A Multi-Agent Reinforcement Learning-Based Framework for Forecasting Terrorist Collaboration and Predicting Future Alliances

Vedat Dogan*, Steven Prestwich, and Barry O'Sullivan

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. Terrorist activity has increased over the years, leading to the rise of new criminal organizations, the persistence of incidents, and increased collaboration and coordination among criminal entities. This study proposes a framework based on multi-agent reinforcement learning (MARL) to forecast terrorism collaboration dynamics from time-series data and predict future collaborations. Firstly, we retrieve data from the Global Terrorist Database for numerous countries and construct a terrorist collaboration network. Subsequently, we employ the cumulative time series data to construct cumulative temporal graphs, thereby facilitating the observation of the evolution of collaboration over time. Then, we design a reward function that quantifies the lethality of terrorist groups, the benefits of collaborations, the group's role in the network and the effectiveness of the partnership. Finally, we use the learned parameters to generate unobserved terrorist collaboration networks and, therefore, to predict the future potential collaborations for terrorist groups. The research findings demonstrate that the MARL approach exhibits superior forecasting performance in predicting terrorist collaboration networks. Future research endeavours should explore the potential of AI in countering terrorist activities.

Keywords: Forecasting Terrorist Collaboration, Multi-agent Reinforcement Learning, Counter-Terrorism, Predictive Models

1 Introduction

In recent years, terrorism has posed a significant threat to global security, manifesting in the emergence of new criminal organizations, the recurrence of terrorist incidents, and the collaboration and coordination among criminal entities. The *shared* ideological origins, behavioural patterns, and pursuit of goals among these organizations have led to a loose coalition of international criminal networks. This situation has made it more challenging to combat terrorism and heightened the risks it poses to the international security environment.

* Corresponding Author.

Terrorist acts have emerged as the predominant form of terrorism over the years. Terrorist organizations often exhibit a networked structure. Consequently, the understanding of the network of cooperation among terrorist organizations and the effective implementation of strategies to disrupt organizational alliances have garnered significant attention from scholars and security agencies worldwide. Experts have advocated for a specialized scientific discipline to analyze conflicts, civil wars, and terrorism computationally [15]. Despite this call, efforts to harness AI for these purposes have been limited. While terrorism is inherently uncertain and unpredictable, transdisciplinary research leverages comprehensive data, sophisticated computational models, and a foundational understanding of terrorist behaviour to provide data-driven solutions [21]. Considering all the above, this study aims to propose a framework based on multi-agent reinforcement learning (MARL) to forecast terrorism collaboration dynamics from time series data and predict future collaborations. For this purpose, we focus on terrorist group dynamics and shared incidents over the years. We retrieve data from the Global Terrorism Database (GDT) for numerous countries and construct a terrorism collaboration network. Subsequently, we employ annual time-series data to construct cumulative temporal graphs, thereby observing the evolution of collaboration over the years. We utilize the created time series graphs to learn collaboration dynamics, and periodic patterns between terrorist groups, such as ideological, lethality-based, etc. and related shared event features, such as tactical, operational, and ideological.

In this study, we assume each node representing a terrorist group is an agent in a reinforcement learning (RL) setting. Each agent has three actions: maintaining the current state, making or deleting a collaboration. We design a multi-objective reward function that considers multi-objective lethality, collaboration benefits, network position, and temporal improvement. We assume that the terrorist groups are willing to collaborate to increase the incidence of casualties, therefore lethality, sharing the same ideology and having the same target types, while designing the reward function. Then, we modified the algorithm to learn the reward functions of each node and the policy to learn the strategy to achieve higher rewards. By using the learned policies, we predict the future potential collaborations of terrorist groups based on the MARL simulation. The contributions of this study are three-fold:

1. We propose a novel methodology and framework for forecasting and predicting criminal collaborations by incorporating a diverse range of interpretable features derived from network science, terrorism research and AI literature.
2. We develop a multi-objective lethality-based reward function that quantifies the lethality of a terrorist group based on a wide range of specifications.
3. Also, we use advanced AI techniques in the MARL setting to encode terrorist group and collaboration features and capture network dynamics over the years to extract information for predictions.

These contributions improve the model's learning and prediction of the performance of future terrorism collaboration networks.

The rest of the paper is organized as follows. Section 2 explains the related works to both network prediction and reinforcement learning literature. The GDT dataset that is used for this study is explained in Section 6 with data preparation process and proposed node level and edge level features. Section 4 describes the creation of the cumulative

temporal terrorist collaboration networks for forecasting purposes and how to put the graphs to time-series. In Section 3, we formulate the problem definition and in Section 5 we explain the proposed MARL framework with the specifically designed components. Section 6 discuss the experiments with experimental settings and results. Finally, Section 7 is devoted to conclusions and future directions of research.

2 Related Works

In this section, we discuss related works in the network prediction and MARL literatures and highlight the differences between existing methods.

Network prediction has become a hot topic through recent advances in AI. Improvements in graph-based deep learning algorithms have especially raised interest in the field. The authors in [10] implemented the GraphSAGE algorithm, which is an improvement on graph convolutional networks (GCN) [27], to generate new node features from the network structure. Also the role classification in criminal networks is studied in [4] by embedding node and edge features with graph neural networks. The ELSM method is proposed in [9] to augment the network structure by using the node's latent variable and the network structure. The STEP algorithm is proposed in [3] to predict the network structure by using temporal and structural information on the network. The DualCast algorithm is proposed in [11] to predict network structure and attributes, and presents competitive results. The NetEvolve framework is a method that uses reinforcement learning to predict social network structure and node features, and it discuss the network types and prediction results in [18]. There are various applications that appear as link prediction tasks in the graph mining literature, and they generally focus on social networks. We refer the reader to [2,25] for more up-to-date surveys related to this area.

Reinforcement learning (RL) is a machine learning technique that enables an agent to acquire decision-making skills through interactions with an environment. The agent receives rewards for positive actions and penalties for negative ones, thereby guiding its learning process [24]. Multi-agent reinforcement learning [1] is a framework that focuses on the behaviour of multiple agents that interact with the environment in either a cooperative or independent way [19]. Several studies applied the MARL framework to graph learning tasks. For example, authors focused on link prediction in [14] by learning structural changes over time. Another study predicts the emergence of new social network structures in a MARL framework [26]. We refer to recent surveys for a more in-depth review related to this topic [8].

