Graph-Based Reinforcement Learning for Minimizing Population Mortality in Epidemic Networks

Zhihao Dong, *Student Member, IEEE,* Yuanzhu Chen, *Senior Member, IEEE,* Somayeh Kafaie, Qiao Kang, Cheng Li, *Senior Member, IEEE*

***Abstract*—The spread of infectious diseases in networked populations poses significant challenges for public health intervention strategies. Traditional centrality-based and heuristic network dismantling approaches prioritize highly connected nodes but often fail to account for individual mortality risk, limiting their effectiveness in minimizing overall fatalities. While recent advances in machine learning have improved intervention strategies, existing models largely focus on reducing disease transmission rather than directly targeting mortality outcomes. To address this gap, we propose a reinforcement learning based framework that integrates graph representation learning to identify and remove high-risk nodes, thereby maximizing network fragmentation while minimizing overall deaths. The framework is trained using synthetic networks and evaluated on five synthetic and four real-world datasets, benchmarking its performance against state-of-the-art network dismantling methods (GDM, GND, and GERL). Experimental results demonstrate that the proposed method consistently outperforms baseline approaches, particularly in scale-free and community-structured networks, where targeted node removal significantly weakens network connectivity and suppresses epidemic spread. Moreover, in real-world networks, the method achieves lower cumulative death rates and higher epidemic thresholds, highlighting its robustness in controlling disease propagation. By incorporating mortality risk into network representation learning, the proposed framework offers a scalable, adaptive, and socially responsible approach to epidemic mitigation, misinformation control, and network resilience enhancement.**

***Index Terms*—Reinforcement learning, graph neural networks, network dismantling, epidemic mitigation, network resilience**

I. Introduction

Pandemics, such as COVID-19, pose severe risks to human life and disrupt the social economy [1], [2]. The rapid and widespread transmission of infectious diseases strains healthcare systems, exacerbates socio-economic disparities, and challenges governments in decision-making regarding public health interventions. Vaccination remains one of the most effective measures for controlling pandemics by reducing both transmission and disease severity. However, the equitable and efficient distribution of vaccines presents a formidable challenge, particularly in the early stages when supply is limited [3]. Ensuring that vaccines reach those who need them most—whether to curb the spread of infection or to prevent fatalities—requires a strategic allocation approach informed by both epidemiological insights and network science.

The application of Social Network Analysis (SNA) in public health, particularly in immunization efforts, represents a crucial intersection between epidemiology and network theory [4]. This approach is based on the premise that human interactions and social structures significantly influence the transmission of infectious diseases. By identifying and targeting influential individuals within these networks, more effective disease control and prevention strategies can be implemented [5]. It is well established that not all nodes in a network hold the same level of importance [6], [7]. Some individuals exert greater influence than others, making them key figures in shaping disease transmission dynamics [8].

Various methodologies have been proposed to identify key individuals for immunization. Centrality-based approaches, such as degree centrality, betweenness centrality [9], and closeness centrality [10], prioritize highly connected nodes to disrupt transmission pathways. However, these methods often oversimplify network dynamics and fail to account for variations in individual risk factors. More advanced techniques, including PageRank [11], Coreness [12], and VoteRank [13], improve upon these limitations but still focus predominantly on reducing infection rates rather than mortality outcomes. An alternative approach involves hierarchical network decomposition, where methods like k-core decomposition [14] are used to rank nodes based on their structural influence. These strategies have been extended in various studies [15]–[18] to improve epidemic control. Additionally, heuristic methods, such as community-based selection [19] and percolation theory-based approaches [20], aim to fragment the network more efficiently. In recent years, data-driven techniques have emerged, leveraging machine learning [21]–[24] and reinforcement learning [25], [26] to optimize vaccination decisions. These methods have demonstrated promising results in dynamically adapting to changing network structures and epidemic conditions.

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through Discovery Grants RGPIN-2026-04763 and RGPIN-2025-04272, and by Simon Fraser University through Grants N001210 and N001230.

Z. Dong is with the School of Computing, Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: z.dong@queensu.ca).

Y. Chen is with the School of Computing, Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: yuanzhu.chen@queensu.ca).

S. Kafaie is with Department of Mathematics and Computing Science, Saint Mary's University, Halifax, NS B3H 3C3, Canada (e-mail: Somayeh.Kafaie@smu.ca).

K. Qiao is with the Department of Earth System Science, Stanford University, Stanford, CA 94305, United States (e-mail: qiaok@stanford.edu).

C. Li is with the School of Engineering Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada (e-mail: cheng_li_5@sfu.ca).

While these approaches can effectively reduce infection spread, they typically optimize for transmission suppression rather than mortality reduction. In practice, minimizing the number of infections does not necessarily minimize the number of deaths. It is well-documented that individuals vary in vulnerability due to factors such as age, comorbidities, and access to healthcare. Those with pre-existing health conditions, limited access to medical care, or advanced age are disproportionately affected by severe disease outcomes, making mortality a crucial consideration in vaccination strategies. Furthermore, socio-economic disparities and geographical barriers further complicate equitable vaccine distribution, leaving high-risk populations at greater peril. The prioritization of highly connected individuals for vaccination may reduce transmission rates but does not necessarily prevent the highest number of deaths. While network-based methods focus on minimizing the spread of disease, they often fail to account for the critical need to protect individuals with the highest likelihood of severe outcomes. Given that pandemics often result in overwhelmed healthcare systems, mitigating mortality can alleviate the burden on hospitals and critical care facilities, ensuring better overall health outcomes. This underscores the need for a more refined strategy that integrates both epidemiological spread and individual mortality risk to ensure a fair and effective allocation of vaccines. By incorporating mortality risk into immunization policies, public health initiatives can achieve a more balanced and impactful response, ultimately saving more lives while still controlling transmission.

Recent work has begun exploring machine learning methods for identifying influential nodes in networks. Graph representation learning and reinforcement learning approaches have shown promise in learning adaptive intervention policies that consider both structural properties and epidemic dynamics. However, existing models primarily focus on maximizing network disruption or minimizing infection spread. The explicit optimization of population-level mortality outcomes remains largely unexplored in graph-based learning frameworks.

To address this gap, this paper proposes a mortality-aware sequential vaccination framework that integrates graph representation learning with reinforcement learning. Instead of ranking nodes solely by structural influence, the proposed method models vaccination as a sequential decision-making process in which an agent selects individuals for vaccination while considering both network topology and mortality risk. A graph neural network (GNN) encoder captures structural relationships among individuals and integrates node-level mortality risk into learned embeddings. A deep reinforcement learning (DRL) agent then learns a vaccination policy that sequentially selects nodes in order to minimize cumulative deaths under epidemic dynamics.

This formulation differs from traditional risk-weighted heuristics in two important ways. First, mortality risk is incorporated directly into the state representation used by the learning agent, allowing the model to consider interactions between vulnerability and network structure. Second, the vaccination strategy is learned as a dynamic policy over evolving network states, rather than as a static ranking of nodes. This enables the framework to adapt decisions as the network structure changes due to vaccinations and epidemic progression. The primary contributions of this study are as follows:

1) We formulate vaccination-oriented network intervention as a mortality-aware graph decision problem, in which the objective is not only to suppress epidemic spread but also to explicitly minimize expected deaths.
2) We design a graph-based reinforcement learning framework that integrates network structure, node-level mortality risk, and intervention status into a unified MDP, allowing the policy to optimize mortality-aware node selection sequentially.
3) We show that the learned policy can be trained on small synthetic graphs yet generalize effectively to larger synthetic and real-world networks, which makes the approach practical despite the cost of simulation-based training.
4) Through experiments on multiple synthetic and real-world networks, we demonstrate that the proposed method achieves lower cumulative death rates while also producing strong network fragmentation compared with existing dismantling baselines.

The novelty of this work does not lie in the isolated use of GNNs or DRL, but in the formulation of epidemic intervention as a mortality-aware sequential graph optimization problem, together with a reward design and graph encoding that directly optimize for death reduction rather than only transmission reduction or structural fragmentation.

II. Related Work

A. Centrality- and Heuristic-Based Intervention Methods

Early approaches to epidemic intervention in networks primarily rely on centrality-based measures, such as degree, betweenness [9], and closeness centrality [10], to identify influential nodes for immunization. These methods aim to disrupt transmission pathways by targeting structurally important nodes. To integrate richer network information for vaccine prioritization, researchers have proposed advanced centrality measures, including PageRank [11], Coreness [12], LocalRank [27], VoteRank [13], ClusterRank [28], LeaderRank [29], and TwitterRank [30]. Each of these metrics provides a distinct perspective for analyzing network influence by incorporating various topological and functional properties. Their effectiveness, however, depends on factors such as network size, density, connectivity patterns, and domain-specific attributes. Although these centrality-based approaches are computationally efficient, they primarily focus on network influence rather than individual vulnerability. As a result, vaccination strategies derived from these metrics typically aim to minimize infection spread rather than reduce mortality among high-risk individuals. The k-core decomposition method [14] and its extensions [15]–[18] refine this approach to better identify influential nodes. Beyond these methods, heuristic approaches have been explored for vaccine allocation. Chen et al. [31] introduced a degree discounting algorithm, improving computational efficiency while maintaining accuracy compared to traditional greedy algorithms [32]. He et al. [19] applied

community detection [33] to prioritize nodes within distinct subgroups, and Morone and Makse [20] leveraged percolation theory to determine the minimal set of nodes necessary to maintain network connectivity. Despite their effectiveness, these methods primarily target reducing disease transmission rather than mortality minimization.

