Engineering Applications of Artificial Intelligence 127 (2024) 107362

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Multi-Agent Learning of Asset Maintenance Plans through Localised Subnetworks[☆]

Marco Pérez Hernández[a,*], Alena Puchkova[b], Ajith K. Parlikad[b]

[a] School of Computing and Creative Technologies, University of the West of England, BS16 1QY , Bristol, UK
[b] Institute for Manufacturing, University of Cambridge, CB3 0FS, Cambridge, UK

ARTICLE INFO

Keywords:
Agent-based maintenance
Network maintenance
Multi-agent reinforcement learning

ABSTRACT

Maintenance planning of networked multi-asset systems is a complex problem due to the inherent individual and collective asset constraints and dynamics as well as the size of the system and interdependencies among assets. Although multi-asset systems have been studied numerous times in the past decades, maintenance planning implications of the system's network characteristics have been barely analysed. Likewise, solutions that consider the network perspective suffer from scalability issues as a network-wide observability is assumed. This paper proposes a network maintenance planning approach based on the decomposition of the multi-asset network into fixed-size localised subnetworks. The overall network maintenance plan is produced by aggregating the subnetwork maintenance plans, which are computed independently via a multi-agent deep reinforcement learning (MARL) algorithm. The results are evaluated against a network-wide approach as well as the commonly-used individual approach. The paper also introduces a systematic approach to integrate the MARL resulting policy in a multi-asset agent-based model. Simulation results of several random asset networks and a large nationwide network infrastructure show that, although a network-wide approach outperforms, on average, other approaches considered, the localised subnetworks approach, provides an acceptable alternative in networks with small-world properties, without the need of a network-wide view.

1. Introduction

Maintaining critical infrastructure assets in a timely manner is key to prolonging their lifespan. This reduces the maintenance operation costs and the overall impact on the service delivery. Most of the attention in research and practice has been around individual asset maintenance and only recently, the sources of complexities of multi-assets systems have been considered for a system-wide maintenance planning (Petchrompo and Parlikad, 2019).

Although progress has been made to understand the collective nature of the maintenance planning problem, there is still a lack of studies covering the implications of maintenance planning in networked multi-assets systems as discussed in Section 2. These studies are relevant to support the management decisions in the context of large networks of assets aggregating subnetworks with multiple topologies such as those present in nationwide critical infrastructures, among others (Zio, 2007). This type of networked multi-asset systems bring additional challenges to the maintenance problem. On the one hand the interdependences among assets and on the other, the scale of the systems restrict the application of (global) network-wide approaches that consider the state of every individual asset and the resulting network dynamics.

This paper explores the use of network-specific maintenance planning with an aim to minimise maintenance costs and the impact on service provision. A network-specific planning approach captures the network characteristics and uses that information to identify opportunities for synergistic maintenance activities in a multi-asset system. Our previous work has shown that, although network-wide approaches might yield best results, compared to individual approaches and thanks to the global view of the system, when the multi-asset system grows, the computational resources needed to generate the plan also scales up (Pérez Hernández et al., 2022). Hence, the motivation of this study is to identify alternative approaches that consider the network characteristics of the multi-asset system, to generate an acceptable maintenance plan without the need of a global network view.

To address this challenge, this paper proposes a maintenance planning approach grounded in breaking down the networked multi-asset

[☆] This research work was funded by University of the West of England (UWE) VC Early Career Researcher Award (UCSC0082) and by the Engineering and Physical Sciences Research Council (EPSRC) through the BT Prosperity Partnership Project: Next Generation Converged Digital Infrastructure under Grant (EP/R004935/1).

[*] Corresponding author.
E-mail address: marco.perezhernandez@uwe.ac.uk (M. Pérez Hernández).

https://doi.org/10.1016/j.engappai.2023.107362
Received 12 June 2023; Received in revised form 31 August 2023; Accepted 23 October 2023
Available online 3 November 2023

system in fixed-size subnetworks and solve separately every subnetwork maintenance plan. This task is approached by framing the planning as a multi-agent reinforcement learning (MARL) problem (Busoniu et al., 2010). The paper also introduces a systematic approach to integrate the MARL resulting policy in an agent-based model that enables simulation of the dynamics of a networked multi-asset system. This approach is evaluated against a network-wide approach and common preventive and corrective individual approaches.

The contents of the paper is structured as follows. The relevant literature on multi-asset maintenance planning is reviewed in Section 2, mainly considering nationwide infrastructures where complex networks characteristics are clear. The formulation of the network maintenance problem is presented in Section 3. A comprehensive Network-wide approach is explained in Section 4. As alternative to this approach, a novel localised approach is introduced in Section 5. This approach uses only information of the local subnetworks to produce the maintenance plan. As this approach relies on Multi-Agent Reinforcement Learning (MARL) to learn the optimal policy, the mechanism to integrate MARL policies into a networked-assets agent-based model is also proposed in Section 6. The network-specific approaches are evaluated against alternatives for several random networks and a large nationwide infrastructure network. The context of evaluation is provided in Section 7. The evaluation results are presented and discussed in Section 8. Finally, future work and conclusions are drawn in Section 9.

2. Multi-asset maintenance planning

In the last years, there has been significant attention to the maintenance planning of multi-asset systems. These are systems of several homogeneous or heterogeneous assets that might depend on each other (Petchrompo and Parlikad, 2019). These dependencies can be physical or logical and these are the source of the existence of networks of assets. Although a network perspective is not always considered in the study of multi-asset systems, this perspective brings tools to capture the static and dynamic properties of the system to better understand its behaviour (Vespignani, 2018). This behaviour is key to determine best maintenance approaches. The network structure and dynamics of the systems can be considered among other factors, at different planning levels. For example, a maintenance strategy encompasses a wide organisational perspective, as highlighted in industry standards such as ISO 55001 (ISO (International Organization for Standardization), 2014), and strategies can be classified according to distinctive approaches of *corrective (breakdown)*, *preventive* and *predictive* maintenance (Poór et al., 2019). Likewise, multiple objectives such as maximising system availability, minimising maintenance costs, or risk of failures, can be considered among these maintenance strategies (Pinciroli et al., 2023).

Civil infrastructures are usually seen as networks with multiple maintenance planning drivers. Researchers have structured a solution based on dynamic programming to plan the maintenance of a bridge network, considering safety and cost objectives (Frangopol and Liu, 2007). Their approach aims to reach optimal solutions, firstly at individual level and secondly at network level. Similarly, minimisation of pavement costs while maintaining the quality requirements has been achieved with a multi-objective optimisation model (Meneses and Ferreira, 2012). Moreover, a model combining analytical and numeric techniques for multi-component multi-system networks is presented in Liang and Parlikad (2020). The model introduces a genetic algorithm where the mutation is based on an agglomerative procedure. All together solving the Markov Decision Process (MDP) for maintaining the entire system. The model is demonstrated in a two-bridge network obtaining reductions of overall maintenance costs. A weighted random forest algorithm also enables maintenance planning decisions in a road network (Han et al., 2022). The decisions are based on the sequence of the conservation plan, the time after maintenance, and specific maintenance indicators.

Maintenance plan optimisation has also been a recurrent challenge in transport industries. Maintenance of railway networks is studied in Mohammadi and He (2022). Authors use double deep Q-networks reinforcement learning to find the policy that optimises maintenance and renewal planning. The aim is to reduce costs and failure occurrence in large railway networks. The case study focused on a 5-year plan for a railway network of 4000 miles, that was discretised into segments as part of the state model. Mixed integer programming has enabled the optimisation of the maintenance activities while taking into account the railway traffic, at the long-term, and also the required cycles for maintenance (Lidén and Joborn, 2017). Another proposal aims to formulate a plan for the chinese railway infrastructure, by considering both, the maintenance requirements and the constraints derived from the railway network schedule (Zhang et al., 2020). These authors use a heuristics algorithm built on the Lagrangian relaxation process for solving the joint optimisation problem. Simulation of these railway asset networks has been also addressed by researchers. Fleets of assets have been simulated to study the condition-based maintenance of critical components (Márquez et al., 2023). Their focus is on optimising the maintenance activities for each asset's critical components based on their remaining useful life (RUL). Moreover, vehicle fleets have been studied as a multi-objective problem (Wang et al., 2022). In their work, authors use the predicted RUL of the vehicle components to compute the maintenance schedule of the vehicle fleet. The schedule is obtained by using a tailored evolutionary algorithm that seeks to reduce repair costs, improve safety and reduce downtime.