3 Problem Definition

In this section, we define the problem. The input terrorism collaboration network is represented as cumulative temporal graphs. Each node represents a terrorist group, and the edge between groups represents the incidents in which they collaborate. The work consist of learning the collaborations between groups over time and predict the future potential alliances using training data.

Definition 1. ***Cumulative Temporal Graphs:*** *A cumulative temporal graph $\mathcal{G}$ is a sequence of T discrete snapshots $\langle \mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_T \rangle$ where $\mathcal{G}_t$ denotes the graph at the time*

t. $\mathcal{G}_t$ is a tuple $\{\mathcal{V}_t, \mathcal{E}_t, \mathbf{X}_t, \mathbf{Y}_t\}$ where $\mathcal{V}$ is a set of vertices, $\mathcal{E}_t$ is the set of undirected edges between vertices at the time t, $\mathbf{X}_t : \mathcal{V} \rightarrow \mathbb{R}^n$ is the feature vector with the size n and $\mathbf{Y}_t : \mathcal{E} \rightarrow \mathbb{R}^k$ is the feature vector with the size k for vertices and edges in the network, respectively.

As we mention in Section 1, each node, $v_t \in \mathcal{V}_t$ represents a terrorist group in the collaboration network at the period of t in this study. An edge $e_t \in \mathcal{E}_t$ is created whenever associated terrorist groups collaborate on the same event at time t. Finally, the node and edge features, $\mathbf{X}_t$ and $\mathbf{Y}_t$, contains the terrorist group and shared event features. We treat predicting alliances as a decision-making problem in the MARL setting, as each agent has three actions to make to maximize the multi-objective reward function.

Definition 2. ***Predicting Alliances:*** *Given a set of cumulative temporal graphs $\{\mathcal{G}_i\}_{t-n}^{t}$ as training instances $\mathbb{D}$, predict the set of edges $\mathcal{E}_{t'}$ at $t' \in T$ for unseen graphs.*

4 Network Construction

In this section, we introduce a terrorist collaboration network that leverages the advantages of cumulative time-series temporal graphs. First, for each time unit t, the vertices $\mathcal{V}_t$ represent the terrorist groups which conducted attacks and edges $\mathcal{E}_t$ represent the attacks that terrorist groups collaborated on during time period t. The nodes are defined as $\mathcal{V}_t = \{v_i | v_i \in g_i\}$ where v_i represents the nodes in group g_i that is included as a node if it participated in at least one event during the period of t. The edges are defined as $\mathcal{E}_t = \{e_{i,j} | (v_i, v_j) \in \mathcal{V}_t^2 \textit{and } v_i \neq v_j\}$ where $e_{i,j}$ indicating a collaboration which is defined as participation in the same event during t.

Each group is identified with various features related to its lethality, strategic, structural information, and tactical diversity, as defined in Section 4.1. We created feature matrix $\mathbf{X}_t$ for each terrorist group, with the dimension of 19 for each time period of t. Then, we created the edge feature matrix $\mathbf{Y}_t$ for the time period of t with the dimension of 8 to capture the collaborated attack characteristics. They are defined as follows:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,F} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,F} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N,1} & x_{N,2} & \cdots & x_{N,F} \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} \mathbf{y}_{1,2}^\top \\ \mathbf{y}_{1,3}^\top \\ \vdots \\ \mathbf{y}_{i,j}^\top \\ \vdots \\ \mathbf{y}_{N-1,N}^\top \end{bmatrix} = \begin{bmatrix} y_{1,2,1} & \cdots & y_{1,2,K} \\ y_{1,3,1} & \cdots & y_{1,3,K} \\ \vdots & \ddots & \vdots \\ y_{i,j,1} & \cdots & y_{i,j,K} \\ \vdots & \ddots & \vdots \\ y_{M-1,M,1} & \cdots & y_{M-1,M,K} \end{bmatrix} \tag{1}$$

where N represents the number of nodes and F represents the number of node features, which is 19 in this study. Also M and K represent the total number of edges and features, respectively. We created features to capture the characteristic information with various aspects of terrorist groups and correlate it with the future collaboration. The features capture collaboration details, tactical and operational similarity of collaborated groups and the lethality of the shared attacks. Finally, we define the cumulative temporal network $\mathcal{G}_t$ as $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t, \mathbf{X}, \mathbf{Y})$ where t represents the defined time period.

In this study, we consider the annual-based time series. For each selected country, the time-series of graphs is created starting from the first date that collaboration has

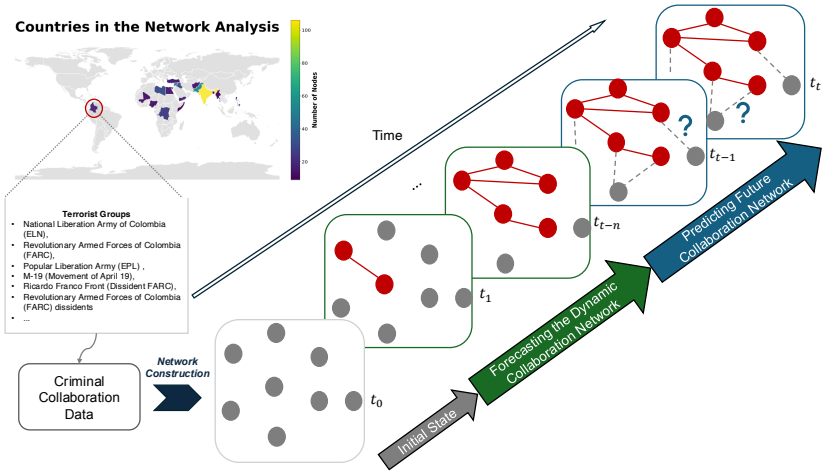

Fig. 1. Illustration for the network construction from tabular data.

appeared until the activity date in the dataset. The nodes and edges in the graphs are considered in a cumulative way to keep track of the historical evolution and temporal networks to capture the existing characteristics. Therefore, each cumulative time-series temporal graph $\mathcal{G}_t$ can be defined as $\mathcal{G}_t = \{\mathcal{G}|\mathcal{G}_{t-1} \in \mathcal{G}_t, 1 \leq t \leq T\}$ where t is a time stamp. For instance, an illustration of the collaboration network graphs is given in Figure 1. The illustrative network belongs to a country and has, presumably, 8 total groups that are collaborated until the end of the period, and the first collaboration appears in t_0. As the last data period is t_t, we use the $t-n+1$ network to train and n networks to test where n is 30% of the total number of graph networks. As we can see from the Figure, the network is evolving over the years, and the proposed algorithm aims to predict the future collaboration by learning from the time-series graphs with MARL.