Heuristic strategies based on percolation theory and network dismantling, including methods such as Collective Influence (CI) [34] and other structural optimization techniques, have also been widely studied due to their computational efficiency and effectiveness in fragmenting networks. However, these approaches are fundamentally *topology-driven* and typically aim to reduce connectivity or slow epidemic spread, without explicitly considering heterogeneity in node-level mortality risk.

B. Learning-Based Network Dismantling and Reinforcement Learning

Recent work has explored learning-based approaches for network dismantling and intervention. Methods such as GDM [35] and GND [36] leverage learned representations or optimization strategies to identify critical nodes more effectively than traditional heuristics. More recently, reinforcement learning (RL) has been applied to sequential node selection problems, enabling adaptive decision-making as the network evolves. A representative example is the work by Fan et al. [37], which formulates the identification of key nodes in complex networks as a reinforcement learning problem. Their approach demonstrates that deep RL can outperform heuristic strategies by learning structural patterns directly from data. Similarly, graph-based reinforcement learning methods such as GERL [26] combine Graph Neural Networks (GNNs) with RL to learn node removal strategies in an end-to-end manner. Despite these advances, most existing learning-based methods focus on optimizing *structural objectives*, such as minimizing the size of the largest connected component or maximizing network fragmentation. Mortality outcomes are typically not explicitly incorporated into the optimization objective.

C. Epidemic Modeling and Intervention

Recent studies have further explored integrating learning-based models with epidemic dynamics. For instance, neural ordinary differential equation (Neural ODE) approaches have been proposed to model the coupled dynamics of epidemic spreading and information diffusion [38]. In addition, reaction-diffusion models on multiplex networks have been developed to capture interactions between epidemic processes, media influence, and human mobility [39]. While these methods provide sophisticated tools for modeling epidemic behavior, they are primarily designed for *prediction* or *analysis* of spreading dynamics, rather than for sequential decision-making and intervention optimization.

D. Research Gap and Motivation

In summary, existing approaches to epidemic intervention in networks can be broadly categorized into heuristic methods, learning-based dismantling strategies, and epidemic modeling frameworks. However, most prior work focuses on reducing infection spread or disrupting network structure, while treating mortality as a secondary evaluation metric. In particular, few methods explicitly incorporate *node-level mortality risk* into both the state representation and reward function of a sequential decision-making framework. As a result, existing approaches are not directly optimized for minimizing population-level deaths. To address this gap, we propose a mortality-aware graph reinforcement learning framework that integrates structural information and node-level risk into a unified decision process, enabling end-to-end optimization of intervention strategies with respect to mortality outcomes.

III. Proposed method

The framework illustrated in Figure 1 integrates reinforcement learning with graph representation learning to optimize decision-making in complex network environments. With death risk incorporated as a node feature, the model can more effectively identify which nodes require protection to minimize overall mortality within the population. This capability enhances the decision-making process by ensuring that the removal or intervention strategy prioritizes safeguarding individuals at the highest risk.

To improve clarity, we describe the overall decision pipeline as follows. At each step, the current graph state is represented by the adjacency matrix and node-level mortality risk. These inputs are encoded by a GNN to produce node embeddings, which capture both structural and risk information. A Q-network then assigns a value to each node, representing its expected contribution to mortality reduction. The node with the highest Q-value is selected for intervention, after which the graph is updated and a reward is computed using SIR simulations. This process is repeated iteratively to construct a sequential intervention strategy.

A. Encoder and Decoder

At the core of the framework is the encoder, which processes the network structure and transforms it into meaningful representations. The encoder is implemented using a Graph Neural Network (GNN), allowing each node to learn from its surrounding connections through iterative message passing. Each node in the graph can be characterized by a set of features, such as age, sex, and income, which can be used to estimate an individual's death risk. In our current experimental setting, this risk is directly provided as a scalar node attribute. Over successive iterations, nodes accumulate information from their surroundings, integrating both structural dependencies and death risk data. This process allows each node to develop an embedding that encodes both local interactions within its neighborhood and broader network-level influences, ultimately mapping the network into a lower-dimensional latent space. The GNN encoder maps the current graph state to node embeddings:

$$h_i^{(t)} = \mathrm{GNN}(A_t, X_t)_i, \tag{1}$$

where A_t is the adjacency matrix at step t, and X_t is the matrix of node features.

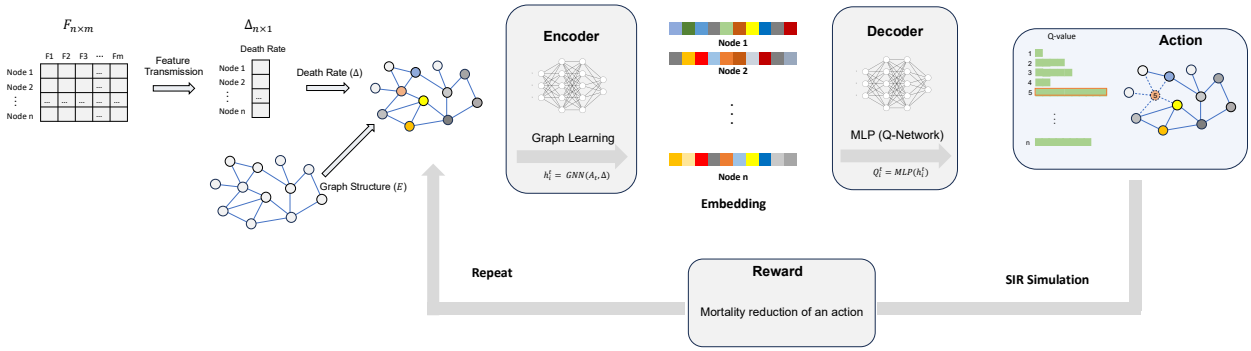


Fig. 1: Overview of the proposed mortality-aware graph reinforcement learning framework. At each step, the current graph state, comprising the adjacency matrix and node-level mortality risk, is encoded by a GNN into node embeddings. A Q-network (MLP) maps each node embedding to an action value. The node with the highest Q-value is selected for intervention. The graph is then updated by removing or vaccinating the selected node, and the reward is computed based on the reduction in mortality estimated via repeated SIR simulations. This process is repeated sequentially until the intervention budget is reached.

By explicitly integrating death risk into the node embeddings, the model gains a crucial advantage in understanding which nodes are most vulnerable. This ensures that decisions about intervention or removal are not based solely on topological importance but also account for the potential human cost of each action. Once node embeddings are generated, the decoder takes these representations as input and computes a score function that guides decision-making. The decoder is implemented as a multi-layer perceptron (MLP), which maps the encoded features to scalar Q-values. Each node embedding $h_i^{(t)}$ captures both structural information and node-level attributes. These embeddings are then passed to a Q-network, implemented as a multi-layer perceptron (MLP), to compute action values:

$$Q(s_t, a = i) = \text{MLP}(h_i^{(t)}). \tag{2}$$

The agent selects the node with the highest Q-value:

$$a_t = \arg\max_{i \in \mathcal{V}_t} Q(s_t, a = i), \tag{3}$$

where $\mathcal{V}_t$ is the set of nodes that have not yet been vaccinated or removed. These values quantify the expected long-term reward associated with selecting a given node for intervention or removal. By embedding death risk directly into the learned representations, the model can prioritize the protection of high-risk nodes, ensuring that the overall mortality in the network is minimized.

In this work, although the framework is designed to incorporate rich demographic and health-related features (e.g., age, sex, and income), the experimental evaluation uses simplified scalar death risk values assigned to each node. This design choice allows us to isolate and evaluate the effectiveness of the proposed reinforcement learning framework in leveraging mortality-aware signals, without introducing additional confounding factors from feature engineering. Extending the model to learn death risk from real-world demographic features is an important direction for future work.

B. Reinforcement Learning Framework for Node Selection

The entire process is structured as an iterative reinforcement learning loop, where an agent interacts with the network environment through sequential decision-making. At each step, the current state represents the network configuration, which dynamically evolves as nodes are protected or removed. The available actions involve selecting a node to target based on its computed Q-value, and the reward function measures the effectiveness of the action in reducing mortality within the population.

During the training phase, synthetic networks are randomly sampled and processed through the framework. The agent learns an optimal intervention strategy by iteratively interacting with different network structures, refining its decision-making process based on observed rewards. The encoder continuously updates node embeddings. The decoder adapts its scoring mechanism to improve the accuracy of action selection, ensuring that the model effectively prioritizes high-risk nodes for protection. Over multiple training episodes, the framework learns to balance topological optimization with human-centered risk mitigation, identifying strategies that minimize overall loss of life. Once training is complete, the trained model can be deployed on real-world networks to guide intervention strategies. When a new network is provided, the framework first encodes its structure into compact vector representations. These embeddings, enriched with death risk information, are then used to estimate Q-values for each node. At each iteration, the node with the highest Q-value—which represents the best trade-off between structural importance and mortality risk—is selected for intervention. The process continues until a predefined stopping criterion is met, such as reaching a target reduction in overall mortality or protecting a critical subset of individuals. The final sequence of protected nodes represents the optimal intervention strategy, ensuring that the fewest possible lives are lost.

A key advantage of this framework is its ability to balance immediate structural gains with long-term human impact.

Traditional heuristic approaches often prioritize nodes based on their connectivity or influence within the network, but they may overlook critical factors such as individual vulnerability. By embedding death risk directly into the learned representations, this model ensures that intervention strategies are more socially responsible and ethically grounded. Furthermore, as the network embeddings are dynamically updated at each step, the approach continuously refines its understanding of the evolving risk landscape, enabling more adaptive and effective decision-making.