In addition to specific asset condition, different aspects including geography, customer needs and risk, among others, are also considered in maintenance planning of different infrastructures. Researchers of the water distribution networks, have proposed a Maintenance Grouping Optimisation model based on genetic algorithms (Li et al., 2014). This approach enables the grouping of adjacent pipelines to plan maintenance, showing cost benefits compared to ungrouped plans. Empirical, single and multi-objective optimisation approaches for the maintenance of pipeline networks have been also evaluated with evolutionary algorithms (Chu et al., 2022). In this case, the "Non-dominated Sorting Genetic Algorithm II" (NSGA-II) with an elitist selection, obtains the maintenance plan that contemplates costs, reliability and overall network health. Online and Offline Deep Q-Networks reinforcement learning has been also used in maintenance planning of water pipes (Bukhsh et al., 2023). Authors train an agent to learn the optimal rehabilitation policies based on pipe deterioration profiles. A risk-oriented perspective has been adopted for the study of natural gas distribution systems in Italy (Leoni et al., 2019). By considering the relevant risks, authors are able to optimise the maintenance time for the components of the analysed gas monitoring stations.

Maintenance planning of power networks has also received significant attention. Customer requirements and long-term economic savings are considered in microgrids (Moradi et al., 2019). Researchers propose a multi-attribute decision making model that enables identification of critical components and their failure rate over time. Focusing more on power distribution networks, another proposal considers not only maintenance planning but also day-ahead scheduling (Matin et al., 2022). In their work, authors demonstrate that an approach, based on the Epsilon-constraint method, can lead to significant reductions in operating costs and improved reliability of multi-microgrids. Authors of Rocchetta et al. (2019) study a deep reinforcement learning framework for planning maintenance and operations of a power grid system. An agent was trained to select the combined operations and maintenance actions. Solutions were found to be comparable with true optimals for a scaled-down grid scenario.

Approaching the maintenance problem as a Markov Decision Process has enabled researchers to use reinforcement learning (RL) algorithms. Researchers have developed a RL model based on neural networks for the maintenance of pavement (Yao et al., 2020). The single-agent deep Q-learning solution approach achieves long-term

Table 1
Summary of reviewed works covering multi-asset infrastructure maintenance planning and selected works using reinforcement learning.

Reference	Type of system	Planning objectives	Solution approach	System characteristics
Frangopol and Liu (2007)	Bridge Network	Safety and cost objectives	Multi-objective Optimisation Two-phase Dynamic programming & Monte Carlo.	Six-bridge regional network
Li et al. (2014)	Water distribution networks	Asset (pipe) replacement & costs	Grouping Optimisation & Genetic algorithms	Adjacent pipelines
Moradi et al. (2019)	Microgrids	Critical components, failure rate & budget	Multi-attribute decision making	46-component network
Matin et al. (2022)	Multi-Microgrids	Operating costs & Minimise interruptions	Bi-level Epsilon-constraint method	69-bus distribution network
Leoni et al. (2019)	Natural gas distribution systems	Optimising maintenance time	Bayesian Risk-oriented method	59 Gas monitoring stations
Wang et al. (2022)	Vehicle fleets	maintenance downtime, workload, costs & No. of failures.	Multi-objective evolutionary algorithm	20 vehicles (Taxi) with 2 workshops
Yao et al. (2020)	Pavement maintenance	Long-term cost-effectiveness	Single-agent deep Q-learning	1974 segments of expressways with no network-level constraints.
Chen and Wang (2023)	K-component mechanical systems	Lifecycle costs reduction	Deep Q-learning	Structural dependencies
Zhang and Si (2020)	Multi-component systems	Cost minimisation with dependent competing risks	Deep reinforcement learning	2-component & 12-component systems mentioned
Lei et al. (2022)	Regional deteriorating bridges	Minimise risks & costs Optimising regional strategy	Deep Q-networks	Highway infrastructure with slab, T-shaped and beam bridges.
Kuhnle et al. (2019)	Parallel production systems	Downtime reductions & opportunity costs	Holistic approach with Multi-agent Q-learning	Identical machines with upstream buffers.
Rodriguez et al. (2022)	Parallel production systems	Uncertainty of multiple machine failures	Markov game with Proximal Policy Optimisation MARL	3 &5 Identical machines
Thomas et al. (2021)	Radio access networks	Costs and Network availability	Multi-Agent Actor Critic MAAC	9-asset single grid topology
Rocchetta et al. (2019)	Scaled-down power grid	Maximise expected profit with random uncertainties.	Deep Q-Networks	2 generators, 5 cables, 2 sources and 2 loads.
Bukhsh et al. (2023)	Water distribution system	optimise average costs and reduce failure probability	Deep Q-Networks	16-pipe system
Mohammadi and He (2022)	Railway Network	Long-term cost effectiveness & risk reduction	Double Deep Q-Networks with prioritised replay	Class I freight railroad of discretised 4000 miles.

cost-effectiveness in the context of Ningchang and Zhenli expressways. Deep Q-learning is also applied in a K-component mechanical systems with structural dependencies, being able to bring policies that reduce system lifecycle costs (Chen and Wang, 2023). Multi-component systems are also approached using Deep Reinforcement Learning (Zhang and Si, 2020). These authors incorporate dependent and competing risks to the problem formulation. Deep Q-networks (DQN) model is also applied for planning maintenance of regional deteriorating bridges (Lei et al., 2022). Their model optimises regional life-cycle strategies according to various budget constraints.

So far, there is a limited volume of works that have adopted a multi-agent perspective, particularly in the maintenance planning of nationwide infrastructure. A multi-agent environment is considered for the optimisation of the maintenance of parallel homogeneous working machines (Kuhnle et al., 2019). In this work, opportunistic agents learn, via proximal policy optimisation, when to trigger maintenance actions as close as possible to breakdown hence reducing downtime and maintenance costs. Although the paper considers interdependencies and interactions between the different machines in the production system, the topology studied is simple, based on parallel machines. Another multi-agent approach is used to coordinate maintenance scheduling among a set of partially-observed machines (Rodriguez et al., 2022). A mix of sequential/parallel and centralised/distributed agent architectures are analysed. The problem is approached as a Markov game that is tackled with the proximal policy optimisation algorithm. Likewise, the Multi-Agent Actor Critic (MAAC) framework has been used to plan the maintenance of radio access networks with a single grid topology (Thomas et al., 2021). Although network dependencies in a particular network are considered, there is no indication of how this could work in different network configurations.

Based on this review, the maintenance of multiple assets has generated more interest in the context of some civil infrastructures such as bridges, power networks and less attention in other infrastructures such as telecommunications. The summary of works reviewed is presented in Table 1. Although network structures are implied when identifying dependencies, the topologies analysed are usually simple, with limited number of assets or simple sequential/parallel structures in the multi-asset systems. Furthermore, there is still limited research on the assessment of the benefits and trade-offs of different maintenance approaches for networked multi-asset systems.