4.1 Proposed Features

To capture the specifications of terrorist groups and events related to collaboration, we needed to develop new feature metrics that would specifically address the *lethality metrics*, *network structure*, *activity metrics*, *tactical diversity* and *temporal collaborations* on the terrorist group level. On the event level, we created *basic collaboration metrics*, *tactical similarity*, *operational similarity* and *combined lethality metrics*. We explained node level features as follows:

- The first feature is the total collaborations of the group so far, represented with C_i.
- The second feature is the proposed multi-objective lethality , and it considers multiple aspects of lethality for the related group. The lethality metric has four metrics:
  - The first lethality metric is *casualty_impact*, which represents the casualties caused by the group and contains a number of *kills* that the group has and the number of *wounded* people.

  - The second lethality metric is *operational_efficiency*, which represents the success rate and survival ratio of the group.
  - The third one is *tactical_sophistication*, which we define as diversity in weapons, targets, tactics and use of suicide or multiple attacks.
  - The last metric is *strategic_impact*, representing the property damage, attack frequency and average kills per attack. We aim to capture the lethality of the group accurately with a detailed multi-objective structure.
- We used the statistical network metrics for each group as a third node level feature. We used three metrics to represent the network structure which are *degree centrality*, *clustering coefficient*, and *betweenness centrality*.
- We used the *activity metric* to capture the tactical sophistication, and we identify it with four features: *Total attacks*, *average casualties success rate* and *property damage rate* represents as it means.
- We define *tactical diversity* with two features, *weapon diversity* and *target diversity*, which represent the types of weapons and types of targets attacked, respectively.

In terms of edge level features, we define the *basic collaboration metrics* with two created features, *weights* and *average casualties* in the shared attacks. *Tactical similarity* features are *attack similarity, weapon similarity*, and *target similarity* of the groups that collaborate on attacks. We define *operational similarity* with two features, which are *geographic overlap* and *temporal overlap*, representing the fraction of overlapping regions and overlapping active time periods for the groups, respectively. Lastly, *combined lethality* is the aggregated lethality metric of shared attacks.

5 Proposed Method

This section describes the proposed method for predicting the future terrorist collaboration network based on MARL using cumulative temporal criminal networks. The illustration of the proposed method for forecasting criminal collaborations is shared on Figure 2. We discuss the environment for graph networks in Section 5.1, and reward and policy function for the agents in Section 5.2 and Section 5.3, respectively.

5.1 Environment

We designed the environment of reinforcement learning (RL) as the Markov Decision Process (MDP), which contains the state $\mathcal{S}_t$, action $\mathcal{A}_t$ and reward $\mathcal{R}_t$. In MDPs, the agents make decisions based on the current state by utilizing a policy function $\pi(\mathcal{A}_t|\mathcal{S}_t, \theta)$ where θ is the parameter of the policy function. After the agents make the actions, the state changes accordingly, and the transition function is defined as $f(\mathcal{S}_{t+1}|\mathcal{S}_t, \mathcal{A}_t)$. The rewards are calculated based on the defined function $r(\mathcal{S}_t|\Psi)$ where Ψ is the parameters of the function. In RL, the objective is to learn the parameter for the defined policy that maximizes the expected reward for the agent's actions.

In this study, we exploit the RL setting to learn from time-series graph networks and we defined the states as the current terrorist collaboration network $\mathcal{G}_t$ and the actions as the change in the graph and defined as $\Delta\mathcal{G}_t = (\Delta\mathcal{E}_t|\mathbf{X}_t, \mathbf{Y}_t)$. The $\Delta\mathcal{E}_t$ represents the

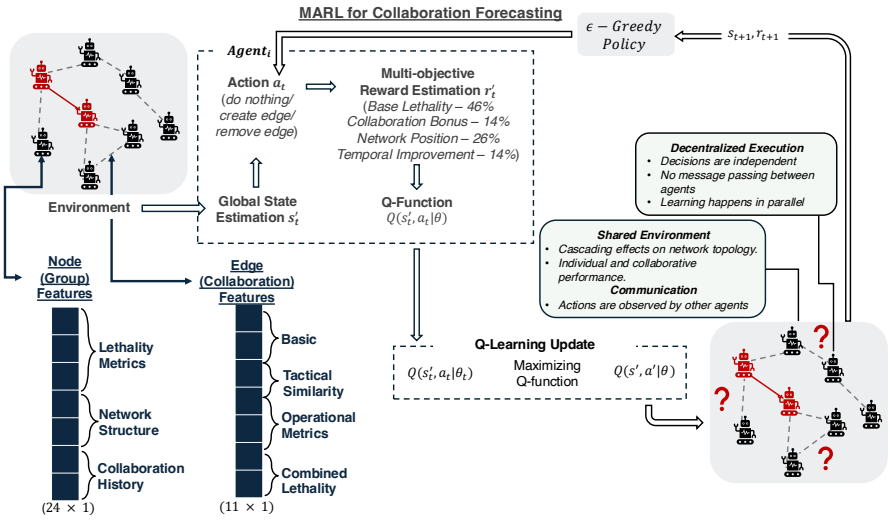

Fig. 2. Framework of the MARL approach for forecasting criminal collaboration.

edge changes in the graph at the period of t and $\mathbf{X}_t, \mathbf{Y}_t$ represents the node and edge attributes, respectively. Therefore, the reward and the policy function can be defined as $r(\mathcal{G}_t|\Psi)$ and $\pi(\Delta\mathcal{G}_t|\mathcal{G}_t, \theta)$, respectively. Now, we can define the state transition for the network as $f(\mathcal{G}_{t+1}|\mathcal{G}_t, \mathcal{A}_t) = \{\mathcal{V}_t, \mathcal{E}_t \cup \Delta\mathcal{E}_t, \mathbf{X}, \mathbf{Y}\}$.