C. MDP Formulation

The process of node selection for intervention is formulated as a Markov Decision Process (MDP), where an agent interacts with a graph-structured environment through sequential decision-making. The reinforcement learning (RL) model optimizes an intervention policy that selects critical nodes for protection or removal, ensuring minimal mortality in the population while considering network topology and individual risk factors.

A Markov Decision Process (MDP) is defined by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$, where $\mathcal{S}$ represents the collection of states, $\mathcal{A}$ corresponds to the collection of available actions, $\mathcal{T} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ denotes the transition function, $\mathcal{R} : \mathcal{S} \rightarrow \mathbb{R}$ represents the reward function, and γ serves as the discount-rate parameter. At each timestep t, the agent observes the current state s_t, selects an action a_t based on the learned policy $\pi(s)$, and transitions to a new state s_{t+1} according to the environment's transition dynamics. The agent receives a reward $R(s_t, a_t)$, which guides the learning process. The objective is to find an optimal policy π^* that maximizes cumulative rewards over time.

State Representation: The state $s_t \in S$ at timestep t represents the network configuration and includes:

- Graph structure: Adjacency matrix $A_t \in \mathbb{R}^{N \times N}$ describing node connectivity. The index t in A_t denotes the intervention step in the sequential decision process, rather than an externally evolving temporal network. The graph structure changes as a result of the agent's actions: when a node is selected for intervention, it is removed (or vaccinated), along with its incident edges, leading to an updated adjacency matrix A_{t+1}.
- Death risk: A vector $\Delta \in \mathbb{R}^N$ indicating the death risk level if infected of each node. During the sequential decision process, the Δ_i remain fixed.
- Vaccination status: A vector $P_t \in \mathbb{R}^N$ indicating the vaccination status of each node, where N is the number of nodes in the network.

The GNN encoder processes this information and generates node embeddings h_t^i for each node i:

$$h_t^i = \mathrm{GNN}\left(A_t, \Delta, P_t\right). \tag{4}$$

With the nodes in the graph already being represented by their embeddings, this state can be thought of as a vector in a d-dimensional space, mathematically expressed as the summation over all nodes:

$$s_t = \sum h_t^i. \tag{5}$$

At any given step, the state corresponds to the current structure of the network, along with the embedded node features (including death risk and vaccination status). As nodes are vaccinated (removed from the network), the network changes dynamically, leading to an evolving state space.

Action Space: The action space consists of selecting a node $v \in V$ at each timestep for intervention. The goal is to choose nodes that maximize long-term benefits in reducing mortality in the network. Using Q-learning, the agent estimates a Q-value $Q(s, a)$ for each node:

$$Q\left(s_t, a_t\right) = R\left(s_t, a_t\right) + \gamma \max_{a'} Q\left(s_{t+1}, a'\right) \tag{6}$$

where $Q\left(s_t, a_t\right)$ represents the expected cumulative reward of selecting node a_t in state s_t, $R\left(s_t, a_t\right)$ is the immediate reward received for the action, $\max_{a'} Q\left(s_{t+1}, a'\right)$ estimates the future value of the next best action. The agent selects the node with the highest Q-value, meaning:

$$a_t = \arg\max_a Q\left(s_t, a\right). \tag{7}$$

Reward Function: The reward function $R(s_t, a_t)$ is designed to prioritize actions that reduce overall mortality. In a formal setting, the reward function quantifies the change in death over the network following the execution of action a, resulting in a transition to the subsequent state s_{t+1}. This can be mathematically expressed as:

$$R(s_t, a_t) = \frac{D(I(G_t, a_t), \Delta) - D\left(I(G_t), \Delta\right)}{N}, \tag{8}$$

where G_t is the network in timestep t, function I gets the infected nodes as determined by an SIR simulation, and function D calculates the death number. Function D is defined as:

$$D(I, \Delta) = I \cdot \Delta^T. \tag{9}$$

where $I \in \mathbb{R}^N$ indicates the infection status of each node and the vector $\Delta \in \mathbb{R}^N$ indicates the death risk level if infected of each node.

Since the reward is computed based on an SIR simulation, nodes can transition through susceptible (S), infected (I), and recovered (R) states during the epidemic process. Recovered nodes are naturally handled by the SIR model as non-infectious individuals who no longer contribute to further transmission. Their influence on the epidemic is implicitly captured through the simulation dynamics prior to recovery. In our formulation, the function $I(G_t, a_t)$ represents the infection outcome over the entire SIR rollout after applying action a_t, indicating whether each node becomes infected at any point during the simulation. The reward is then computed based on the resulting mortality outcome, rather than on intermediate infection states. Therefore, recovered nodes are not explicitly modeled as active spreaders in the reward calculation, but their impact is fully reflected in the simulated epidemic trajectory.

Since the SIR process is inherently stochastic, the reward signal can exhibit variance across different simulation runs. To obtain a more stable estimate of the reward, we compute it as the average outcome over multiple independent SIR simulations. In our implementation, we use 50 simulation runs per action, which provides a good balance between

estimation stability and computational efficiency. Using too few simulations leads to high-variance reward estimates that can negatively affect the stability of Q-learning, while using a significantly larger number of simulations yields diminishing returns in variance reduction relative to the additional computational cost. Empirically, we found that 50 simulations are sufficient to ensure stable training without incurring excessive overhead.

Policy Optimization and Training:

The policy $\pi(a \mid s)$ defines a mapping from the current graph state s to a probability distribution over candidate nodes for intervention. The objective of the reinforcement learning process is to learn an optimal policy π^* that maximizes the expected cumulative reward over the intervention sequence. In the DQN-based implementation used in this work, the policy is implicitly represented by the learned Q-function. During execution, the policy is applied in a greedy manner by selecting the node with the highest Q-value:

$$a_t = \arg\max_{a \in \mathcal{A}_t} Q(s_t, a), \tag{10}$$

where $\mathcal{A}_t$ is the set of available nodes at step t. Thus, the learned policy corresponds to a sequential node-selection strategy that determines which node to vaccinate or remove at each step, conditioned on the current graph state. The final output of the policy is an ordered sequence of selected nodes representing the intervention plan.

D. Complexity Analysis

The computational complexity of the proposed framework for minimizing population mortality depends on multiple factors, including the graph representation learning process, the reinforcement learning updates, and the decision-making process during deployment. Below, we analyze the complexity of each major component.

Graph Representation Learning Complexity: The encoder is based on a GNN, which processes node features and updates embeddings through message passing. Assuming a graph with N nodes and E edges, the time complexity of a Graph Convolutional Network (GCN) with L layers can be approximated as: $O(L(E + N))$. Each node aggregates information from its neighbors in every layer, leading to a complexity of $O(E)$ for message passing. The transformation of node features in each layer contributes an additional $O(N)$ factor. Since the process runs for L layers, the total complexity scales with $O(L(E + N))$.

Reinforcement Learning Complexity: The reinforcement learning process involves updating Q-values and optimizing the policy over multiple training episodes. The complexity is influenced by:

- State Space Size: The number of possible network configurations (S).
- Action Space Size: The number of possible node selections $(A \approx N)$.
- Q-Value Computation: Using a neural network, typically modeled as an MLP, which has a complexity of $O(N_{hidden} \times A)$, where N_{hidden} is the number of hidden neurons.
- Policy Updates: If trained using Deep Q-Learning, each update step has a complexity of $O(A)$ per iteration.

Over T training iterations, the complexity becomes $O(T \times A)$. Considering all components, the overall complexity of reinforcement learning is approximately: $O(T \times N \times N_{hidden})$, where T is the number of training iterations. The value of N_{hidden} depends on the neural network architecture but is typically much smaller than N, ensuring that training remains feasible.

Decision-Making Complexity During Deployment: Once training is complete, applying the model to real-world networks involves:

- Encoding the network using the trained GNN: $O(L(E + N))$
- Computing Q-values for all nodes using the trained policy: $O(N_{hidden} \times N)$
- Selecting the highest Q-value node at each iteration: $O(N)$
- Updating the network and repeating until termination (typically K iterations): $O(KN)$

The overall complexity of the deployment phase is approximately: $O(L(E + N) + K(N + N_{hidden}))$, where K is the number of node selections before reaching the stopping criterion. Since $K \ll N$, the decision-making process remains efficient.

The proposed framework balances expressiveness and efficiency. While the training phase involves significant computational overhead due to reinforcement learning, this is mitigated by pre-training on synthetic networks. During deployment, the method remains efficient, scaling well with large networks and allowing for real-time decision-making in practical applications.

IV. Experiments

We compare the proposed GBRL method with several representative baselines, including Graph Dismantling with Machine Learning (GDM) [35], Generalized Network Dismantling (GND) [36], and GERL [26], as well as the Collective Influence (CI) [34] method. GDM utilizes machine learning techniques, including graph convolutional layers, to identify and eliminate critical nodes for network dismantling. GND focuses on mitigating network fragmentation by selectively removing nodes based on cost considerations, incorporating both a spectral bisection strategy and a weighted vertex cover method. GERL introduces a novel framework that combines Graph Neural Networks (GNNs) with Deep Reinforcement Learning (DRL) to strategically disrupt network connectivity and curb the spread of diseases. CI is a widely used and highly effective heuristic based on optimal percolation theory, and has been shown to perform strongly in network dismantling tasks. It serves as a strong classical baseline for identifying influential nodes based on structural importance. Together, these baselines provide a comprehensive comparison across both heuristic and learning-based approaches for network intervention.