3. Network maintenance problem

The goal of the maintenance planner is to identify the optimal plan for maintaining a portfolio of assets. In a network setting, not only the characteristics of the assets (network elements) but also the structure of the connections among assets (network topology) become relevant when considering the potential impact of the maintenance plan. The optimal plan should consider minimum maintenance costs but also minimum impact on the quality of services enabled by the assets. For simplification, this study considers a network of assets, where the asset heterogeneity is limited to the speed of the deterioration patterns, but following a common linear deterioration function. The focus is on the role of network topology, the effect on throughput, as the key quality indicator, and the overall cost per cycle. The total maintenance cost function is made up of the downtime cost, labour cost, lost life cost and cost of parts. Pérez Hernández et al. (2022) provide a detailed description and discussion of the cost function and the parameters used in this study.

To tackle this problem, well-known corrective and preventive approaches can be used to support decisions from an individual perspective of the assets. Beyond that, multi-asset approaches can incorporate the dynamics of multiple assets into the decision problem, however these approaches do not capture in detail the network properties of the system.

The following sections introduce two approaches that exploit those properties, enabling solutions that are tailored to every network of assets. For the sake of clarity, the approaches are explained in the context of a Telecommunications network, however, the abstractions used and solution approach enable to capture the dynamics of other complex network systems such as nationwide critical infrastructures i.e. transport, water, energy or others. Note this problem focuses on the network perspective of the multi-asset systems. This approach enables the formulation and use of techniques for analysis that are common to multiple domains, however, a comprehensive planning requires also elements that are specific of the domain.

4. Network-wide approach

This approach assumes complete observability of the relevant features of network. A centralised optimiser is fed, periodically, with snapshots of the condition of individual assets and the state of the traffic flows during a time window. In a Telecommunication network, the traffic flows pertain to the data packets being transported across multiple network equipment from the source of the data, i.e. servers, camera feeds, sensors, etc, to the consumers e.g. mobile phones, industrial computers, laptops, tv, cars, etc.

We represent the telecommunication network as nodes $v \in V$ linked by arcs $a \in A$ with a set $\{(k,l)\}$ corresponding to services, where traffic flow starts from a source node k and ends at a destination node l. Nodes that are likely to fail are denoted by $\bar{V} \subset V$. A number of control variables are defined to model traffic flow and maintenance decisions as follows:

$x^t_{k,l,a}$ - traffic amount flowing through arc a at time t from source k to destination l,

$$w^t_v = \begin{cases} 1, & \text{if node } v \in \bar{V} \text{ is shut down for predictive maintenance at time } t, \\ 0, & \text{otherwise}, \end{cases}$$

$$z^t_v = \begin{cases} 1, & \text{if predictive maintenance job on node } v \in \bar{V} \text{ starts at time } t, \\ 0, & \text{otherwise}. \end{cases}$$

To guarantee that decision variables behave as desired, the following constraints need to be introduced. Constraints (1) ensure that the sum of all traffic coming from node k is equal to traffic demand d^t_{kl} for each service. The sum of traffic coming from any intermediate node does not exceed the sum of traffic flowing to this node, see constraints (2). Constraints (3) imply that the sum of traffic going through arc a cannot exceed its capacity cap_a, as well as that traffic flow is not permitted on those arcs that are incident with the node undergoing maintenance (i.e. when $w^t_v = 1$).

A node $v \in \bar{V}$ on maintenance is shut down for the duration of predictive maintenance job t^{pred}_v, see (4). In this model we consider continuous maintenance without pre-emption as represented by constraints (5).

$$\sum_{a \in out(k)} x^t_{k,l,a} = d^t_{kl}, \tag{1}$$

$$\sum_{a \in in(v)} x^t_{k,l,a} \geq \sum_{a \in out(v)} x^t_{k,l,a} \quad \forall v \neq k, l, \tag{2}$$

$$\sum_{k,l} x^t_{k,l,a} \leq (1 - w^t_v) cap_a \quad \forall a \in in(v) \cup out(v), \tag{3}$$

$$\left(\sum_t w^t_v - t^{pred}_v \right) I^T_v = 0, \quad I^t_v = \sum_{s=0}^{t-1} z^s_v, \tag{4}$$

$$z^t_v \geq w^t_v - w^{t-1}_v \quad \forall t \geq 1, \quad z^0_v \geq w^0_v, \quad \sum_t z^t_v = 1. \tag{5}$$

The optimisation model aims to identify the best values of decision variables defined earlier that minimise the total cost consisting of maintenance cost, cost of traffic loss and rerouted traffic cost:

$$\begin{aligned} \text{Minimise } J = & \sum_{v \in \bar{V}} \sum_t p^t_v I^t_v C^{pred}_v + \sum_{v \in \bar{V}} \sum_t p^t_v (1 - I^t_v) C^{corr}_v \\ & + \sum_{v \in \bar{V}} \sum_t p^t_v (1 - I^t_v) \sum_{(k,l)} \sum_{a \in in(v) \cup out(v)} \sum_{s=t}^{t+t^{corr}_v} x^s_{k,l,a} \\ & + \sum_t \sum_{(k,l)} \sum_{a \in A} w_a x^t_{k,l,a} \end{aligned} \tag{6}$$

where p^t_v is failure probability of node $v \in \bar{V}$.

5. Local-networks (localised) approach

Due to scalability of assets in the network or specific deployment limitations, there are cases where full observability of the network cannot be guaranteed or there is not enough capacity to process and compute timely the data collected from the entire network. In these cases, alternative approaches considering only partial network information are necessary. The Local-networks (Localised) approach defines the network maintenance problem as a Multi-Agent Reinforcement Learning (MARL) problem.

The reinforcement learning framework enables an agent to learn, from the interactions with the environment, a sequence of actions that maximise a given cumulative reward (Sutton and Barto, 2018). A RL problem is formally defined as a Markov Decision Process (MDP) and RL algorithms aim to find a policy that drives the agent-decision making process (Sutton and Barto, 2018). MARL is an extension of the single-agent reinforcement learning (RL) problem where multiple agents are interacting with the environment and taking actions, hence potentially having influence on each other (Busoniu et al., 2010).

The overall network maintenance plan is built from the localised maintenance decisions that independent agents take, based on the observability of their local networks. Agents observe their environment and learn decentralised policies that seek to maximise individual rewards. Collectively, the aggregation of individual rewards yields a system-level reward. The idea of applying this approach is to determine the ability of agents to learn an acceptable policy and understand the magnitude of the compromise that maintenance planning decision-maker faces when a Network-wide approach is not feasible. The rationale for this approach is to reduce the dependency on the full network information to drive maintenance decisions. Mathematically the problem can be formulated as an adaptation of the stochastic game definition (Busoniu et al., 2010), as follows:

$$\Gamma = \langle S, Z, U, f, r \rangle \tag{7}$$

where Γ is the environment, of which S are the possible states, from which a number of n agents observes Z_i (See Section 5.1), such as $Z = Z_1, Z_2, \ldots, Z_n$. Every agent is able to take actions U, note that original MARL definition is simplified by assuming that $U_1 = U_2 = \cdots U_n$, in other words, all the agents have the same action space, which is discrete with u_0 *:Do nothing* and u_1 *: Start maintenance*. Likewise, f and r are functions that capture the transition probabilities and represent the collective reward, respectively. In this case the collective reward is a simple aggregation of the individual rewards of every agent: $r = \sum_{i=1}^{n} R_i$.