In the MARL setting, we assume each agent acts as a node v_i in a network. The reward and policy functions for each agent are defined as $r_i(\mathcal{G}_t|\Psi_i)$ and $\pi_i(\Delta\mathcal{G}_t|\mathcal{G}_t, \theta_i)$ respectively. In this study, we assume the reward function for the overall graph is the summation of each node's reward and product of each node's policy, defined as;

$$r(\mathcal{G}_t|\Psi) = \sum_{v_i \in \mathcal{V}} r_i(\mathcal{G}_t|\Psi_i) \quad , \pi(\Delta\mathcal{G}_t|\mathcal{G}_t, \theta) = \prod_{v_i \in \mathcal{V}} \pi_i(\Delta\mathcal{G}_t|\mathcal{G}_t, \theta_i) \tag{2}$$

where $\Psi = \{\psi_i\}_{v_i \in \mathcal{V}}$ and $\theta = \{\theta_i\}_{v_i \in \mathcal{V}}$ are the reward and policy function's parameter, respectively.

5.2 Reward Function

This section describes the design of the reward function for each agent (*node*) for a given terrorist collaboration network. We design the reward function to measure the lethality of a terrorist group, the benefits of collaborations, the group's role in the network and the effectiveness of the partnership.

We define the reward function in a terrorism collaboration network $\mathcal{G}_t$ at the time of t for each agent (node) as follows:

$$r_i(\mathcal{G}_t|\Psi_i) = w_1 L_{i,t} + w_2 \Delta L_{i,t} + w_3 P_{i,t} + w_4 C_{i,t} \tag{3}$$

where $L_{i,t}$, $\Delta L_{i,t}$, $P_{i,t}$ and $C_{i,t}$ are base lethality score of v_i, improvement in lethality due to new collaborations, network position reward based on centrality and clustering and collaboration reward by motivating effective partnerships. The components of multi-objective reward functions are defined as follows. *The base lethality score* $L_{i,t}$ captures the operational effectiveness of a terrorist group at the time period of t, which is defined as $L_{i,t} = w_c C_i + w_o O_i + w_t T_i + w_s S_i$ where C_i, O_i, T_i and S_i are causality impact, operational efficiency, tactical sophistication and strategic impact, respectively. *Lethality improvement*, $\Delta L_{i,t}$, encourages the formation of mutually beneficial partnerships and defined as $\Delta L_{i,t} = \max(0, L_{i,j,t} - (L_{i,t} + L_{i,j}))$ where $L_{i,j,t}$ is combined lethality of the terrorist groups i and j after collaboration. *Network position reward* encourages the agents to improve their network roles and is defined as;

$$P_{i,t} = \alpha \times DegreeCentrality_{i,t} + \beta \times ClusteringCoefficient_{i,t} \tag{4}$$

Collaboration reward, $C_{i,t}$ gives the reward for successful collaborations and defined as follows:

$$C_{i,t} = \lambda \times SuccessRate_{i,j,t} \tag{5}$$

The rewards defined above motivate agents to improve their lethality, forge mutually beneficial partnerships, and enhance their network's position. Moreover, the reward function incorporates weighted components to ensure that multiple objectives are balanced. We optimized the weights of the components, and shared in Table 1. The weights are the hyperparameters for the MARL approach, and the details of the optimization are shared in Section 6. Additionally, the design of a multi-objective reward function guarantees that agents act in a way that collectively enhances the network's effectiveness, considering both individual node dynamics and network-wide dynamics.

Table 1. The optimized weights (best) of the multi-objective reward function.

Reward Component	**Weight (%)**
Base Lethality	46%
Collaboration Reward	14%
Network Position	26%
Temporal Improvement	14%

5.3 Policy

We describe the deep Q-network based policy in this section. We design the policy function by assuming that each agent (node v_i) can take three actions: *maintain the position, make an edge* and *delete an edge*.

The policy is derived from the Q-value function, $\mathcal{Q}(s_i, a_i, ; \theta)$, which represents the expected cumulative reward when an agent takes action a in the state s. It is defined as:

$$\mathcal{Q}(s_i, a_i; \theta) = \mathbb{E}\pi\left[\sum_{t=0}^{T-1} \gamma^t R_{i,t} | s_{i,t} = s_i, a_{i,t} = a_i\right] \tag{6}$$

where $R_{i,t}$, γ and θ are the rewards at time t, a discount factor ($0 < \gamma \leq 1$) that weights future rewards, and the parameters of the Q-network, respectively. So, we define the

policy function as $\pi(s_i) = \arg\max Q(s_i, a_i; \theta)$ with deterministic policy. According to that, the action a_i with the highest Q-value is chosen. During the training policy, we used ϵ-greedy policy for exploration, which is defined as;

$$\pi_\epsilon(s_i) = \begin{cases} \text{random action } \textit{with probability } \epsilon \\ \arg\max Q(s_i, a_i; \theta) \textit{ with probability } 1 - \epsilon \end{cases} \tag{7}$$

where ϵ is the exploration rate and it decays over time to balance exploration and exploitation. To approximate the $Q(s_i, a_i; \theta)$, the Q-network takes the state vector, which has 24 dimensions in this study (*16 historical metrics, 4 network features and 4 collaboration history of the group*), extracts the features with hidden layers and gives the Q-value for each action $a_i \in \mathcal{A}_i$. It is defined as $Q_\theta(s_i, a_i) = f_\theta(g_\phi(s_i))$ where g_ϕ is feature encoder vector, f_θ is Q-network and ϕ and θ are network parameters. The network is trained using the Bellman equation, defined as $Q_i(s_i, a_i; \theta) = R_i(s_i, a_i) + \gamma \max_{a'_i} Q_i(s'_i, a'_i; \theta^-)$ where $R(s_i, a_i)$ is immediate reward, s' is the following state and θ^- is the parameters of the target Q-network. The loss function is the mean squared error between the predicted Q-values and target Q-values, and defined as:

$$L(\theta) = \mathbb{E}_{s_i, a_i, r_i, s'_i}\left[(y - Q(s_i, a_i; \theta))^2\right] \tag{8}$$

where the target defined as $y = R(s_i, a_i) + \gamma \max_{a'_i} Q(s'_i, a'_i; \theta^-)$. The experience replay is used in a way that transitions (s_i, a_i, r_i, s'_i) for the terrorist group (the agent v_i) in a replay buffer and batches are sampled randomly for training to break correlation between consecutive experiences. The target Q-network parameter θ^- is updated periodically to stabilize training, and gradients of the loss are computed and used to update the parameters θ, defined as $\theta \leftarrow \theta - \eta \Delta_\theta L(\theta)$ where η is the learning rate. In multi-agent extension, each agent maintains its own Q-network, meaning that each terrorist group (as a node in the network) has their own feature embeddings. Each state, action and reward of the terrorist groups can be defined as $\mathcal{S} = \{s_1, \ldots, s_N\}$ and $\mathcal{A} = \{a_1, \ldots, a_N\}$ where N is the number of terrorist groups, meaning agents, and each agent optimizes its own Q-function with the Bellman equation. The learned parameters are used to generate unobserved terrorist collaboration networks and, therefore, to predict the future collaboration of terrorist groups.

6 Experiments

We evaluated our proposed approach to capture future collaborations on real-world criminal groups. We use the GTD dataset to filter the terrorist attacks in the top countries where the terrorist groups collaborated the most. These countries are *India, Pakistan, Afghanistan, Colombia, Philippines, the Democratic Republic of the Congo, Iraq, Israel, Turkey, West Bank and Gaza Strip, Bangladesh, Myanmar, USA* and *Egypt*. For every dataset, we consider the following settings. We selected the beginning of the time period, t_0 of as the first date of the first collaboration appears, and the end of time period, t_k is the last attack that occurred according to the dataset. Then, we created the cumulative temporal graph networks as described in Section 4. The date of the first

collaboration appears is different for each country as we can see in Figure 4. So, it affected the selection of the training and testing time periods and the amount of years accordingly. For the implementation, the datasets are available online[1].

Dataset The data is retrieved from the GTD, maintained by the START research centre at the University of Maryland [13]. The GTD is an event-level database with over 200,000 records of terrorist attacks worldwide for the dates between 1970 and 2020. It is the world's most comprehensive and detailed open-access dataset on terrorist events. START releases an updated version annually. To be included, an event must meet specific criteria divided into two levels. The first level includes three essential aspects: intentionality and violence level and the sub-national nature of terrorist actors. The second level has three criteria, but at least two must be met. These relate to political, economic, religious, or social objectives, coercion or intimidation, and legitimate warfare activities. An event meets the criteria and is included, but there is a control mechanism for conflicting information or acts not exclusively of a terrorist nature. Each event is associated with various variables, including geographic and temporal information, event characteristics, and economic damages, and attack perpetrators' identities.

After retrieving the data, we analyse the incidents with the most collaboration and the countries that occurred. We can see the top 20 countries that contain the highest number of terrorist groups and collaborative events in Figure 3. Each event has three group features that represent the names of connected terrorist groups, *gname, gname2* and *gname3*. We filled the missing values with 'unknown'. After handling other missing numerical and categorical values, we scaled the features and made the necessary normalization. To create the collaboration network and extract the useful information from data, we propose node and edge level features, and present them in Section 4.1.

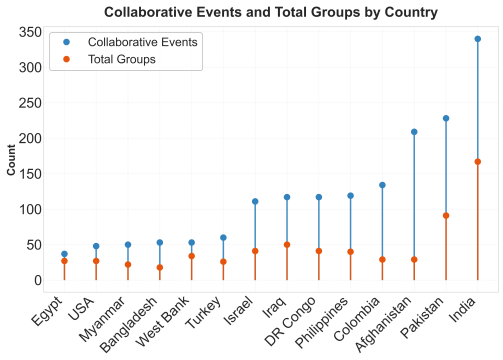

Fig. 3. Total terrorist groups in each country at the end of the time period—nodes in the network—and the number of events that these terrorist groups are collaborated—edges in the network—for each country.

Experimental Setting In the experiments, we manually set the parameters of the reward and the policy functions, and then we observe how the proposed algorithm updates it. We expect that the proposed algorithm updates the parameters such that the

[1] https://www.start.umd.edu/data-tools/GTD

future collaboration network to observe potential alliances of a terrorist group. The node counts collaboration networks is different for each country. Also, we set the node and edge features for model to learn terrorist group and related attack characteristics to measure the components to built the reward function, as described in Section 5.2. We shared all formulations and descriptions of the features in Section 2 in supplementary material. For instance, in Iraq, there are 50 total groups that are collaborated and 117 recorded collaborated events between 2003 and 2021. Therefore, for the forecasting collaboration network for Iraq, there are 50 nodes and 117 edges at the end of the period. We compare the quality of the predictions of the algorithms using the area under the ROC curve (ROC-AUC) [6] and prediction accuracy. We run our experiments on a machine with an Apple M4 Pro chip with a 12-core CPU, 16-core GPU and 16-core Neural Engine computer.

Train/Test Split: We evaluate the methods by running 20 times pre train/test split and we report the average results with standard deviation. For each dataset, we selected the annual time periods and we used the 70% of the time segments as training, 30% of the time segments for testing the proposed framework's prediction performances. We shared the splits for each country in Figure 4. Train/test splits are applied for all countries are listed seperately starting from the date of the first collaboration.

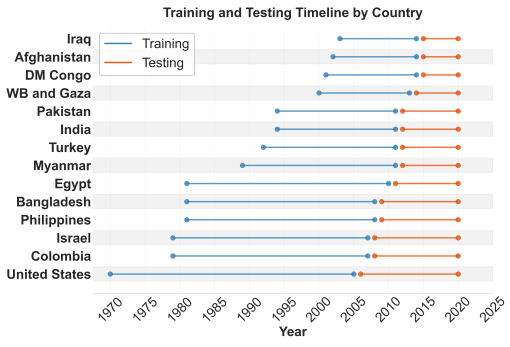

Fig. 4. Train-test (70%/30%) split of the data for each country. The first snapshot of the criminal group is considered when the first collaboration appears between groups.