All experiments were performed on a Lenovo ThinkStation with Ubuntu 20.04.4 LTS, an Intel i9-14900KF processor, 64

GB of RAM, and an NVIDIA RTX 4090 GPU with 24 GB of memory. For training, the model was trained on 10 small synthetic random networks, each consisting of 50 nodes with an average degree of 6. Each node was assigned a death risk value randomly selected from 0.1, 0.3, 0.5, 0.7. The total training duration was approximately six hours.

It is important to note that the reported training time is primarily dominated by the repeated SIR simulations required for reward evaluation during reinforcement learning. Each action evaluation involves simulating epidemic dynamics on the network, which introduces significant computational overhead. However, this cost is incurred only during training. Once trained, the model can be efficiently deployed, as it only requires a forward pass through the GNN and MLP to compute node priorities.

Furthermore, the proposed framework does not require large-scale networks for training. In our experiments, training on small synthetic networks (e.g., 50–100 nodes) is sufficient for the model to learn effective intervention strategies that generalize well to larger and structurally diverse networks. This is because the model learns structural patterns and local interaction dynamics, which are transferable across networks of different sizes. This highlights an important advantage of the method: it can be trained in a data-efficient manner while still achieving strong performance on real-world networks.

A. Datasets

To assess the adaptability and performance of our proposed method, we carried out experiments on both synthetic and real-world network datasets. A summary of the statistical properties of these networks is provided in Table I. For synthetic networks, we examined three distinct types: random, scale-free, and small-world, each characterized by unique structural features. Additionally, we introduced community structures in random and scale-free networks to explore their influence on network spreading dynamics. However, due to its intrinsic connectivity pattern, the small-world network does not support a community-based variant. The synthetic networks were generated using established graph models: Erdős–Rényi (ER) model [40] for random networks, Barabási–Albert (BA) model [41] for scale-free networks, and Watts–Strogatz (WS) model [42] for small-world networks. These networks were constructed using NetworkX [43]. To generate networks with community structures, we employed a random modular network generator [44], configuring it with a modularity coefficient of $Q = 0.5$ and three distinct communities ($m = 3$).

We used four real-world networks: Netscience, Weaver, Mammalia, and Tortoise. Netscience [45] represents a collaboration network of researchers in network science, where nodes correspond to scientists and edges denote co-authored publications. Weaver [46] captures the social interactions of weaver birds, where nodes represent individual birds and edges indicate shared use of nest chambers. Mammalia [47] is a vole interaction network, in which nodes represent voles and edges are formed between individuals that were trapped together during primary trapping sessions. Tortoise [48] maps burrow-sharing interactions among tortoises, with edges connecting individuals that occupied the same burrow. This diverse selection of datasets enables a comprehensive evaluation of our method across networks with different structural properties and real-world applications.

B. Metrics of interest

The proposed method aims to achieve maximum network disruption with minimal node removals. To evaluate its effectiveness, we utilize five key performance metrics that measure the degree of network fragmentation following node elimination. The results obtained from our approach are compared against those of baseline methods to assess relative performance.

Edge Count: The total number of connections in a network. A lower edge count after node removal indicates stronger network disruption, limiting disease or information spread.

Network Components: The number of disconnected sub-networks. A higher count means greater fragmentation, reducing propagation. To ensure consistency in result visualization, we use the reciprocal of the component count, where a smaller reciprocal value implies greater efficiency in hindering pathogen transmission.

Largest Component Size: The size of the biggest remaining connected sub-network. A smaller largest component reflects greater network breakdown, restricting large-scale transmission.

Epidemic Threshold [49]: The epidemic threshold measures a network's resistance to pathogen spread. An outbreak occurs when the spreading rate exceeds this threshold. Using the Susceptible-Infected-Susceptible (SIS) model, it is defined as:

$$\tau = \frac{1}{\lambda_{\max}}, \tag{11}$$

where $\lambda_{\max}$ represents the largest eigenvalue of the network's adjacency matrix. A higher epidemic threshold indicates a more resilient network, making disease propagation more difficult.

Average Node Connectivity [50]: Average node connectivity measures the expected number of node removals needed to disconnect non-adjacent nodes, indicating network robustness. It is defined as:

$$\bar{\kappa}(G) = \frac{\sum_{u,v} \kappa_G(u,v)}{\binom{n}{2}}, \tag{12}$$

where $\kappa_G(u,v)$ represents the maximum number of node-disjoint paths connecting node u to node v, and n is the total number of nodes in the network. A higher value signifies stronger network resilience, while a lower one reflects greater structural disruption.

Death Rate: In a network vaccination model, where nodes represent individuals and edges represent potential transmission pathways, the death rate quantifies the proportion of individuals (nodes) who die as a result of infection during the simulated epidemic process. It is defined as:

$$R_{death} = \frac{N_{dead}}{N} = \frac{I \cdot \Delta^T}{N}, \tag{13}$$

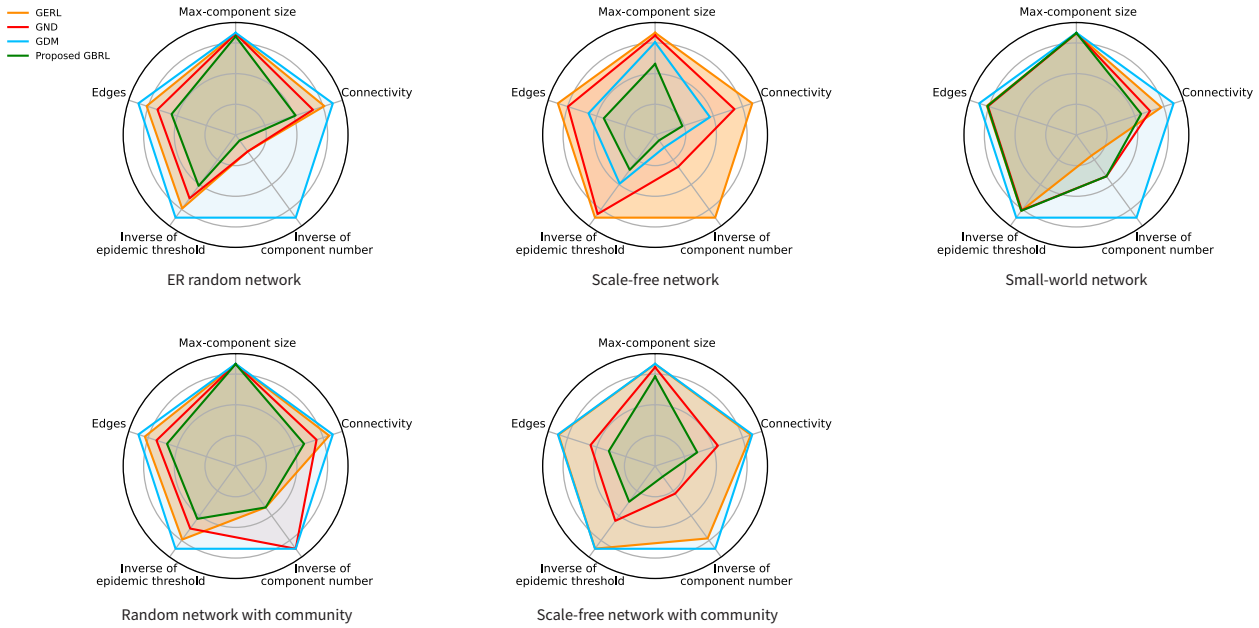


Fig. 2: Impact of 20% node removal on network degradation across different models in synthetic datasets

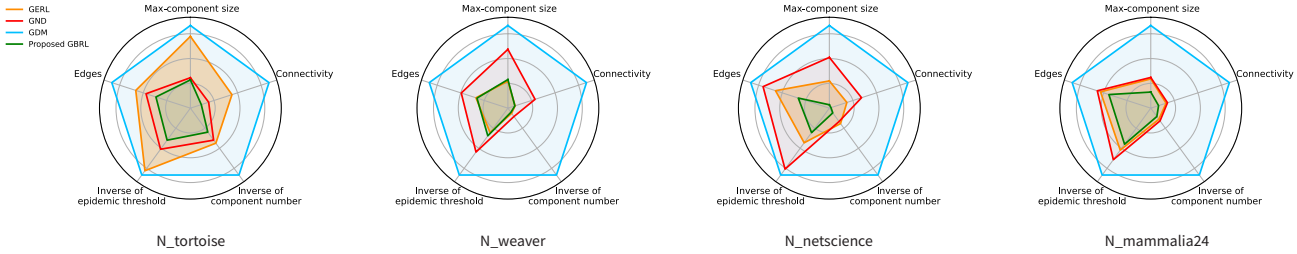


Fig. 3: Impact of 20% node removal on network degradation across different models in real-world datasets

where $I \in \mathbb{R}^N$ indicates the infection status of each node, the vector $\Delta \in \mathbb{R}^N$ indicates the death risk level if infected of each node, and N is the total number of nodes.

C. Results

1) Comparative Evaluation of Network Structural Disruption: The connectivity of a network serves as a critical pathway for the dissemination of pathogens, influencing the extent to which they can spread. The topological structure of a network plays a fundamental role in determining its susceptibility to pathogen transmission or information diffusion. This section examines the impact of the proposed framework on network integrity, utilizing five core metrics: edge quantity, connectivity, epidemic threshold, largest connected component size, and number of network components. These measures collectively offer a detailed evaluation of how well the framework disrupts network cohesion. Each metric captures a distinct structural characteristic, and all assessments are conducted on the remaining network after a specified number of node removals.

