4663±0.10	0.5308±0.12	0.5187±0.12
	Recall	0.5190±0.15	0.4476±0.11	0.4952±0.12	0.4857±0.10
	F1-Score	0.5097±0.15	0.4429±0.10	0.5174±0.11	0.4894±0.09
AlShabaab	*Precision*	0.6330±0.04	0.6680±0.09	0.6471±0.09	0.6553±0.06
	Recall	0.6350±0.05	0.4250±0.12	0.6000±0.12	0.6000±0.12
	F1-Score	0.6328±0.04	0.4913±0.10	0.6144±0.10	0.6095±0.09

of the general network by GNNs. We can see the class accuracies in Table 3 for ideological subgroups. GraphSAGE model reached $64\%, 52\%$ and 65% as the best accuracies compared with other methods for all Nativist, Islamic and Anti-government groups for classifying the leader task in the networks.

Performances on Specific Terrorist Groups We aim to analyse the performance of classification algorithms including GNNs considering the specific groups belonging to the same and also different ideologies. We selected nine different terrorist groups. The groups are subnetworks of the whole network, and their nodes and features are used for training the selected models. To deal with the imbalanced classes in the data, we integrated the weighted loss function into our GNN models. The weights are defined based on the inverse class frequencies that give more importance to the minority classes. The reason that we did not use the SMOTE method is the lack of data in the minority class.

Observations Table 4 shows the node classification performance with different terrorist group networks having different or the same ideologies. We observe that

the GraphSAGE algorithm have the highest F1-scores for the role classification of individuals in ISIL, Ku Klux Klan, and Sovereign Citizens groups. We can observe that the reason of a good performance score of the GraphSAGE algorithm on ISIL and JDL groups is high k_{max} values with similar network density. On the contrary, the QAnon group has the lowest k_{max} value and MLP reached better results than GNN, which is expected as GNNs struggle to learn from network structure when there is low network density. When we look at the network topology of these groups in Table 1, we can see GNNs are successful when the degree of network and maximum degree in the network are both high. Also, node and edge sizes are important for GNNs to learn the graph structure along with the clustering coefficient and degree. These four groups for GNNs and two groups for MLP have high k and c values in Table 1. As long as the network size gets smaller, we can see the performance of GNN algorithms is worse. More in-depth analysis is studied in [15], and they declare that high imbalance classes in the data along with small network size might lead negative affect on classification performance, because a small network with a high imbalance of data can lead to wrong patterns in the learning characteristics of the classes.

4.5 Explainability of the GNN Models

We select the networks in which GNN algorithms succeeded compared to other methods, and perform an analysis to evaluate the feature importance of GNN predictions for each graph network. We use the GNNExplainer library [30] for analysing the explainability of GNNs, which is a library that uses a model-agnostic approach to provide interpretable explanations for predictions of any GNN-based model. It does this by taking a trained GNN model and its predictions and returning an explanation in the form of a small subgraph of the input graph together with the node features that are most influential for the prediction. The results of the analysis include a set of feature contributions to the prediction that can be used to understand and interpret the model's behaviour. Figure 2 shows the feature importance for the best GNN algorithms for classification tasks on the Nativist, Anti-government, ISIL, Ku Klux Klan, AlQaeda and JDL groups.

For the Nativist group, the GraphSAGE algorithm reached the best F1-score and we can see from Figure 2 the *Extent_ Plot*, *Foreign_ Govt_ Leader* and *Lenght_ Group* are the top-3 features that influenced the decision. *Extent_ Plot* represents if the individual's extremist activity involved a violent plot and 5 represents the successful execution of the plot, meaning that it plays a crucial role in identifying the role of a person in the group. *Foreign_ Govt_ Leader* represents whether an individual's radicalization is connected to specific actions by a foreign government or particular foreign leaders, which again seems important for GraphSAGE's ability to classify the role of individuals in the Nativist group.

The UniMP model reached the best F1-scores scores as we see in Table 2. We can observe the feature influences for decision-making in Figure 2. Again the *Extent_ Plot* feature was the most important one for the classification. *Foreign_ Govt_ Leader* and *Recruiter1* follow it, meaning that the individual is actively recruited by family, friends or another member of the group. We observe