Baselines: We consider different approaches, including traditional and modern machine learning algorithms proposed for graph network forecasting. We choose the logistic regression and random forest to present the performance if the data is tabular, and these methods are successful methods to predict linear and non-linear relationships. The temporal graph convolutional network (TempGCN) will present the performance with graph embeddings and compare our result with the graph neural network-based model.

– **Random Forest (RF) [5]** is a successful technique that is used for various purposes including forecasting on temporal data [20,16]. It creates multiple decision trees, and each tree votes on collaboration likelihood. For the final prediction, it combines the votes and uses feature randomization for a better generalization. The main advantage of RF is it captures the complex, non-linear relationships.

– **Logistic Regression (LR) [7]** is a widely used common technique for forecasting in the literature [17]. LR uses a linear combination of the explained and proposed features in Section 4 and applies a sigmoid function to estimate the probability of collaboration. The data is used as tabular data, and the problem is assumed as a binary classification problem as the algorithm decides if there is a collaboration.
– **TempGCN [27]** combines the graph convolutional network (GCN) and gated recurrent unit. GCN is used to capture the topological structure of the network, and a gated recurrent unit is used to capture the dynamic change of the graph. It has been used to forecast numerous domain in the literature [23,22].

Hyperparameters: We use the deep Q-network based agents and the feature encoder network is 3-layer MLP, set as "State_dim $\rightarrow$ **MLP**(64) $\rightarrow$ *ReLu* $\rightarrow$ **MLP**(32) $\rightarrow$ *ReLu* $\rightarrow$ **MLP**(3)" where **MLP**(n) means a fully-connected layer with output size of n, and *ReLu* means Rectified Linear Units. We set *memory buffer size* as $10,000$ and the learning rate as 0.001. We used ϵ-greedy policy for exploration and set epsilon decay to 0.995 with a minimum epsilon value of 0.01. For the reward function components, the weights are the hyperparameters of the MARL approach, and we aim to increase the prediction accuracy by optimizing the weights with the Bayesian optimization (BO) [12] algorithm, which is successful hyperparameter optimization (HPO) technique. For the HPO, we created a simulation dataset with random country's networks and applied BO over these datasets. We shared the distribution density and space exploration on Figure 5. We can see from the figure how different weights affect the accuracy, and we select the best weight configuration for our experiments.

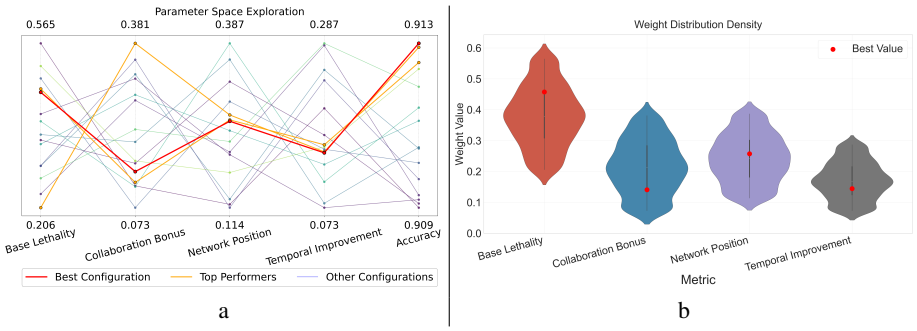

Fig. 5. The search space exploration for the weights of the reward components (a) and the final weight distribution density (b) of the BO for HPO process.

6.1 Predicting Collaboration Accuracy

Table 2 shows the prediction accuracy interms of ROC AUC and prediction accuracy scores. Across all countries, the MARL approach reaches 0.853 and 0.879 accuracy and ROC AUC scores on average. Results also show that the proposed MARL approach outperforms the competing approaches to predict collaboration in criminal networks for all

Table 2. ROC AUC scores and accuracies of criminal network forecasting across the baseline. The highest and second-highest values for each column are in bold and underlined, respectively. We performed a paired t-test comparing the best model against the others (markers ** and * indicate p-value $< .01$ and $< .05$, respectively).

Method	Metric	India	Pakistan	Afghanistan	Colombia	Philippines	Iraq	DR Congo
TempGCN	Accuracy	$0.769^{**} \pm 0.043$	$0.770^{**} \pm 0.120$	$\underline{0.769}^{*} \pm 0.091$	$0.792^{**} \pm 0.046$	$\underline{0.801} \pm 0.025$	$0.802^{**} \pm 0.037$	$0.797^{*} \pm 0.124$
	ROC AUC	$\underline{0.840}^{**} \pm 0.023$	$\underline{0.830} \pm 0.072$	$\underline{0.819} \pm 0.080$	$0.808^{**} \pm 0.027$	$0.845^{*} \pm 0.026$	$0.756^{**} \pm 0.045$	$\underline{0.841} \pm 0.092$
RF	Accuracy	$\underline{0.834}^{**} \pm 0.001$	$\underline{0.818}^{**} \pm 0.013$	$0.759^{**} \pm 0.013$	$\underline{0.837}^{*} \pm 0.001$	$\mathbf{0.802} \pm 0.021$	$0.813^{**} \pm 0.003$	$\mathbf{0.846} \pm 0.014$
	ROC AUC	$0.825^{**} \pm 0.003$	$0.796^{**} \pm 0.021$	$0.755^{**} \pm 0.001$	$\underline{0.827}^{**} \pm 0.002$	$0.769^{**} \pm 0.018$	$\underline{0.786}^{**} \pm 0.008$	$0.782^{**} \pm 0.027$
LR	Accuracy	$0.789^{**} \pm 0.023$	$0.752^{**} \pm 0.091$	$0.751^{**} \pm 0.001$	$0.802^{**} \pm 0.003$	$0.772^{*} \pm 0.061$	$\underline{0.825}^{*} \pm 0.043$	$0.801^{**} \pm 0.144$
	ROC AUC	$0.756^{**} \pm 0.035$	$0.783^{**} \pm 0.156$	$0.757^{**} \pm 0.014$	$0.769^{**} \pm 0.009$	0.804 ± 0.099	$0.754^{**} \pm 0.071$	$0.778^{*} \pm 0.176$
MARL (Proposed)	Accuracy	$\mathbf{0.916} \pm 0.044$	$\mathbf{0.883} \pm 0.120$	$\mathbf{0.821} \pm 0.091$	$\mathbf{0.854} \pm 0.047$	$0.758^{**} \pm 0.025$	$\mathbf{0.877} \pm 0.038$	0.852 ± 0.125
	ROC AUC	$\mathbf{0.932} \pm 0.024$	$\mathbf{0.913} \pm 0.072$	$\mathbf{0.871} \pm 0.080$	$\mathbf{0.895} \pm 0.027$	$\mathbf{0.848} \pm 0.029$	$\mathbf{0.909} \pm 0.046$	$\mathbf{0.894} \pm 0.092$