5.1. Networked asset state

The environment state is formed by continuous, discrete and network representation components. This is defined by the tuple: $Z_i =$

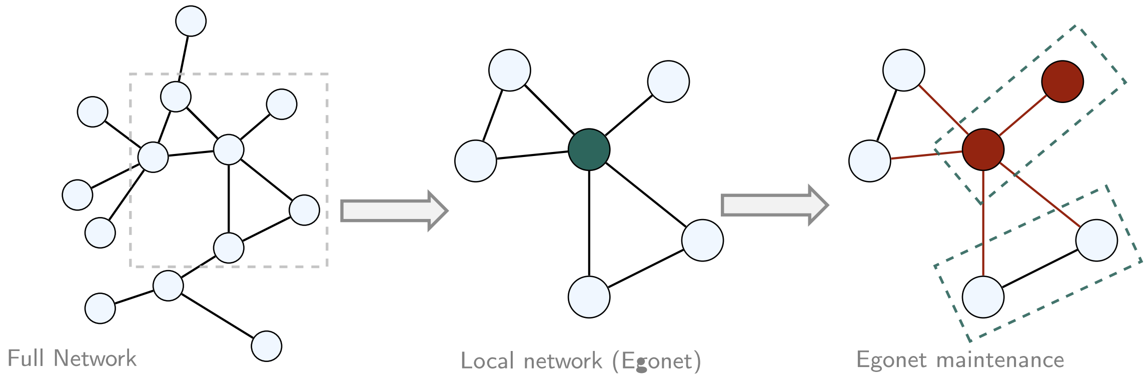

Fig. 1. Local network observation process. Local networks are extracted for each node of the network, representing assets and their links. Then agents are trained using information about nearby unavailable assets (red nodes) to learn when maintenance is more detrimental, according local network robustness metrics.

$\langle x, p, g\rangle$, where x is the current condition of the asset, p is the individual tracker of the maintenance state of the asset (*up, down* or *on maintenance*) and g is the local network state. Every agent is assumed to have visibility of the local network of the asset. This also known as an ego-centric perspective or *egonet* (Scott and Carrington, 2011). There is more information available (x and p) for the ego (focal) node and the number of nodes of the egonet varies depending on the depth. Both the size of the egonet and the information shared by neighbours are limited to control the communication overhead required in the learning process. As every agent is computing the maintenance plan for an egonet, this subnetwork serves as a limit for the quantity of assets considered in every maintenance plan. Similar approaches of using limited-range network centrality measurements have been used to address the complexity of characterising the structure of large networks in other domains (Ercsey-Ravasz and Toroczkai, 2010).

Fig. 1 illustrates the local network observation process. Starting with the identification of the *egonets* (left) to the exploration of different scenarios where the neighbouring nodes are undergoing maintenance (right), every agent observes the individual maintenance tracker of the local network of assets. This allows the agent to determine what assets are undergoing maintenance (red nodes in the figure) and uses this information to drive the policy learning process. Following a standard reinforcement learning process with Deep Q-learning, the agents are trained by observing their network state, taking individual actions and receiving rewards according to the state–action pair. As one of the main drivers of this approach is to offer a low-overhead localised alternative, the state space is transformed to a uniform continuous space. This is possible by computing the "health" of the local network based on the available edges at every time. This metric is regarded as network density (Newman, 2018): $\delta = e/v(v-1)$ which corresponds to the connected edges e over the possible connections $v(v-1)$. The rationale of using this metric is that it can inform the agent about the impact of going to maintenance at time t at local network level. Note there is no consideration of flows in the state space, this is also a design decision to reduce the overhead as live traffic is also computationally expensive to collect and process.

5.2. Reward function

The reward function of every agent ($R_{i,t}$) at a time step t is inversely derived from the cost function (Section 3). Moreover, an additional term is added to account for the local network information. When the u_1 (Start maintenance) action is selected, the agent is penalised proportionally according to the density of the egonet at the current step. This is multiplied by a factor β that considers the importance of the egonet information in the agent's reward. Accordingly the function is defined as:

$$R_{i,t} = -1 \cdot \left(C_{i,t} + \beta \cdot \left(1 - \frac{\delta_t}{\delta^*} \right) \right), \delta^* > 0, \delta_t \in [0, \delta^*] \tag{8}$$

where $C_{i,t}$ is the total cost of maintenance calculated according to Section 3. The density δ_t of the (egonet) subnetwork at time t, is influenced by the condition of the assets that are part of the subnetwork. Particularly, if assets have failed, δ_t will be lower than the expected density δ^* when all assets are working. The expected effect of this function is that the agent learns to balance the decision of when the maintenance is due because the asset condition has deteriorated while also discouraging maintenance when there are other assets, within its local network, on maintenance or failed.

5.3. Independent DQN

A reinforcement learning (RL) algorithm finds the set of actions, namely policy π, that maximises the agent's cumulative reward. Algorithms are normally suited for a particular environment, including specific action and state spaces. Independent Deep Q-Networks (I-DQN) (Tampuu et al., 2017) is a foundational algorithm that has been used in multi-agent environments with discrete action spaces. I-DQN is a multi-agent adaptation of the single-agent DQN algorithm that has been benchmarked in different environments (Mnih et al., 2015). As expected, in a multi-agent environment, this algorithm does not provide guarantees for convergence to an optimal global policy due to non-stationarity. However, similar approaches have been successfully used in two-player games (Foerster et al., 2016). This algorithm was selected for the suitability for the discrete action space, its simplicity and the decentralised nature which is aligned to the aim of learning a policy based only on the local subnetworks.

Algorithm 1 is derived from the multi-agent I-DQN adapted by Tampuu et al. (2017). It aims to find the policy π starting with the observation of the current environment state. Every agent selects an action by using the Q-value function that estimates the quality of each action $u_{i,t}$ at the given state s_t. To allow for exploration, instead of always picking the action that maximises the Q-value, the ϵ-greedy method allows to pick random actions with some frequency. This selection can be adjusted using the ϵ parameter. After taking each action agents store their experiences –reward obtained and new state– in an experience buffer D. The experiences are then sampled and used to approximate the Q function with non-linear neural networks that minimise a loss function which is defined by:

$$L(\theta) = \mathbb{E}\left[\left(r + \gamma \cdot \max_{a'} Q(s', u'; \theta') - Q(s, u; \theta) \right)^2 \right] \tag{9}$$

Here the Q-value *prediction* is subtracted from the Q-value *target*.

This algorithm addresses the problem of learning instability by sampling uniformly from an experience buffer (replay memory) hence avoiding local minima by considering uncorrelated experiences (Tampuu et al., 2017).

Algorithm 1: Independent DQN. Derived from (Tampuu et al., 2017). Starting from observation of the environment state, agents estimate Q-value of the actions and select one, keeping record of their experiences every time. Then deep neural networks are used to approximate the Q function by minimising the loss $L(\theta)$.

```
Data: Environment $\Gamma$
Result: policy $\pi$
Initialise experience buffer $D$;
Initialise parameters for Q-networks (weights $\theta$) of each agent;
for $t = 1$ to max_steps do
    $S_t \leftarrow$ Obtain present state ;
    // For each agent obtain action
    for $a_i \in A$ do
        $Z_i \leftarrow$ observe $Si$ ;
        $u_{i,t} \leftarrow$ Select action based on $\epsilon$-greedy policy using
          $Q_i(s_t, u_{i,t}; \theta_i)$ ;
        push $u_{i,t}$ to $U_t$ ;
    end
    $(r_t, S_{t+1}) \leftarrow$ execute $U_t$ ;
    $D \leftarrow$ store $(r_t, S_{t+1}, U_t)$ ;
    sample $D$ ;
    for $a_i \in A$ do
        Approximate Q-function by minimising $L(\theta)$ (Eq. (9)) ;
        Update Q-networks ;
    end
end
```

The I-DQN algorithm is used for the particular Network Maintenance problem, by training the algorithm with different random networks varying in their size and topology. Egonet density metrics are calculated and composed as part of the environment state observed by the training agents. Once agents are trained offline, the maintenance plan is obtained by evaluating their policies in a set of specific networks and using these plans in an agent-based simulation model.