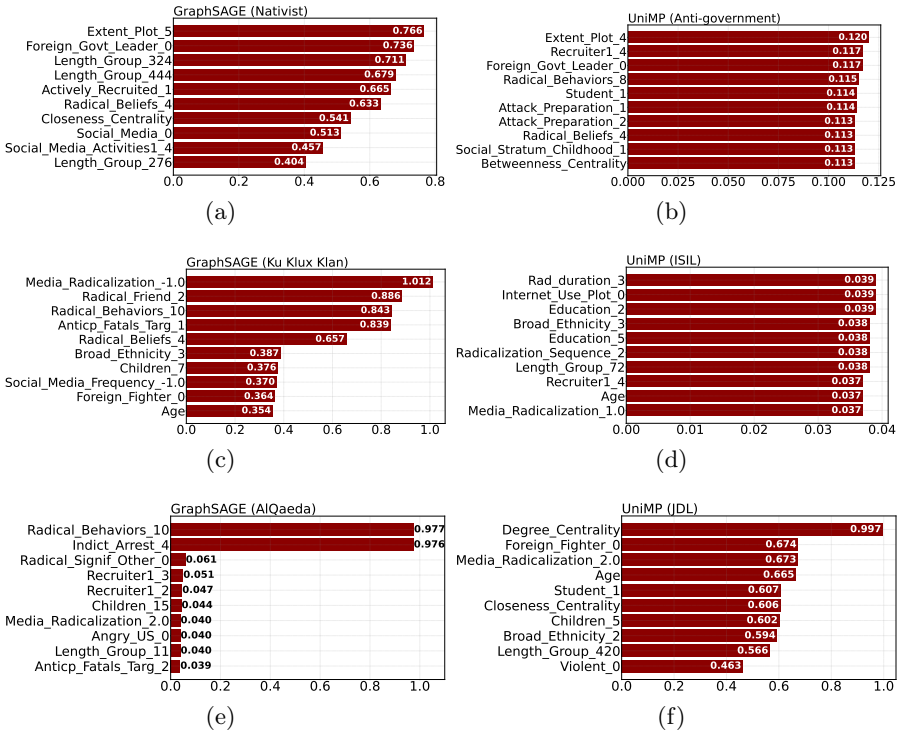


Fig. 2. Feature importance graphs of succeeded GNN algorithms for Nativist (a), Anti-government (b), Ku Klux Klan (c), ISIL (d), AlQaeda (e) and JDL (f) groups.

the most important two features for the individual's role classification for both Nativist and Anti-government groups are *Extent_ Plot* and *Foreign_ Govt_ Leader*, as explained above. We observe that violent activities and foreign government-related connections are crucial to identifying the role of an individual in these groups, according to the winning GNN models.

We see from Table 2 and Figure 2 that the GraphSAGE algorithm has the best classification scores, and one of the top 3 important features is *Media_ Radicalization* which represents whether the media besides the internet (books, movies, television shows, radio) plays a role in the individual's radicalization. *Radical_ Friend* represents a close friend of the individual involved in radical activities. Last of the top 3 is *Radical_ Behaviors* which is a feature that tells us the maximum extent of radicalization apparent in the individual's behaviours, such as changing daily lifestyle according to a radical belief system. We can observe directly that radicalization behaviours and social activities play a vital role in identifying the role of an individual for the Ku Klux Klan group.

The UniMP model succeeded in the role classification task for the ISIL group, and the feature importance for the classification is shared in Figure 2. We see that *Red_ duration, Internet_ Use_ Plot* and *Education* are the top 3 most

important features. The *Rad_ duration* feature gives us information about the time period between the first time radicalization and exposure, basically how long the individual was considered as in a radical group. The *Internet_ Use_ Plot* and *Education* features indicate whether an individual used the internet for communications or logistics while preparing for a plot, and what is the highest level of education of the individual. We observe that for the ISIL group role identification, *long* duration of radicalization, *none* of the internet sources usage and *low* education are the important features of the role identification task.

As we see from Figure 2, the most two important features are *Radical_ Behaviors* and *Indict_ Arrest* features for AlQaeda group whose member is successfully classified by GraphSAGE model. We mentioned above about the kind of information the *Radical_ Behaviors* feature gives above. The *Indict_ Arrest* feature tells us whether the individual was arrested for an ideologically motivated crime as a result of their public exposure. We see that the *indicted and arrested* category for this feature gives a good amount of information for classification by the GraphSAGE model, along with the active participation of individuals in operational actions that result in casualties. We can clearly see from the results that behavioural activities are vital for identifying the role of the AlQaeda group.

Figure 2 shows that the most important feature of the UniMP model for classifying the role of individuals in the JDL group is *Degree_ Centrality*. Degree centrality measure is simply the degree of the node in a network. We can see from Table 1 that the JDL group has a max degree of 10 while the density is around 0.05. Even though there is a small number of nodes, we can see that the nodes are relatively connected and it makes sense that network structural information can yield good information about an individual's role in a group. The following two most important features are *Foreign_ Fighter* which tells us whether the individual takes actionable steps to travel abroad to join a foreign terrorist organization and *Media_ Radicalization* which is about whether the media besides the internet played a role while the individual is radicalizing.

We see that each group has its own network structure, according to the philosophies that they follow. Hence the feature that mostly affects classification performance for role identification tasks changes accordingly. In this section, we explained the most important features and analysed their effects on the performance. More details about the features can be found in [24] with explanations.