Among these metrics, the epidemic threshold is particularly significant, as it represents a network's resistance to pathogen invasion. A pathogen can only spread if its transmission rate exceeds this threshold [51], meaning that networks with higher thresholds exhibit stronger resilience against outbreaks. Another key indicator is the largest connected component, which refers to the most extensive subset of nodes where all members remain reachable from one another. This component plays a central role in maintaining overall network connectivity and reflects the system's robustness. Additionally, the number of network components reflects the degree of fragmentation within the network, indicating how many separate, internally connected groups exist. Each of these groups remains isolated from others, providing insight into the network's structural coherence and resilience.

To benchmark the effectiveness of the proposed GBRL, its performance is compared against three established network dismantling strategies: GDM, GND, and GERL. To assess the generalizability of the proposed approach, experiments were performed on five synthetic network models and four real-world networks. The synthetic models include ER random

TABLE I: Properties of network datasets

Network	Type	Nodes	Edges	Average degree	Max-degree	Clustering coefficient	Diameter
SF-100	Synthetic	100	292	5.9	24	0.14	4
SF-500	Synthetic	500	1491	6.0	52	0.04	6
ER-100	Synthetic	100	306	6.1	15	0.05	6
ER-500	Synthetic	500	1498	6.0	15	0.01	7
SW-100	Synthetic	100	300	6.0	7	0.5	8
SW-500	Synthetic	500	1500	6.0	8	0.46	10
SF-C-100	Synthetic	100	358	7.2	21	0.23	6
SF-C-500	Synthetic	500	1852	7.4	35	0.06	7
ER-C-100	Synthetic	100	315	6.3	12	0.12	5
ER-C-500	Synthetic	500	1573	6.3	14	0.01	7
netscience	Real-world	380	914	4.8	34	0.74	17
weaver	Real-world	64	177	5.5	21	0.60	6
mammalia	Real-world	172	363	4.2	12	0.69	23
tortoise	Real-world	283	418	3.0	11	0.43	8

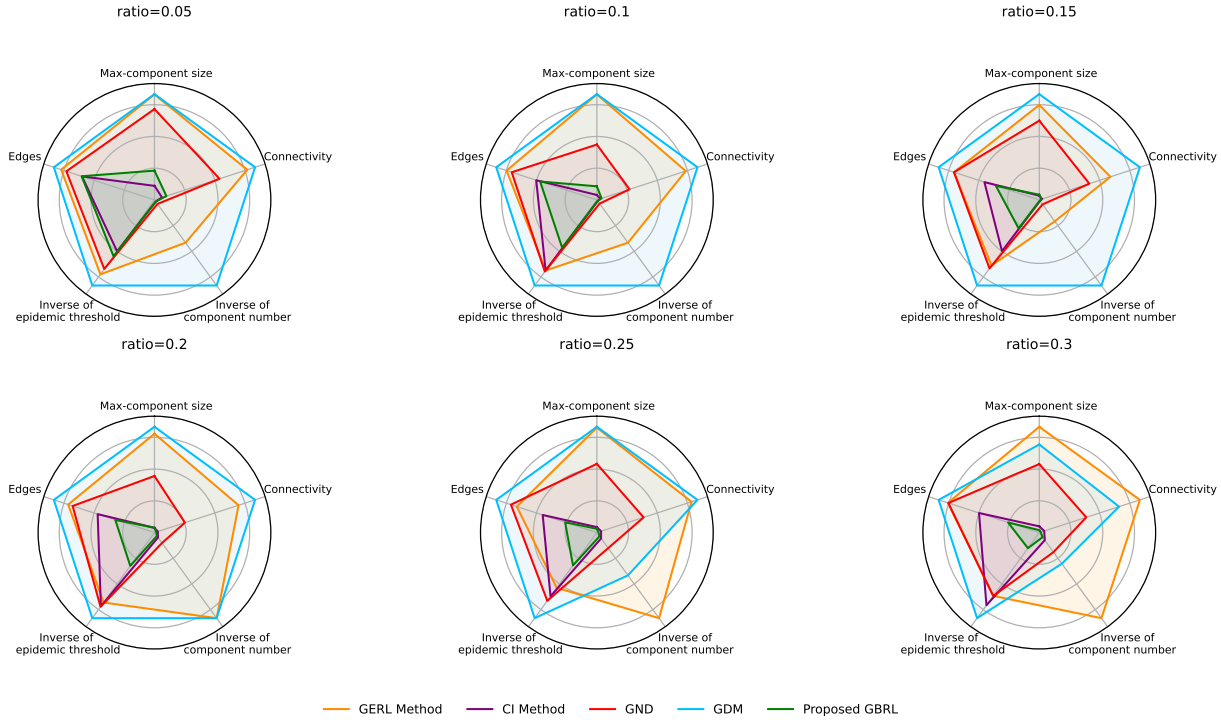


Fig. 4: Network degradation comparison across ratios

networks, ER networks with community structures, scale-free networks, scale-free networks with communities, and small-world networks. Community structures play a crucial role in spreading dynamics, as nodes within the same community tend to form dense internal connections while having fewer external links. To investigate this impact, ER and scale-free networks were modified to incorporate communities, allowing for a more comprehensive evaluation of the framework's adaptability across different network topologies.

Figures 2 and 3 illustrate the superior performance of the proposed method over GERL, GND, and GDM in disrupting both synthetic and real-world networks following the removal of 20% of nodes. In synthetic networks, the proposed approach proves particularly effective in scale-free and scale-free with community networks, where strong internal linkages typically enhance network robustness. The radar charts reveal that the proposed method consistently achieves lower edge counts, reduced connectivity, smaller max-component sizes, and a higher number of disconnected components, indicating that it fragments networks more efficiently than competing methods. In real-world datasets, the proposed method continues to exhibit strong network fragmentation, as evidenced by smaller largest connected components and increased epidemic thresholds, making the remaining network less prone to the spread of diseases or information. In contrast, GERL, GND and GDM show varied effectiveness depending on network topology, with GND being more competitive in certain cases but inconsistent overall. Notably, small-world networks demon-

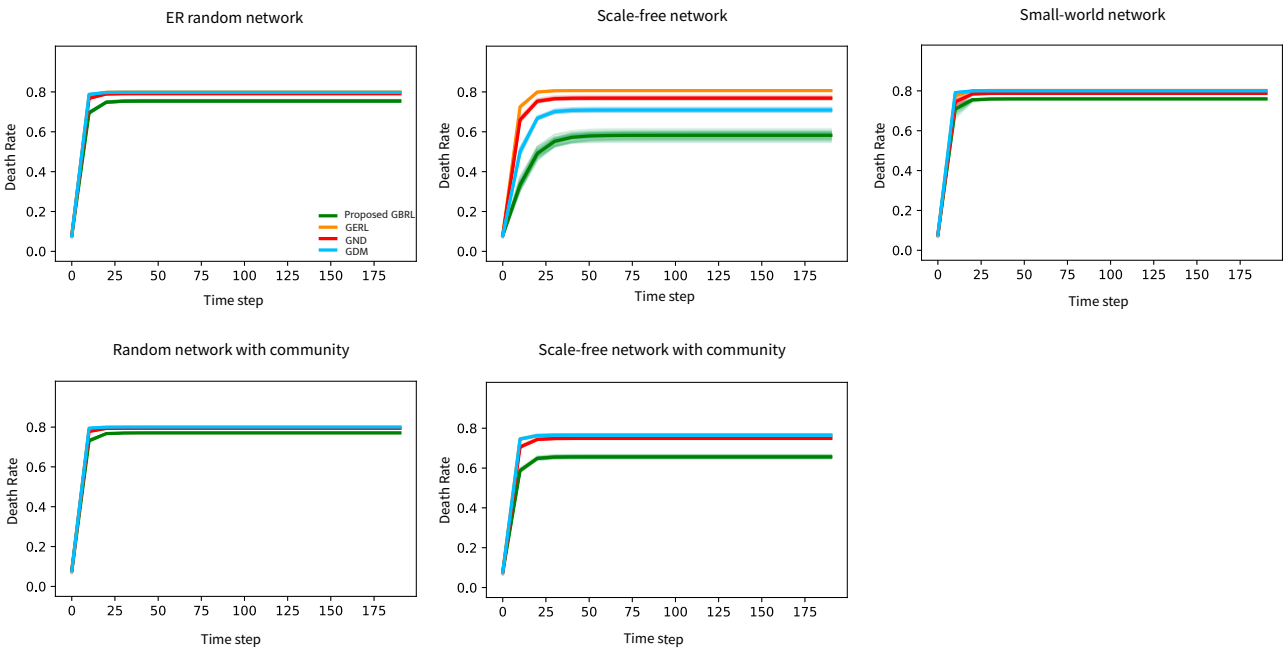


Fig. 5: Cumulative death rate on five synthetic networks after 20% node removal by different models, with $\beta = 0.1$, $\gamma = 0.01$, and $\rho = 0.1$ in networks of 500 nodes

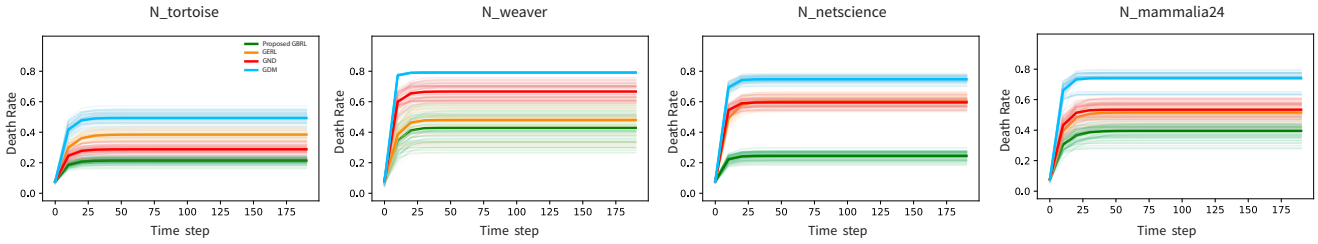


Fig. 6: Cumulative death rate on four real-world networks after 20% node removal by different models, with $\beta = 0.1$, $\gamma = 0.01$, and $\rho = 0.1$

strate higher resilience to structural disruption due to their homogeneous structure and absence of dominant hubs. Their short path lengths facilitate efficient communication between nodes, making them harder to dismantle through the selective removal of a limited number of nodes. In contrast, scale-free networks are the most vulnerable to targeted node removal, as they rely heavily on high-degree hubs to sustain connectivity. Immunizing or removing these hubs significantly weakens network integrity and impedes propagation, highlighting the effectiveness of targeted intervention strategies in scale-free structures. By maximizing network fragmentation while increasing the epidemic threshold, the proposed method effectively amplifies network vulnerability to targeted interventions, curbing large-scale spread and minimizing overall risk. These findings underscore the adaptability and robustness of the proposed approach, making it a highly effective strategy for applications in network dismantling, epidemiological control, and risk mitigation.