6. Reinforcement learning in the multi-asset agent-based model

NAssets.jl is an agent-based model and simulator (ABMS) introduced in Pérez Hernández et al. (2022) that enables modelling and simulation of networked multi-assets systems. This model represents assets as agents whose condition deteriorates along the time, following a defined model. Likewise, NAssets.jl allows for configuration of network topologies identifying assets as vertices and the edges among them enable traffic flows. This simulator enables the introduction of an agent-based control system that manages the maintenance operations and routing of the underlying network of assets. This control systems could be defined by a single agent or by several arranged in their own control network. As part of this work, NAssets.jl model is extended to enable integration of offline Network-specific approach described in Sections 4 and 5.

Network-specific maintenance approaches rely on an offline planning phase, which uses available network data to determine the maintenance plan. Network topology and condition deterioration functions are used as starting points in both Network-wide and Localised approaches. In the Localised approach the complete topology is only necessary when evaluating the learned policy on the network of interest, thus subnetworks are obtained around every critical asset. Once the maintenance plan is generated the NAssets.jl model simulates traffic and condition deterioration dynamics during a defined observation time.

The integration of Reinforcement Learning approaches into Agent-based Models (ABMs) has been identified as a way to support decision-making processes within the simulation of complex systems. Similar techniques have been explored in domains different to network maintenance planning, for example in Vargas-Pérez et al. (2023) and Lee et al. (2017). For network maintenance, MARL processes are integrated into ABM according to the flow presented in Fig. 2.

There are two main phases in this flow. During the first *Offline Planning* phase, the MARL agents are trained according to the process described in Section 5.3. Thus, the networks of interest are pushed to the agents that use the reward (Section 5.2) function to drive the policy learning process that determines the maintenance actions. Once a policy is learned by the agents, test networks are used to generate maintenance plans for a required period. Resulting plans are consolidated in a single plan with the form of a $m \times n$ binary matrix, with m assets and n time steps and set to 1 when maintenance is due.

During the *Network Dynamics Simulation* phase, the consolidated plan is loaded into NAssets.jl which is also configured according to the network topology, the service portfolio supported by each network, the condition deterioration model for the assets and the traffic dynamics parameters. At start, the ABM configures maintenance activities and traffic re-routing as events in line with the input plan. The agent-based control system monitors asset's condition and acts according to events planned.

7. Case study: Multi-asset networks in nationwide digital infrastructure

The nationwide digital infrastructure is a multi-asset system where routers, mobile antennas, ad hoc computing resources and many other devices enable data packet transport along the country. This infrastructure is also a large complex network of networks that is carefully designed considering several requirements such as performance, quality, reliability and cost-efficiency. Particularly, it is expected that data packets across the network only travel a few hops until the destination, this is known as the small-world effect (Newman, 2018). Likewise, others have highlighted that traffic flows within these type of networks follow a scale-free model (Pastor-Satorras and Vespignani, 2004).

The infrastructure makes possible the transport of data between providers and consumers. Service requirements specify data transfer expectations from individual and business customers. Likewise, mobile or broadband operators use the nationwide digital infrastructure to support their own service portfolio (Amin et al., 2000). There is a Service Level Agreement (SLA) for each service that also includes Key Performance Indicators (KPIs), facilitating evaluation of the delivered quality of service against specification (Kosinski et al., 2008). As multiple KPIs are monitored depending on the service, the focus of this case is on one of the most common: Throughput, which indicates the rate of data packets delivered over time from end to end (providers to consumer).

Although the network perspective is not constrained to a particular planning level, the case focus is on the tactical maintenance planning. This planning assumes a stable network of assets and a set of fixed contracted services according to the network capacity for a medium term period, e.g. six to twelve months. A challenging task at this level is to balance the maintenance costs while keeping an adequate quality of service across the infrastructure, built from geographically distributed assets. The infrastructure follows a hierarchical architecture organised in network segments with different technologies and protocols (Tanenbaum, 2003). Access networks enable users, either data providers or consumers, to join the network, while metro/regional networks connect specific geographical areas to the core/backbone network which ensures national long-distance data packet transfer (Stavdas, 2010).

The environment for evaluation of the localised maintenance approach is motivated by the characteristics and dynamics of the UK's nationwide digital infrastructure. At the small scale, random networks are generated to resemble some of the networks present in this infrastructure. Particularly, the Barabasi–Albert (BA) model (Barabási and Albert, 1999) facilitates the generation of scale-free networks and the

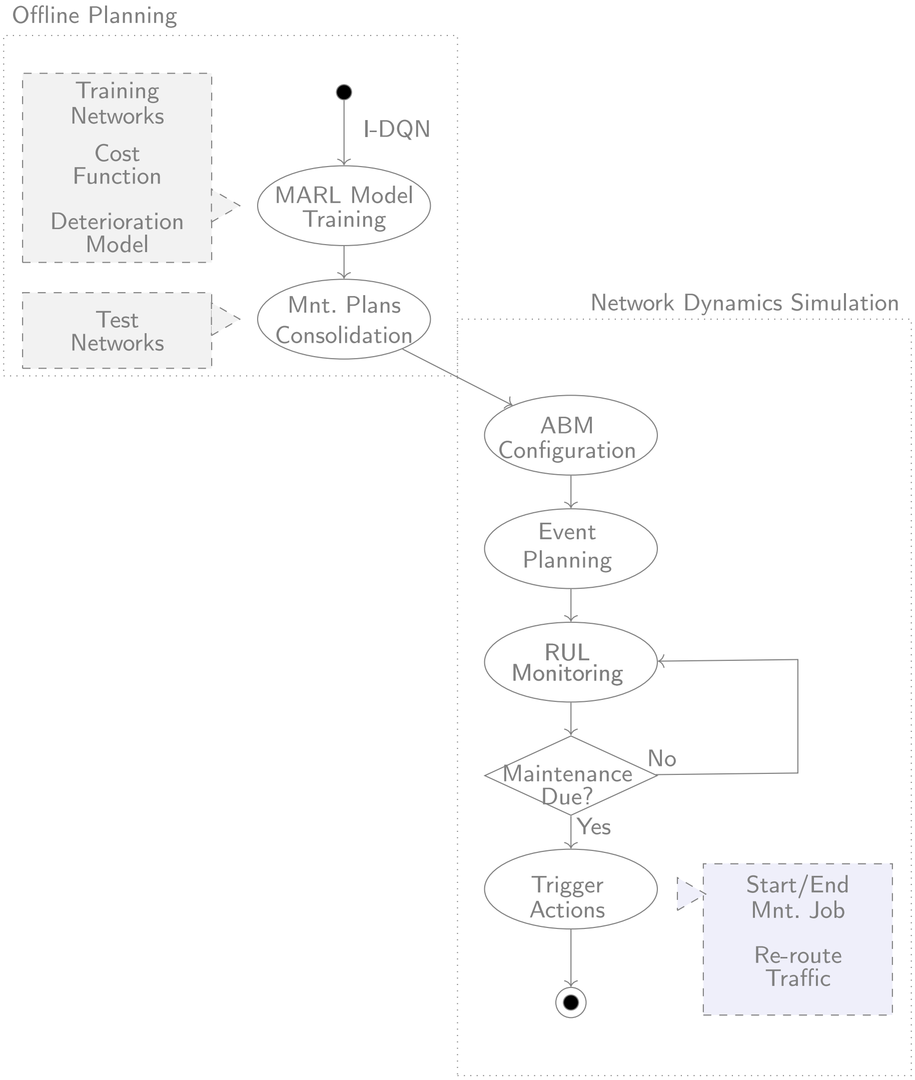

Fig. 2. Integration of Models: Multi-Agent Reinforcement Learning (MARL) of network maintenance plans and networked multi-asset agent-based (ABM) simulation. MARL agents are trained offline to learn a policy that is later used to generate maintenance plans of specific test networks. These plans are pushed to the ABM simulator enabling examination of networked multi-asset system dynamics for the given plan.