		Israel	Turkey	Bangladesh	West Bank	Myanmar	US	Egypt
TempGCN	Accuracy	$0.828^{*} \pm 0.051$	$\mathbf{0.849} \pm 0.025$	$0.773^{**} \pm 0.115$	$0.793^{**} \pm 0.049$	$0.805^{*} \pm 0.044$	$0.796^{**} \pm 0.042$	$0.790^{*} \pm 0.077$
	ROC AUC	$0.774^{**} \pm 0.048$	$0.810^{**} \pm 0.026$	0.759 ± 0.080	$0.826^{**} \pm 0.016$	0.820 ± 0.064	$0.796^{**} \pm 0.012$	0.829 ± 0.082
RF	Accuracy	$\underline{0.839}^{**} \pm 0.001$	$\underline{0.835}^{*} \pm 0.005$	$\mathbf{0.812} \pm 0.001$	$\underline{0.844} \pm 0.001$	$\mathbf{0.832} \pm 0.002$	$0.753^{**} \pm 0.005$	$\underline{0.803}^{*} \pm 0.007$
	ROC AUC	$\underline{0.831}^{**} \pm 0.002$	$\underline{0.835}^{**} \pm 0.009$	$0.778^{**} \pm 0.002$	$\underline{0.848}^{**} \pm 0.001$	0.820 ± 0.004	$0.738^{**} \pm 0.001$	$0.784^{**} \pm 0.009$
LR	Accuracy	$0.777^{**} \pm 0.015$	0.811 ± 0.192	$\underline{0.790}^{**} \pm 0.008$	$0.755^{**} \pm 0.001$	$\underline{0.831}^{**} \pm 0.002$	$0.741^{**} \pm 0.002$	$\mathbf{0.810} \pm 0.022$
	ROC AUC	$0.779^{**} \pm 0.016$	$0.782^{**} \pm 0.102$	$0.757^{**} \pm 0.007$	$0.750^{**} \pm 0.002$	0.820 ± 0.006	$0.750^{**} \pm 0.000$	0.823 ± 0.013
MARL (Proposed)	Accuracy	$\mathbf{0.861} \pm 0.052$	$0.801^{**} \pm 0.025$	$0.770^{**} \pm 0.116$	$\mathbf{0.852} \pm 0.050$	$0.765^{**} \pm 0.044$	$\mathbf{0.882} \pm 0.042$	$0.759^{*} \pm 0.078$
	ROC AUC	$\mathbf{0.897} \pm 0.048$	$\mathbf{0.864} \pm 0.026$	$\mathbf{0.834} \pm 0.080$	$\mathbf{0.885} \pm 0.016$	$\mathbf{0.848} \pm 0.065$	$\mathbf{0.911} \pm 0.012$	$\mathbf{0.847} \pm 0.083$

countries in terms of ROC AUC score. In particular, MARL outperforms TempGCN—the best baseline—by 21%, 10.95%, 10% for predicting alliances of terrorist groups in the US, India and Pakistan in terms of ROC AUC score. Also, MARL improves TempGCN model prediction accuracy by 10.8% in terms of accuracy. The countries which has the highest number fo groups and collaborations are Afghanistan, Pakistan and India. Across all countries, the MARL approach improves the best baseline about 7.7% and 3.8% in terms of ROC AUC and accuracy scores, respectively.

Table 3 shows the performance improvements of the best performed model comparing with the second best one for each country. We can see that RF is the best model for the networks in Bangladesh, Myanmar, and the Philippines, TempGCN for Turkey and LR for Egypt. Myanmar, Bangladesh, and Egypt have relatively small node and edge numbers in the criminal network that they have as they are in the smallest 4 networks group in Figure 3. We believe the good performance of RF and LR compared with graph-structured models is based on the small size of the data for these countries, as the TempGCN needs a graph structure to capture the dynamics, and MARL needs to train more episodes for small and limited

Table 3. Performance improvements (%) of the best model over the second-best model.

Country	Accuracy			ROC AUC		
	%Imp	Best	2nd	%Imp	Best	2nd
US	10.80	MARL	TempGCN	21.47	MARL	TempGCN
Iraq	6.30	MARL	LR	15.65	MARL	RF
India	9.83	MARL	RF	10.95	MARL	TempGCN
Pakistan	7.95	MARL	RF	10.00	MARL	TempGCN
Afghanistan	6.76	MARL	TempGCN	6.35	MARL	TempGCN
Israel	2.62	MARL	RF	7.94	MARL	RF
Colombia	2.03	MARL	RF	8.22	MARL	RF
Bangladesh	2.78	RF	LR	7.20	MARL	RF
DR Congo	0.71	MARL	RF	6.30	MARL	TempGCN
Turkey	1.68	TempGCN	RF	3.47	MARL	RF
West Bank	0.47	MARL	RF	4.36	MARL	RF
Myanmar	0.12	RF	LR	3.41	MARL	TempGCN
Egypt	0.87	LR	RF	2.17	MARL	TempGCN
Philippines	0.12	RF	TempGCN	0.36	MARL	TempGCN

data. Other than these countries, we can see from Table 3, our proposed MARL algorithm improved the performance for both metrics, starting from 0.41% to 16.6% for the worst and best performance, respectively. The experiments can be extended to the desired country in the same setting, and future collaborations can be predicted. As each agent is a terrorist group, also only a terrorist group's alliance preference can be observed by testing single agent's prediction.