2) Performance Across Different Node Removal Ratios: To evaluate the robustness of the proposed method beyond a single intervention budget, we extend our analysis to a range of node removal ratios from 5% to 30%. Figure 4 illustrates the detailed performance trends for a representative dataset, while Tables II–V summarizes the results across all the real-world datasets.

As shown in Figure 4, the proposed GBRL method demonstrates consistent performance improvements across all removal ratios. At lower removal levels (e.g., 5% and 10%), the differences between methods are relatively modest, as the network structure remains largely intact. Nevertheless, the proposed method already shows competitive performance in reducing connectivity and increasing the epidemic threshold. As the removal ratio increases (15%–30%), the advantages of the proposed method become more pronounced. In particular, GBRL achieves smaller maximum component sizes, reduced connectivity, and higher epidemic thresholds, indicating more effective network fragmentation and improved resistance to epidemic spreading. To further validate these observations, Tables II–V report the quantitative results across multiple datasets. The proposed method consistently outper-

TABLE II: Network degradation comparison across ratios on mammalia24 (lower is better)

Metric	Method	0.05	0.10	0.15	0.20	0.25	0.30
Conn.	CI	0.1878	**0.1137**	**0.0933**	0.0858	0.0890	0.0953
	GDM	0.6711	0.7474	0.7953	0.8643	0.9277	0.9811
	GERL	0.3859	0.4009	0.3652	0.3400	0.3175	0.2974
	GND	0.4268	0.5069	0.2015	0.2747	0.1741	0.0942
	GBRL	**0.1671**	0.1442	0.1019	**0.0748**	**0.0564**	**0.0525**
Edges	CI	309	253	225	200	182	170
	GDM	356	349	339	326	305	283
	GERL	330	305	280	256	234	217
	GND	316	306	252	237	205	157
	GBRL	**286**	**241**	**193**	**159**	**125**	**106**
Inv.Comp.	CI	0.0588	0.0400	0.0333	0.0312	0.0286	0.0303
	GDM	0.1429	0.1667	0.2500	0.3333	0.3333	0.3333
	GERL	0.0909	0.0909	0.0909	0.0909	0.0909	0.0909
	GND	0.0667	0.0833	0.0476	0.0588	0.0476	0.0400
	GBRL	**0.0476**	**0.0333**	**0.0294**	**0.0250**	**0.0208**	**0.0204**
Inv.Epid.	CI	4.2507	**3.3514**	**2.2778**	6.0667	6.0667	6.0667
	GDM	4.8940	4.8940	4.9283	4.9670	4.9775	4.9650
	GERL	4.9271	4.9271	4.7964	4.5874	4.5702	4.5962
	GND	4.4265	4.6083	4.0897	4.3058	3.8308	2.8800
	GBRL	**4.2267**	4.2267	3.0566	**2.6857**	**2.8333**	**2.8333**
MaxComp.	CI	36	36	32	25	**19**	**19**
	GDM	119	119	116	113	110	106
	GERL	81	81	73	64	55	47
	GND	94	94	37	56	32	15
	GBRL	**35**	**35**	**30**	**22**	18	18

TABLE III: Network degradation comparison across ratios on netscience (lower is better)

Metric	Method	0.05	0.10	0.15	0.20	0.25	0.30
Conn.	CI	**0.0925**	**0.0440**	0.0370	0.0356	0.0364	0.0371
	GDM	1.2281	1.2555	1.2816	1.0581	0.8716	0.5977
	GERL	1.1329	1.1090	0.9068	0.8845	0.8201	0.7533
	GND	0.7944	0.4077	0.6380	0.3202	0.4050	0.3519
	GBRL	0.1459	0.0586	**0.0254**	**0.0180**	**0.0132**	**0.0109**
Edges	CI	**631**	510	439	393	358	324
	GDM	876	846	805	696	665	541
	GERL	812	759	683	597	529	491
	GND	767	716	682	566	568	489
	GBRL	634	**477**	**347**	**271**	**210**	**168**
Inv.Comp.	CI	0.0204	0.0135	0.0115	0.0114	0.0114	0.0111
	GDM	1.0000	1.0000	1.0000	0.2000	0.0833	0.0455
	GERL	0.5000	0.5000	0.2500	0.2000	0.1667	0.1250
	GND	0.0476	0.0435	0.0526	0.0244	0.0303	0.0286
	GBRL	**0.0192**	**0.0112**	**0.0088**	**0.0078**	**0.0071**	**0.0067**
Inv.Epid.	CI	**4.3000**	6.0000	4.4211	5.5926	5.5926	5.5926
	GDM	7.1575	7.2813	7.3801	6.4940	7.5075	6.5847
	GERL	6.2229	6.0105	5.6623	5.3078	4.8462	4.8713
	GND	5.7767	6.0918	5.9052	5.6297	5.9839	4.8758
	GBRL	4.7108	**4.0652**	**2.4545**	**2.5263**	**2.8824**	**1.2000**
MaxComp.	CI	118	**16**	**13**	13	13	13
	GDM	361	342	323	288	251	184
	GERL	360	341	290	269	249	221
	GND	310	179	242	154	163	143
	GBRL	**100**	44	16	**12**	**9**	**5**

TABLE IV: Network degradation comparison across ratios on tortoise (lower is better)

Metric	Method	0.05	0.10	0.15	0.20	0.25	0.30
Conn.	CI	**0.0449**	0.0388	0.0386	0.0423	0.0454	0.0337
	GDM	0.0906	0.1006	0.1124	0.1265	0.1431	0.1633
	GERL	0.0754	0.0719	0.0613	0.0635	0.0651	0.0459
	GND	0.0704	0.0453	0.0655	0.0278	0.0235	0.0182
	GBRL	0.0596	**0.0376**	**0.0267**	**0.0199**	**0.0103**	**0.0078**
Edges	CI	352	303	272	259	240	192
	GDM	411	404	397	390	381	372
	GERL	395	357	328	305	279	220
	GND	342	275	263	219	192	156
	GBRL	**329**	**266**	**207**	**164**	**128**	**99**
Inv.Comp.	CI	**0.0141**	**0.0125**	**0.0116**	0.0123	0.0127	0.0132
	GDM	0.0185	0.0213	0.0250	0.0303	0.0357	0.0435
	GERL	0.0164	0.0159	0.0156	0.0156	0.0156	0.0156
	GND	0.0164	0.0147	0.0167	0.0143	0.0141	0.0132
	GBRL	0.0147	0.0127	0.0120	**0.0105**	**0.0097**	**0.0090**
Inv.Epid.	CI	2.6000	7.0000	7.0000	7.0000	7.0000	7.0000
	GDM	3.6176	3.6176	3.6176	3.6176	3.6176	3.6176
	GERL	3.1304	3.1304	3.5938	3.5200	3.2727	5.8125
	GND	3.6176	2.3947	3.6176	2.5000	**2.3077**	1.2857
	GBRL	**3.2182**	**2.2593**	**2.2593**	**2.2593**	2.3750	**1.2000**
MaxComp.	CI	**11**	21	21	21	9	9
	GDM	38	38	38	38	38	38
	GERL	28	28	**17**	24	13	11
	GND	38	27	38	26	9	7
	GBRL	34	**20**	20	**20**	**5**	**5**

TABLE V: Network degradation comparison across ratios on weaver (lower is better)

Metric	Method	0.05	0.10	0.15	0.20	0.25	0.30
Conn.	CI	**1.6049**	**0.6213**	**0.3704**	0.2798	0.2252	0.2313
	GDM	2.5126	2.5808	2.4882	1.7142	2.6543	2.4091
	GERL	2.3973	2.4513	2.2707	1.7813	1.5488	1.5828
	GND	1.5579	0.8185	0.8121	0.8914	1.0346	1.1667
	GBRL	1.6109	0.6739	0.3710	**0.2760**	**0.1348**	**0.0778**
Edges	CI	**130**	103	86	76	59	53
	GDM	173	166	155	114	140	132
	GERL	168	161	141	122	105	98
	GND	133	112	93	89	84	81
	GBRL	132	**102**	**75**	**59**	**43**	**31**
Inv.Comp.	CI	0.5000	0.1667	0.1000	0.0909	0.0556	0.0556
	GDM	1.0000	1.0000	1.0000	0.2500	1.0000	1.0000
	GERL	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	GND	0.2500	0.2000	0.1250	0.1250	0.1250	0.1667
	GBRL	**0.3333**	**0.1111**	**0.0909**	**0.0625**	**0.0476**	**0.0417**
Inv.Epid.	CI	**5.2000**	**3.6000**	**2.8387**	**2.4615**	5.0455	5.0455
	GDM	7.4971	7.5843	7.8387	5.4298	7.9786	8.1591
	GERL	7.4286	7.3540	6.6950	6.3934	6.0762	6.0102
	GND	5.4524	5.0909	5.0909	5.0909	5.0909	5.0909
	GBRL	5.4167	4.9615	3.3409	3.0833	**3.3684**	**1.4000**
MaxComp.	CI	60	**33**	**21**	**10**	**8**	8
	GDM	61	58	55	49	48	45
	GERL	61	58	55	52	48	45
	GND	**54**	37	37	37	37	37
	GBRL	59	35	26	22	11	**5**

forms or matches strong baselines under different removal ratios, demonstrating that its effectiveness is not limited to a single network instance. Overall, these results indicate that the proposed method is robust across a wide range of intervention