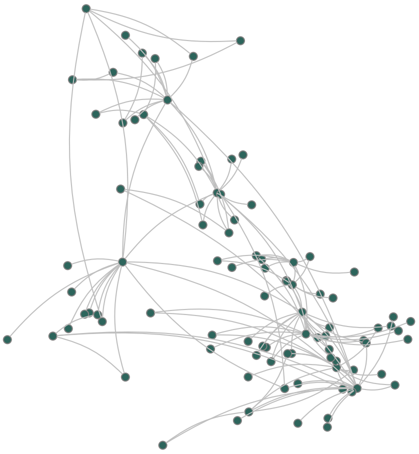

Fig. 3. Nationwide backbone network. Nodes are network elements distributed geographically across the country. Exact geographic location of the nodes has been randomised.

Watts–Strogatz (WS) model (Watts and Strogatz, 1998) is used for networks that exhibit the small-world properties. At the large scale, the UK's metro-core network presented in Fig. 3 is used as an example of a backbone network existing in the digital infrastructure. This network comprises 92 nodes and 131 links. Note that spatial location of nodes has been randomised.

The evaluation considers quality and cost requirements across the networks identified. These networks facilitate the analysis by comparing performance of the network-wide maintenance approach (Pérez Hernández et al., 2022), as well as the individual corrective and preventive approaches. For every network, a set of services is created, such that every service requires the data packet transport between a source and a destination at a given expected rate. The quality indicator is then based on the reduction of throughput owing to the maintenance activities. The cost is calculated according to the parameters presented in Section 3, with values 5, 10 and 20 for three levels (*low*, *medium* and *high*) for each cost component. The Localised approach is implemented using Julia Reinforcement Learning library (Tian and other-contributors, 2020). The parameters of the training process are presented in Table 2.

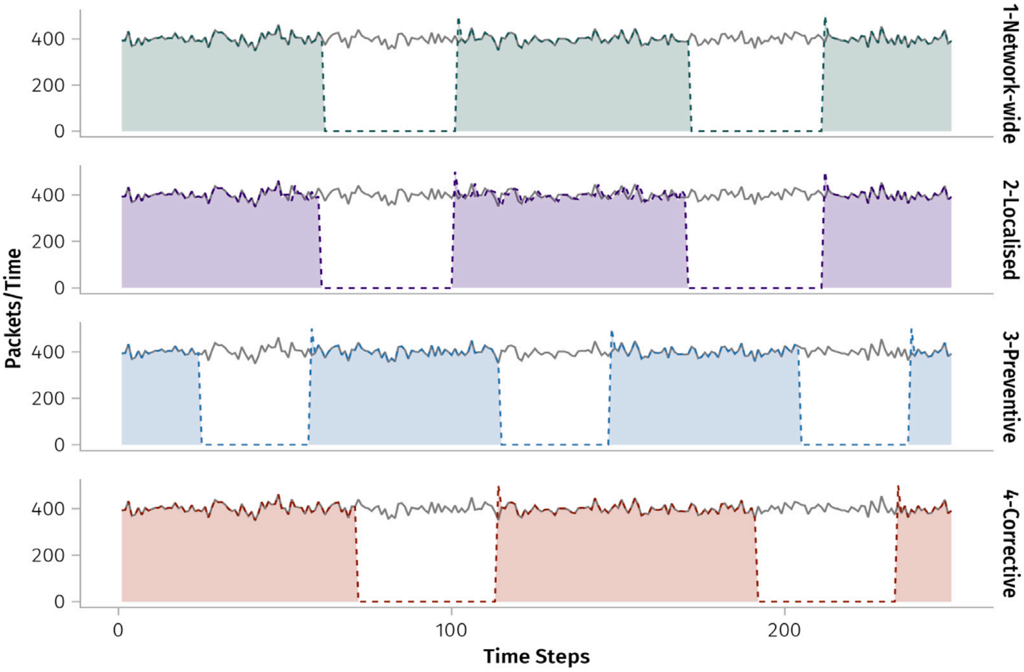

Fig. 4. Impact of maintenance activities on service quality. Extreme example of throughput (packets/time) reduction against expected (Grey line) in a simulated service. Due to the lack of alternative paths when active network elements are on maintenance, throughput drops to 0. Differences in timing of maintenance according to each approach.

Table 2
Independent DQN training process parameters.

Parameter	Value	Units
Inner neural network layers	2 × 64	
Initialisation	Glorot normal	
ϵ-greedy policy	0.06	
Episode length	1000	steps
Episodes training	30 × 10^6	steps
Batch size	600	
Learning rate	0.01	
Target network update frequency	150	steps
Replay history	1000	
Input network topologies	10 × random BA 10 × random WS	

8. Results & discussion

Maintenance activities might affect the availability of the network elements and hence the routes used to transport data. Unavailable network elements –e.g. routers or switches– can cause packets to be dropped or delayed and ultimately affect the quality of the services. Thus, maintenance activities have different impact on the overall service throughput depending on the availability of alternative paths to reroute traffic and the timing of the maintenance activities. The most serious impact of the maintenance activities is presented in Fig. 4 when due to the lack of backup paths, the service throughput drops to zero. The figure shows the different timing of the maintenance activities, according to each approach. For example, if maintenance starts too early and there are no alternative paths, more maintenance activities are required during the same time frame as shown in the *preventive* approach.

The descriptive statistics of the throughput reduction due to maintenance, in the networks studied, are presented in Table 3. Moreover, Figs. 5–6 show the throughput reduction for 10 BA and 10 WS random networks. Each point in the plots represents the reduction of throughput of one of the services provisioned within one random network. As expected, the Network-wide approach yields the minimum average impact on the throughput with a mean of only 0.12 (reduction of the expected throughput) for BA networks and 0.22 for WS. The standard deviation shows the dispersion of the measured impact, of a given maintenance approach, among networks of similar characteristics. For the network-wide approach the standard deviation is 0.084 for BA and 0.097 for WS networks, showing significant dispersion of the measured impact although lower than the standard deviation observed in other approaches. This shows that impact of this maintenance approach is slightly more consistent than others, across the various sets of services and networks analysed. Although not fully shown in the plots, in few cases only, the Network-wide approach causes higher impact, on specific services running in BA networks, than other approaches. Particularly, the corrective approach is the best one in these cases. This might be due to the greater availability of alternative paths for certain nodes in BA networks hence the path chosen after a node fails leads to lower throughput reduction than the anticipated path chosen in the network-wide approach. More details are presented in Pérez Hernández et al. (2022).

Table 3
Descriptive statistics of throughput reduction due to maintenance, grouped per type of network.

	Barabasi–Albert		Watts–Strogatz		Backbone core	
	mean	std	mean	std	mean	std
Network-wide	0.115	0.084	0.221	0.097	0.143	0.095
Localised	0.241	0.130	0.257	0.128	0.156	0.102
Preventive	0.158	0.118	0.276	0.117	0.179	0.124
Corrective	0.180	0.112	0.311	0.109	0.229	0.156

For the BA networks, the greater throughput reductions are obtained by the Localised and Corrective approaches, respectively with 0.24 and 0.18 less throughput than expected. For WS random networks, the performance of the Network-wide decreases, making the Localised approach an acceptable alternative, slightly better than the preventive approach. However, the standard deviation is the highest with 0.12 for both types of networks. Note that as per statistics in Table 3, the performance of the Localised approach is stable across BA and WS