5 Conclusion

In this paper, we first present a summary of counter-terrorism, predicting a role in a terrorist group and its importance along with the node classification task in AI and ML literature. Second, we present a GNN approach to node classification in terrorism networks, which existing literature did not study well, using three state-of-the-art GNN algorithms according to different learning strategies: GCN, GraphSAGE and UniMP. We address the challenges of terrorist network data such as highly imbalanced classes. We conduct the experiments on thirteen real-world terrorist networks including one whole, three ideological, and nine

specific groups. It demonstrates that GNNs outperform the MLP algorithms on certain network topologies. The results of the study open up promising areas for future research in the community, such as extracting structural information and classifying the individuals in terms of future potential to play an active role in an attack, hidden roles of individuals in real terrorist groups which have not been identified so far, etc. Moreover, the network can be structured with more side information such as more edge features, structural node features, etc. and the model performance could be improved. Even though GNNs have been a hot topic almost for the last decade, exploring different types of GNNs could improve the model's performance.

Open Questions As recent studies on the class imbalance problem show [6,7,15], it reduces the classification performance of the majority of GNN algorithms. By the nature of counter-terrorism datasets, especially those represented as knowledge graph networks, the roles of the individuals are naturally highly imbalanced. Hence, it is still an open question how to handle imbalanced classes in terrorist networks without losing some information during the training process, as this is sensitive information in the counter-terrorism domain. Another need might be a better network representation. As it is directly connected with the learning performance of GNNs, we believe it is an interesting area to explore.

Limitations Due to the limited information and many uncovered areas in GNN literature, there is still much to explore, such as how to handle overfitting when we have large terrorism networks, and how much we can count on hidden relationships between individuals or groups. Another limitation comes from the nature of counter-terrorism, such as radicalization that occurs immediately after an unexpected political decision. The identification of this anomaly might be a hard task with the current improvements in the algorithms we discussed in this work. Our work focused on homogeneous graphs, not heterogeneous graphs that contain different types of interactions with the nodes at the same time, such as multiple connections at the same time. Adding edge features could give more information to GNNs that we did not consider in this work. It would be an interesting idea to experiment with how different heterogeneous graphs or improved datasets could affect the model's performance.

Through this work, we believe we open a new direction for counter-terrorism and the AI community to further explore, and we hope that the proposed work will contribute to counter-terrorism as a useful tool for further operations leasing to a more secure community.

Acknowledgments. This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 12/RC/2289-P2 at Insight the SFI Research Centre for Data Analytics at UCC, which is co-funded under the European Regional Development Fund. For Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bangerter, M.L., Fenza, G., Gallo, M., Loia, V., Petrone, A., Volpe, A.: Terrorist organization identification using link prediction over heterogeneous gnn. Human-centric Computing and Information Sciences **12** (2022). https://doi.org/10.22967/HCIS.2022.12.012
2. Bello, I., Pham, H., Le, Q.V., Norouzi, M., Bengio, S.: Neural combinatorial optimization with reinforcement learning. CoRR **abs/1611.09940** (2016), http://arxiv.org/abs/1611.09940
3. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research **16**, 321–357 (Jun 2002). https://doi.org/10.1613/jair.953
4. David Bright, C.W., Harris-Hogan, S.: On the durability of terrorist networks: Revealing the hidden connections between jihadist cells. Studies in Conflict & Terrorism **43**(7), 638–656 (2020). https://doi.org/10.1080/1057610X.2018.1494411
5. Hamilton, W.L., Ying, R., Leskovec, J.: Inductive representation learning on large graphs (2018)
6. Huang, Z., Tang, Y., Chen, Y.: A graph neural network-based node classification model on class-imbalanced graph data. Knowledge-Based Systems **244**, 108538 (2022). https://doi.org/https://doi.org/10.1016/j.knosys.2022.108538
7. Juan, X., Zhou, F., Wang, W., Jin, W., Tang, J., Wang, X.: Ins-gnn: Improving graph imbalance learning with self-supervision. Information Sciences **637**, 118935 (2023). https://doi.org/https://doi.org/10.1016/j.ins.2023.118935
8. Kearnes, S., McCloskey, K., Berndl, M., Pande, V., Riley, P.: Molecular graph convolutions: moving beyond fingerprints. Journal of Computer-Aided Molecular Design **30**(8), 595–608 (Aug 2016). https://doi.org/10.1007/s10822-016-9938-8
9. Kengpol, A., Neungrit, P.: A decision support methodology with risk assessment on prediction of terrorism insurgency distribution range radius and elapsing time: An empirical case study in thailand. Computers & Industrial Engineering **75**, 55–67 (2014). https://doi.org/https://doi.org/10.1016/j.cie.2014.06.003
10. Khoshraftar, S., An, A.: A survey on graph representation learning methods. ACM Trans. Intell. Syst. Technol. **15**(1) (jan 2024). https://doi.org/10.1145/3633518
11. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings. OpenReview.net (2017), https://openreview.net/forum?id=SJU4ayYgl
12. Kool, W., van Hoof, H., Welling, M.: Attention, learn to solve routing problems! (2019)
13. Lautenschlager, J., Ruvinsky, A., Warfield, I., Kettler, B.: Group profiling automation for crime and terrorism (gpact). Procedia Manufacturing **3**, 3933–3940 (2015). https://doi.org/https://doi.org/10.1016/j.promfg.2015.07.922, 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015
14. Li, Z., Sun, D.y., Guo, S.q., Li, B.: Detecting key individuals in terrorist network based on fanp model. In: 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014). pp. 724–727 (2014). https://doi.org/10.1109/ASONAM.2014.6921666
15. Maekawa, S., Noda, K., Sasaki, Y., Onizuka, M.: Beyond real-world benchmark datasets: An empirical study of node classification with gnns (2022)