7 Conclusion

This paper focused on predicting collaborations in a dynamic criminal networks. We proposed a novel framework based on MARL to forecast criminal collaborations and predict future alliances. We also propose components to design a reward function for criminal networks based on characteristic details and event information. The MARL-based framework with the specifically designed reward and policy functions explicitly models the criminal collaboration networks to forecast and predict future potential partnerships. Using the learned parameters, the model can be derived for other networks for other countries, groups and territories throughout a series of event observations. We compared our results against multiple baselines. Our approach outperformed the baselines in terms of ROC AUC score and prediction accuracy. Experiments show that the proposed algorithm can learn the group and event characteristics with various aspect, and use them to forecast criminal network collaborations and predict future networks. It also can be used for collaboration network reshaping in the case of long term inactivity of a specific criminal group. The proposed algorithm fills a gap in the AI and counter-terrorism literature and accurately predicts the behaviour of real world criminal networks. For future work, we aim to increase explainability, use designed reward and policy functions to reshape the networks in the case of emergency and observe collaborative and competitive agents behaviour along with the model performance.

References

1. Albrecht, S.V., Christianos, F., Schäfer, L.: Multi-Agent Reinforcement Learning: Foundations and Modern Approaches. MIT Press (2024)
2. Arrar, D., Kamel, N., Lakhfif, A.: A comprehensive survey of link prediction methods. J. Supercomput. **80**(3), 3902–3942 (Sep 2023)
3. Chen, H., Li, J.: Exploiting structural and temporal evolution in dynamic link prediction. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. p. 427–436. CIKM '18, Association for Computing Machinery, New York, NY, USA (2018)
4. Dogan, V., Prestwich, S.: Graph neural network-based role classification in criminal networks. In: The International Conference on Computational Science and Computational Intelligence (CSCI) Research Track on Cyber Warfare, Cyber Defense & Cyber Security. Springer Nature Switzerland, Cham (2024)
5. Dudek, G.: A comprehensive study of random forest for short-term load forecasting. Energies **15**(20) (2022)
6. Fawcett, T.: An introduction to roc analysis. Pattern Recognition Letters **27**(8), 861–874 (2006), rOC Analysis in Pattern Recognition

7. Feng, R., Wang, J., Wu, W., Liu, S., Liu, A., Xie, S.: Saturated load forecasting based on improved logistic regression and affinity propagation. Electric Power Systems Research **237**, 110953 (2024)
8. Gronauer S., D.K.: Multi-agent deep reinforcement learning: a survey. Artif Intell Rev 55, 895–943 (2022)
9. Gupta, S., Sharma, G., Dukkipati, A.: A generative model for dynamic networks with applications. Proceedings of the AAAI Conference on Artificial Intelligence **33**(01), 7842–7849 (Jul 2019)
10. Hamilton, W.L., Ying, R., Leskovec, J.: Inductive representation learning on large graphs. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. p. 1025–1035. NIPS'17, Curran Associates Inc., Red Hook, NY, USA (2017)
11. Ito, H., Faloutsos, C.: Dualcast: Friendship-preference co-evolution forecasting for attributed networks. In: SIAM (04 2022)
12. Jones, D., Schonlau, M., Welch, W.: Efficient global optimization of expensive black-box functions. Journal of Global Optimization **13**, 455–492 (12 1998). https://doi.org/10.1023/A:1008306431147
13. LaFree, Gary, L.D.: Introducing the global terrorism database. Terrorism and Political Violence 19 (May): 181-204 (2007)
14. Lim, M., Abdullah, A., Jhanjhi, N., Khurram Khan, M., Supramaniam, M.: Link prediction in time-evolving criminal network with deep reinforcement learning technique. IEEE Access **7**, 184797–184807 (2019)
15. McKendrick, K.: Artificial intelligence prediction and counterterrorism. London: The Royal Institute of International Affairs-Chatham House **9** (2019)
16. Meher, B.K., Singh, M., Birau, R., Anand, A.: Forecasting stock prices of fintech companies of india using random forest with high-frequency data. Journal of Open Innovation: Technology, Market, and Complexity **10**(1), 100180 (2024)
17. Mittal, H.K., Dalal, P., Garg, P., Joon, R.: Forecasting pollution trends: Comparing linear, logistic regression, and neural networks. In: 2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP). pp. 411–419 (2024)
18. Miyake, K., Ito, H., Faloutsos, C., Matsumoto, H., Morishima, A.: Netevolve: Social network forecasting using multi-agent reinforcement learning with interpretable features. In: Proceedings of the ACM Web Conference 2024. p. 2542–2551. WWW '24, Association for Computing Machinery, New York, NY, USA (2024)
19. Ning, Z., Xie, L.: A survey on multi-agent reinforcement learning and its application. Journal of Automation and Intelligence **3**(2), 73–91 (2024)
20. Olcay, K., Gíray Tunca, S., Aríf Özgür, M.: Forecasting and performance analysis of energy production in solar power plants using long short-term memory (lstm) and random forest models. IEEE Access **12**, 103299–103312 (2024)
21. Schiermeier, Q.: Attempts to predict terrorist attacks hit limits. Nature 517, no. 7535 (2015)
22. Sun, C., Ning, Y., Shen, D., Nie, T.: Graph neural network-based short-term load forecasting with temporal convolution. Data Science and Engineering **9**, 1–20 (11 2023)
23. Sun, L., Liu, M., Liu, G., Chen, X., Yu, X.: Fd-tgcn: Fast and dynamic temporal graph convolution network for traffic flow prediction. Information Fusion **106**, 102291 (2024)
24. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. The MIT Press, second edn. (2018)
25. Xia, F., Sun, K., Yu, S., Aziz, A., Wan, L., Pan, S., Liu, H.: Graph learning: A survey. IEEE Transactions on Artificial Intelligence **2**(2), 109–127 (2021)
26. Yu, L., Song, J., Ermon, S.: Multi-agent adversarial inverse reinforcement learning. ArXiv **abs/1907.13220** (2019)
27. Zhao, L., Song, Y., Deng, M., Li, H.: Temporal graph convolutional network for urban traffic flow prediction method. CoRR **abs/1811.05320** (2018)