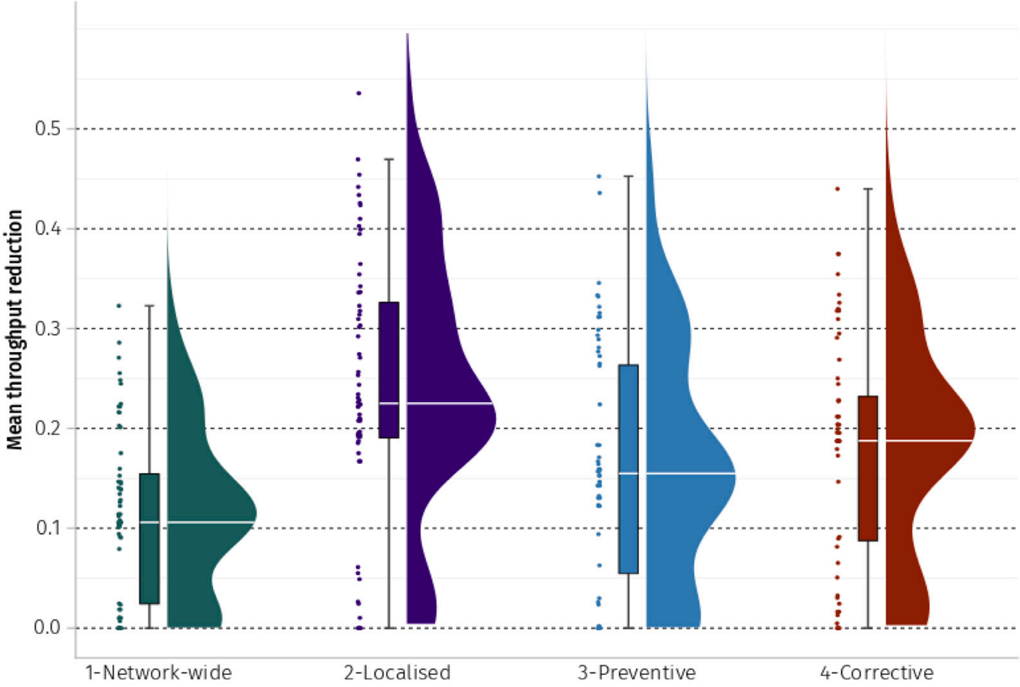

Fig. 5. Reduction of service throughput owing to maintenance operations in 10 Barabási–Albert networks.

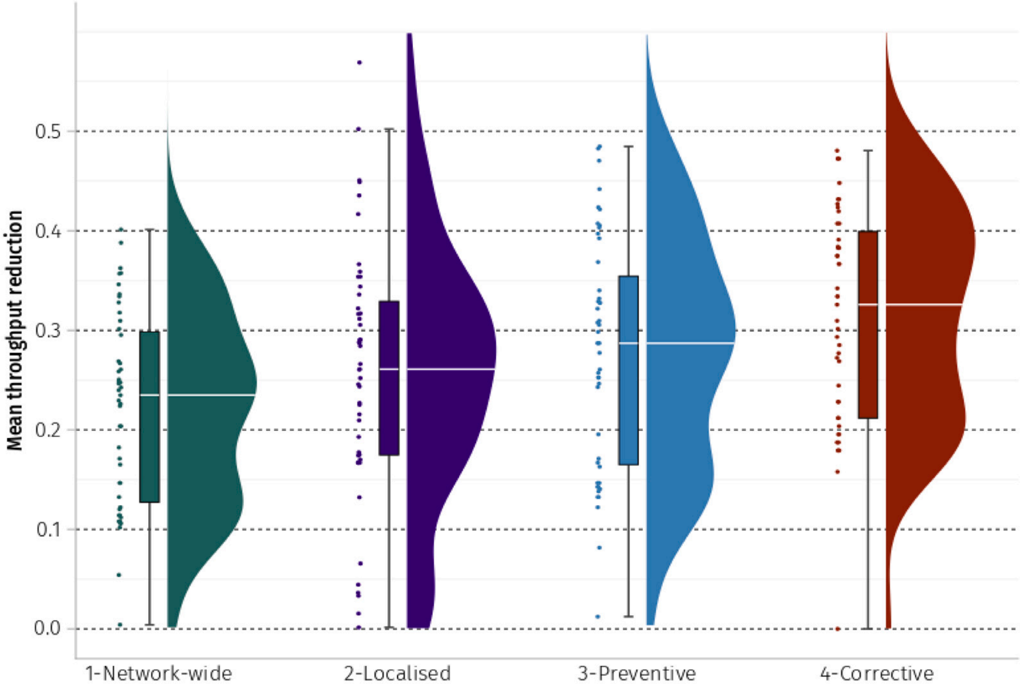

Fig. 6. Reduction of service throughput owing to maintenance operations in 10 Watts–Strogatz networks.

networks. This might be explained as both types of networks were used when training the agent that yields plans according to this approach. Note also that standard deviation is high in comparison to the ranges of the reductions obtained, which limits the power to generalise the behaviours observed. This needs to be investigated further and could be due to throughput reduction being more specific to the characteristics of the selection of services simulated and the network configuration used.

The cross-network analysis shows that average throughput reduction, as a measure of the impact of the maintenance approach, is higher in random networks created with the WS model than those created with the BA model. Extreme Low reduction or no reduction at all in some services is due to the availability of backup plans that can be used to re-route traffic during maintenance operations. This is evidenced by the overlapping markers close to 0.0 for several services and across all approaches in Fig. 5. The lack of backup paths seems to affect the performance of the Network-wide approach while it does not show substantial impact in the Localised approach. Overall, the Network-wide approach performs better, on average, than the alternatives, with larger differences in the BA networks. These results show that in BA

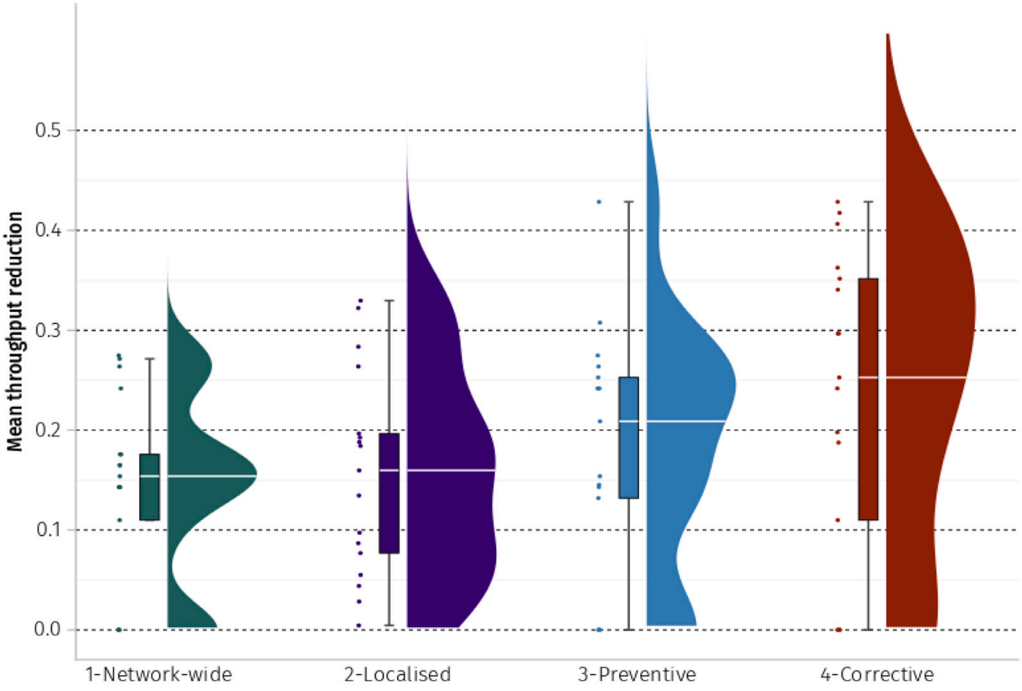

Fig. 7. Reduction of service throughput owing to maintenance operations in the backbone network.