16. Matsunaga, D., Suzumura, T., Takahashi, T.: Exploring graph neural networks for stock market predictions with rolling window analysis. CoRR **abs/1909.10660** (2019), http://arxiv.org/abs/1909.10660
17. Qu, L., Zhu, H., Zheng, R., Shi, Y., Yin, H.: Imgagn: Imbalanced network embedding via generative adversarial graph networks. CoRR **abs/2106.02817** (2021), https://arxiv.org/abs/2106.02817
18. Riba, P., Fischer, A., Lladós, J., Fornés, A.: Learning graph distances with message passing neural networks. In: 2018 24th International Conference on Pattern Recognition (ICPR). pp. 2239–2244 (2018). https://doi.org/10.1109/ICPR.2018.8545310
19. Saini, J.K., Bansal, D.: Computational techniques to counter terrorism: a systematic survey. Multimedia Tools Appl. **83**(1), 1189–1214 (jun 2023). https://doi.org/10.1007/s11042-023-15545-0
20. Sanchez-Gonzalez, A., Heess, N., Springenberg, J.T., Merel, J., Riedmiller, M.A., Hadsell, R., Battaglia, P.W.: Graph networks as learnable physics engines for inference and control. CoRR **abs/1806.01242** (2018), http://arxiv.org/abs/1806.01242
21. Sato, R., Yamada, M., Kashima, H.: Approximation ratios of graph neural networks for combinatorial problems. CoRR **abs/1905.10261** (2019), http://arxiv.org/abs/1905.10261
22. Selsam, D., Lamm, M., Bünz, B., Liang, P., de Moura, L., Dill, D.L.: Learning a SAT solver from single-bit supervision. CoRR **abs/1802.03685** (2018), http://arxiv.org/abs/1802.03685
23. Shi, Y., Huang, Z., Feng, S., Zhong, H., Wang, W., Sun, Y.: Masked label prediction: Unified message passing model for semi-supervised classification (2021)
24. START: National consortium for the study of terrorism and responses to terrorism (start) profiles of individual radicalization in the united states. retrieved from (http://www.start.umd.edu/pirus). (2023)
25. TE-SAT: European union terrorism situation and trend report (2023)
26. Uddin, M.I., Zada, N., Aziz, F., Saeed, Y., Zeb, A., Shah, S.A.A., Al-Khasawneh, M.A., Mahmoud, M.: Prediction of future terrorist activities using deep neural networks. Complex. **2020**, 1373087:1–1373087:16 (2020), https://api.semanticscholar.org/CorpusID:218578815
27. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks (2018)
28. Waikhom, L., Patgiri, R.: A survey of graph neural networks in various learning paradigms: methods, applications, and challenges. Artificial Intelligence Review **56** (11 2022). https://doi.org/10.1007/s10462-022-10321-2
29. Xiao, S., Wang, S., Dai, Y., Guo, W.: Graph neural networks in node classification: survey and evaluation. Mach. Vision Appl. **33**(1) (jan 2022). https://doi.org/10.1007/s00138-021-01251-0
30. Ying, R., Bourgeois, D., You, J., Zitnik, M., Leskovec, J.: GNN explainer: A tool for post-hoc explanation of graph neural networks. CoRR **abs/1903.03894** (2019), http://arxiv.org/abs/1903.03894
31. Yu, B., Yin, H., Zhu, Z.: Spatio-temporal graph convolutional neural network: A deep learning framework for traffic forecasting. CoRR **abs/1709.04875** (2017), http://arxiv.org/abs/1709.04875
32. Zheng, C., Fan, X., Wang, C., Qi, J.: Gman: A graph multi-attention network for traffic prediction (2019)