TABLE VI: Death rates and standard deviations across various models

Method	GDM	GND	GERL	Our Method
SYN-SF-100	73.1±0.8	77.3±0.6	71.1±0.7	**62.2±2.0**
SYN-ER-100	79.1±0.0	80.0±0.1	79.6±0.3	**78.4±0.2**
SYN-SW-100	81.6±0.0	76.4±0.4	79.4±0.1	**76.2±0.2**
SYN-ER-C-100	78.2±0.4	77.5±0.1	77.8±0.5	**75.5±0.1**
SYN-SF-C-100	72.3±0.6	76.9±0.9	78.8±0.0	**67.0±1.3**
SYN-SF-500	69.3±0.5	75.8±0.4	80.1±0.1	**55.7±1.3**
SYN-ER-500	79.7±0.0	78.9±0.2	79.7±0.1	**75.0±0.3**
SYN-SW-500	80.0±0.0	78.6±0.2	80.1±0.1	**75.9±0.1**
SYN-ER-C-500	80.0±0.0	79.3±0.0	79.7±0.0	**76.8±0.1**
SYN-SF-C-500	76.1±0.2	74.1±0.3	76.1±0.3	**65.0±0.5**
SYN-SF-1000	66.3±0.5	74.8±0.4	82.1±0.2	**53.2±1.0**
SYN-ER-1000	79.4±0.0	79.9±0.2	79.7±0.1	**74.0±0.4**
SYN-SW-1000	79.0±0.0	77.6±0.3	82.1±0.1	**75.3±0.2**
SYN-ER-C-1000	81.0±0.0	80.3±0.0	79.5±0.0	**77.0±0.1**
SYN-SF-C-1000	77.1±0.2	73.1±0.1	74.1±0.2	**64.0±0.4**
netscience	73.9±2.1	58.9±3.5	58.7±2.6	**23.8±2.4**
weaver	78.9±0.2	65.4±5.2	44.4±12.3	**41.4±8.8**
tortoise	48.8±4.0	27.9±2.9	37.2±3.9	**20.5±1.9**
mammalia24	73.6±4.2	50.3±4.7	49.2±6.2	**37.7±5.9**

budgets.

3) Simulation of Disease Spread in Immunized Networks: The Susceptible-Infected-Recovered (SIR) model, as implemented within the EoN (EpidemicsOnNetworks) framework [52], [53], is utilized to simulate disease propagation across five synthetic network types and four real-world networks. Figures 5 and 6 display the cumulative death rate over time for the SIR simulations, comparing the performance of the GDM, GND, GERL, and the proposed method across synthetic and real-world datasets.

In synthetic networks, the proposed method proves particularly effective in scale-free and scale-free with community networks, where high-degree hubs play a crucial role in spreading infections. By strategically targeting these critical nodes, the proposed approach significantly lowers the cumulative death rate, outperforming all baseline methods. The random and small-world networks exhibit less variation among models, as their more uniform degree distribution and redundant short paths allow diseases to spread efficiently even after node removals. Real-world networks further validate these findings, with the proposed method achieving the lowest death rates across all datasets, particularly in N_netscience, where its ability to disrupt connectivity is most pronounced. A summary of overall cumulative death rate for all methods is provided in Table VI. These findings further underscore the robustness and adaptability of the proposed method in containing epidemic outbreaks across diverse network structures. These findings reinforce the adaptability of the proposed approach, demonstrating its ability to contain epidemic outbreaks, minimize mortality, and optimize intervention strategies across diverse network structures.

V. Conclusion

This study introduced a reinforcement learning-based framework that integrates graph representation learning to optimize network dismantling and epidemic mitigation strategies. Unlike existing approaches, the proposed method incorporates mortality risk into decision-making, enabling socially responsible and adaptive interventions. Extensive experiments on five synthetic and four real-world networks demonstrated that the proposed approach outperforms state-of-the-art models (GDM, GND, GERL) by achieving lower cumulative death rates, higher epidemic thresholds, and greater network fragmentation, particularly in scale-free and community-structured networks, where targeted interventions significantly weaken connectivity. The results confirm that removing high-degree hubs is highly effective in disrupting disease transmission, while small-world networks remain more resistant due to their homogeneous structure and redundant short paths. By leveraging reinforcement learning, the model dynamically learns optimal intervention policies, making it highly adaptable to diverse network structures.

Although the proposed method demonstrates promising results on networks up to 1,000 nodes, evaluating its performance on significantly larger networks (e.g., tens of thousands of nodes) remains an important direction for future work. Potential extensions include sampling-based GNN techniques, parallel processing, and approximate epidemic simulation methods to further improve scalability. Future work will focus on scaling the framework to larger and evolving networks, integrating additional real-world constraints, and exploring policy transferability across different application domains such as cybersecurity, epidemiology, and misinformation control. This research highlights the potential of AI-driven network dismantling as a scalable and intelligent approach for mitigating large-scale disruptions in complex systems.

References

[1] "Who coronavirus (covid-19) dashboard," (Accessed July 14, 2023), https://covid19.who.int/.

[2] Q. Kang, X. Song, X. Xin, B. Chen, Y. Chen, X. Ye, and B. Zhang, "Machine learning-aided causal inference framework for environmental data analysis: a covid-19 case study," *Environmental Science & Technology*, vol. 55, no. 19, pp. 13 400–13 410, 2021.

[3] T. Padma, "Covid vaccines to reach poorest countries in 2023 – despite recent pledges," *Nature*, vol. 595, no. 7867, pp. 342–343, 2021.

[4] Z. Dong, Y. Chen, T. S. Tricco, C. Li, and T. Hu, "Practical strategy of acquaintance immunization without contact tracing," in *Proceedings of the 13th IEEE International Conference on Social Computing and Networking (SocialCom)*, 2020, pp. 845–851.

[5] ——, "Hunting for vital nodes in complex networks using local information," *Scientific Reports*, vol. 11, no. 1, p. 9190, 2021.

[6] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, and T. Zhou, "Vital nodes identification in complex networks," *Physics Reports*, vol. 650, pp. 1–63, 2016.

[7] M. Lalou, M. A. Tahraoui, and H. Kheddouci, "The critical node detection problem in networks: A survey," *Computer Science Review*, vol. 28, pp. 92–117, 2018.

[8] D. Kempe, J. Kleinberg, and É. Tardos, "Influential nodes in a diffusion model for social networks," in *Automata, Languages and Programming: 32nd International Colloquium (ICALP)*, 2005, pp. 1127–1138.

[9] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978.

[10] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, 1966.

[11] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer Networks and ISDN Systems*, vol. 30, no. 1, pp. 107–117, 1998.

[12] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature Physics*, vol. 6, no. 11, pp. 888–893, 2010.

[13] J.-X. Zhang, D.-B. Chen, Q. Dong, and Z.-D. Zhao, "Identifying a set of influential spreaders in complex networks," *Scientific Reports*, vol. 6, no. 1, p. 27823, 2016.

[14] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature Physics*, vol. 6, no. 11, pp. 888–893, 2010.

[15] Z. Wang, Y. Zhao, J. Xi, and C. Du, "Fast ranking influential nodes in complex networks using a k-shell iteration factor," *Physica A: Statistical Mechanics and its Applications*, vol. 461, pp. 171–181, 2016.

[16] Y.-P. Wan, J. Wang, D.-G. Zhang, H.-Y. Dong, and Q.-H. Ren, "Ranking the spreading capability of nodes in complex networks based on link significance," *Physica A: Statistical Mechanics and its Applications*, vol. 503, pp. 929–937, 2018.

[17] F. Yang, R. Zhang, Z. Yang, R. Hu, M. Li, Y. Yuan, and K. Li, "Identifying the most influential spreaders in complex networks by an extended local k-shell sum," *International Journal of Modern Physics C*, vol. 28, no. 01, p. 1750014, 2017.

[18] L.-l. Ma, C. Ma, H.-F. Zhang, and B.-H. Wang, "Identifying influential spreaders in complex networks based on gravity formula," *Physica A: Statistical Mechanics and its Applications*, vol. 451, pp. 205–212, 2016.

[19] J.-L. He, Y. Fu, and D.-B. Chen, "A novel top-k strategy for influence maximization in complex networks with community structure," *PloS One*, vol. 10, no. 12, p. e0145283, 2015.

[20] F. Morone and H. A. Makse, "Influence maximization in complex networks through optimal percolation," *Nature*, vol. 524, no. 7563, pp. 65–68, 2015.

[21] G. Zhao, P. Jia, C. Huang, A. Zhou, and Y. Fang, "A machine learning based framework for identifying influential nodes in complex networks," *IEEE Access*, vol. 8, pp. 65 462–65 471, 2020.

[22] E.-Y. Yu, Y. Fu, X. Chen, M. Xie, and D.-B. Chen, "Identifying critical nodes in temporal networks by network embedding," *Scientific Reports*, vol. 10, no. 1, p. 12494, 2020.

[23] M. Khajehnejad, A. A. Rezaei, M. Babaei, J. Hoffmann, M. Jalili, and A. Weller, "Adversarial graph embeddings for fair influence maximization over social networks," *arXiv:2005.04074*, 2020.

[24] Z. Dong, Y. Chen, T. S. Tricco, C. Li, and T. Hu, "Ego-aware graph neural network," *IEEE Transactions on Network Science and Engineering*, vol. 11, no. 2, pp. 1756–1770, 2024.

[25] E. Khalil, H. Dai, Y. Zhang, B. Dilkina, and L. Song, "Learning combinatorial optimization algorithms over graphs," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[26] Z. Dong, Y. Chen, C. Li, T. S. Tricco, and T. Hu, "Integrating graph and reinforcement learning for vaccination strategies in complex networks," *Scientific Reports*, vol. 14, no. 1, pp. 1–13, 2024.

[27] D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang, and T. Zhou, "Identifying influential nodes in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 4, pp. 1777–1787, 2012.

[28] D.-B. Chen, H. Gao, L. Lü, and T. Zhou, "Identifying influential nodes in large-scale directed networks: the role of clustering," *PloS One*, vol. 8, no. 10, p. e77455, 2013.

[29] L. Lü, Y.-C. Zhang, C. H. Yeung, and T. Zhou, "Leaders in social networks, the delicious case," *PloS One*, vol. 6, no. 6, p. e21202, 2011.

[30] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," in *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining*, 2010, pp. 261–270.

[31] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009, pp. 199–208.

[32] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2001, pp. 57–66.

[33] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, p. 026113, 2004.

[34] F. Morone and H. A. Makse, "Influence maximization in complex networks through optimal percolation," *Nature*, vol. 524, no. 7563, pp. 65–68, 2015.

[35] M. Grassia, M. De Domenico, and G. Mangioni, "Machine learning dismantling and early-warning signals of disintegration in complex systems," *Nature Communications*, vol. 12, no. 1, p. 5190, Aug 2021.

[36] X.-L. Ren, N. Gleinig, D. Helbing, and N. Antulov-Fantulin, "Generalized network dismantling," *Proceedings of the National Academy of Sciences*, vol. 116, no. 14, pp. 6554–6559, 2019.

[37] C. Fan, L. Zeng, Y. Sun, and Y.-Y. Liu, "Finding key players in complex networks through deep reinforcement learning," *Nature Machine Intelligence*, vol. 2, no. 6, pp. 317–324, 2020.

[38] D. Han, S. Jin, and C. Li, "Controlling epidemics with information dynamics: A neural ode approach," *IEEE Transactions on Computational Social Systems*, 2025.

[39] G. Mei, Y. Cai, S.-S. Zhang, Y. Huang, C. Liu, and X.-X. Zhan, "Modeling coupled epidemic-information dynamics via reaction-diffusion processes on multiplex networks with media and mobility effects," *IEEE Transactions on Network Science and Engineering*, vol. 13, pp. 3369–3390, 2025.

[40] P. ERDdS and A. R&wi, "On random graphs I," *Publ. Math. Debrecen*, vol. 6, no. 290-297, p. 18, 1959.

[41] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[42] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[43] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using networkx," in *Proceedings of the 7th Python in Science Conference*, 2008, pp. 11–15.

[44] P. Sah, L. O. Singh, A. Clauset, and S. Bansal, "Exploring community structure in biological networks with random graphs," *BMC Bioinformatics*, vol. 15, no. 1, pp. 1–14, 2014.

[45] M. E. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Physical Review E*, vol. 74, no. 3, p. 036104, 2006.

[46] R. E. Van Dijk, J. C. Kaden, A. Argüelles-Ticó, D. A. Dawson, T. Burke, and B. J. Hatchwell, "Cooperative investment in public goods is kin directed in communal nests of social birds," *Ecology Letters*, vol. 17, no. 9, pp. 1141–1148, 2014.

[47] S. Davis, B. Abbasi, S. Shah, S. Telfer, and M. Begon, "Spatial analyses of wildlife contact networks," *Journal of the Royal Society Interface*, vol. 12, no. 102, p. 20141004, 2015.

[48] P. Sah, K. E. Nussear, T. C. Esque, C. M. Aiello, P. J. Hudson, and S. Bansal, "Inferring social structure and its drivers from refuge use in the desert tortoise, a relatively solitary species," *Behavioral Ecology and Sociobiology*, vol. 70, pp. 1277–1289, 2016.

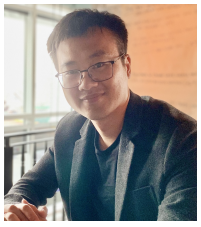
[49] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, "Epidemic thresholds in real networks," *ACM Transactions on Information and System Security*, vol. 10, no. 4, pp. 1–26, 2008.

[50] L. W. Beineke, O. R. Oellermann, and R. E. Pippert, "The average connectivity of a graph," *Discrete Mathematics*, vol. 252, no. 1-3, pp. 31–45, 2002.

[51] A.-L. Barabási, *Network Science*. Cambridge University Press, 2016.

[52] I. Z. Kiss, J. C. Miller, P. L. Simon *et al.*, "Mathematics of epidemics on networks," *Cham: Springer*, vol. 598, 2017.

[53] J. C. Miller and T. Ting, "EoN (Epidemics on Networks): a fast, flexible python package for simulation, analytic approximation, and analysis of epidemics on networks," *arXiv:2001.02436*, 2020.

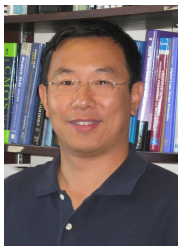


Zhihao Dong received the B.Eng. and M.Eng. degrees from Chongqing University, Chongqing, China, and the Ph.D. degree from Queen's University, Kingston, ON, Canada. His research interests include the intersection of machine learning and network science, network spreading processes, and social network analysis.

Yuanzhu Chen received the B.Sc. degree from Peking University in 1999, and the Ph.D. degree from Simon Fraser University in 2004. He has been a Professor of computing science since 2005 and is currently affiliated with School of Computing, Queen's University. From 2004 to 2005, he was a Postdoctoral Researcher with Simon Fraser University. In 2005, Dr. Chen joined Memorial University as Assistant Professor. While at Memorial, he was the Deputy Head for Undergraduate Studies and for Graduate Studies from 2012 to 2015 and from 2016 to 2019, respectively, and Department Head from 2019 to 2021. He then joined Queen's School of Computing in 2021. Dr. Chen's research interests include complex networks, computer networking, online social networks, mobile computing, graph theory, and evolutionary computation, with funding from national agencies and various university programs and awards. He was a recipient of the President's Award for Distinguished Teaching in 2018, and is an IEEE Senior Member.

Somayeh Kafaie received the M.Sc. degree in Computer Engineering (Software) from Iran University of Science and Technology, Iran (2011), and the Ph.D. degree in Computer Engineering from Memorial University of Newfoundland, Canada (2017). She was a postdoctoral fellow in the Department of Computer Science at Memorial University (2017–2019) and in the Faculty of Computer Science at Dalhousie University (2020–2021). She has been an Assistant Professor in the Department of Mathematics and Computing Science at Saint Mary's University, Halifax, Canada, since 2019. Her research interests include complex networks, knowledge graphs, explainable AI, and bioinformatics. She has received several awards, including the Best Editor Award from IEEE Wireless Communications Letters, the Fellow of SGS distinction, and the GSU Award for Excellence in Teaching.

Qiao Kang is a postdoctoral researcher in the Climate and Energy Policy Program (CEPP) and the Department of Earth System Science at Stanford University. His work focuses on developing new tools, methods, and performance metrics to improve causal inference, interpretable machine learning, and other data-driven approaches for climate-related challenges. Before joining Stanford, he earned his Ph.D. in Environmental Engineering from Memorial University in Canada.



Cheng Li received his B.Eng. and M.Eng. degrees from Harbin Institute of Technology, Harbin, China, in 1992 and 1995, respectively, and his Ph.D. degree in Electrical and Computer Engineering from Memorial University, St. John's, Canada, in 2004. He is currently a Full Professor and Director of the School of Engineering Science of Simon Fraser University, Vancouver, Canada. His research interests include wireless communications and networking, communications signal processing, underwater communications and networks, and mobile ad hoc and wireless sensor networks. He is an IEEE Communications Society Distinguished Lecturer for the 2021-23 term. He is an associate editor of the IEEE Transactions on Communications, IEEE Internet-of-Things Journal, and the IEEE Network Magazine. He has served as the General Co-Chair of the ICNC'23, Q2SWinet'20, WINCOM'19, and AICON'19, and the TPC Co-Chair for the ICNC'20, ADHOCNETS'19, ACOSIS'19, WiCON'17, MSWiM'14, WiMob'11 and QBSC'10. He also served many times as a Technical Program Co-Chair for various technical symposia/tracks of international conferences, including the IEEE GLOBECOM, ICC, WCNC, and VTC. He is the recipient of the Best Paper Award in IEEE ICC'23, Globecom'17 and ICC'10, and the Technical Achievement Award of the IEEE Communications Society Communications Software Technical Committee in 2018. Dr. Li is a registered Professional Engineer (P. Eng.) in Canada and is a Senior Member of the IEEE and a member of the IEEE Communication Society, Vehicular Technology Society, Signal Processing Society, and Ocean Engineering Society.