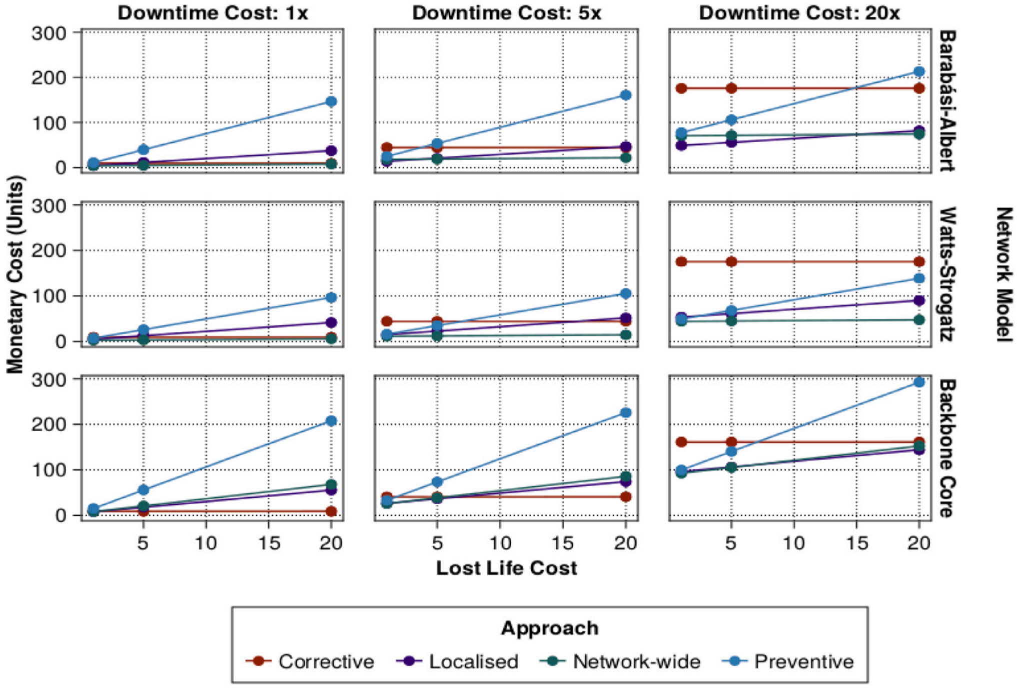

Fig. 8. Sensitivity to downtime and lost life cost parameters of the maintenance costs per cycle for the network models analysed.

networks the maintenance planner is better off using an individual preventive or corrective approach, in case the Network-wide is not possible. In WS, the differences among approaches is only 0.09 of reduction, however the Localised approach offers performance close to the Network-wide with lower overhead. Likewise, the distribution of the throughput reduction for the Localised approach in the WS networks shows a close-to-normal shape, which is useful to enable assumptions looking at wider simulation scenarios.

The comparative performance of the approaches in the Backbone (Metro-core) network is presented in Fig. 7 and the descriptive statistics also in Table 3. The trend is similar to that observed in the WS random networks. The Network-wide has the lowest impact on the services with only 0.14 of throughput reduction, the next best performance is by the Localised with 0.15. In this case the standard deviation is slightly higher for the individual approaches and lower in the Network-wide and Localised approaches.

The results of the maintenance costs per cycle, indicate that the sensitivity among approaches was minimal for the parts costs. Likewise the labour costs behaviour was similar to the downtime costs. For the sake of clarity, only downtime and lost life costs for the three types of networks analysed are presented in Fig. 8.

The positive slope of the preventive approach confirms that this approach is highly sensitive to lost life cost. It is the most expensive when lost life costs increase, as shown by the blue lines above others at right side of the subplots. When downtime costs are high (20x) and lost life costs are medium (5x) or low (1x), the corrective approach shows the highest maintenance costs, which is expected. Costs patterns for all the approaches are similar to WS and BA networks. Network-wide and Localised approach show medium-low sensitivity to both lost life and downtime costs as can be observed for the slope of the lines and the minimum shift to the top from left to right in the plot grid.

As a result of the cases studied, the Network-wide approach is the one that leads to average lowest impact on the quality of the services and the most cost-effective. Cost-wise, there are minimal difference between the Network-wide and the Localised approach for the cost parameters analysed. This is explained as both approaches are designed to optimise, or learn the policy that optimises, the defined cost function. The spread of the data obtained suggests that specific analysis is required for different portfolio of services and network configurations. This analysis discourages the use of an asset's individual preventive approach for networked assets as the costs and impact on quality are higher than the Network-specific approaches. Individual corrective causes the highest impact on the quality and the cost per cycle, is only as low as network-specific alternatives when the downtime costs are also low. However, corrective is the simplest approach to implement. Hence, when there is tolerance to quality reduction and downtime costs are low, the corrective approach seems an acceptable alternative.

The Network-specific approaches are more complex to implement. Particularly, as the Network-wide requires a comprehensive view of the network assets' state, the maintenance plan generation is more computationally demanding than individual approaches. In this case the planning process is highly sensitive to the scale the network, because the greater the number of assets to consider the higher the computational resources needed to both store condition trajectories of every asset and calculate alternative paths along a large network. The Localised approach although also computationally demanding for the training phase, is not as sensitive to the scale as the Network-wide as the scale of the subnetwork of assets, considered for planning, is capped to the size of the egonet with a fixed depth. This reduces and breaks down the demand of computational resources, compared to computing the plan for the entire network and then offers an acceptable alternative for networks when the topology shows small-world properties (WS model).

The Localised approach still shows room of improvement as only one Independent DQN algorithm was evaluated, while this is an active area of research. Likewise, alternative approaches for the approximation of the agent policy can be based on Graph Neural Networks (Cappart et al., 2021) which are naturally suited to represent local network state. Although this approach seems promising for the network maintenance problem, additional computational and environment design overhead must be also considered when using these approaches.

9. Conclusion and future work

This paper explores the use of network properties to plan the maintenance of multi-asset systems, aiming to reduce the impact of maintenance operations on the quality of services and the overall costs. Two network-specific maintenance planning approaches are introduced: A Network-wide and a Localised approach. The former considers the network topology and the dynamic traffic flows of the multi-asset system to jointly plan maintenance operations and re-route traffic flows accordingly while optimising impact reduction and costs. The latter approach, identifies local subnetworks and uses the Independent Deep Q-Learning Networks (I-DQN) algorithm, to learn a policy that generates the maintenance plan for each local subnetwork. The purpose of this latter approach is to reduce the overhead of considering the full network topology, the flows and every asset's condition when planning maintenance operations by providing an alternative, working in smaller subnetworks with a fixed size.

The performance of the proposed approaches is evaluated against individual corrective and preventive approaches over twenty random networks and an example of the UK's nationwide digital infrastructure backbone network. For evaluation of the Localised approach, an approach for integration of Multi-Agent Reinforcement Learning (MARL) and a Multi-asset agent-based model is also introduced. The Network-wide approach yields, on average, the lowest reduction on service throughput across all approaches and networks analysed. In networks with small-world properties, particularly the random networks generated from the Watts–Strogatz model and the backbone core, the Localised approach shows a performance close to the Network-wide with less overhead. Cost analysis across all networks and covering various combinations of parameters show minimal differences between the network-specific approaches, which are less sensitive to network and parameter changes, in contrast to individual approaches.

The current work sets the basis for the design of network-specific maintenance approaches using, agent-based modelling, mathematical optimisation and multi-agent reinforcement learning. Further work is required to evaluate approaches in a wider mix of network topologies and dynamics as the standard deviation of the results obtained in this study is high. Moreover, the Localised approach shows promising results and alternative MARL algorithms should be evaluated. More complex scenarios where assets have heterogeneous capacity and traffic can be distributed among more than one assets deserve further exploration as these resemble more closely existing nationwide infrastructure networks.

CRediT authorship contribution statement

Marco Pérez Hernández: Concept, Methods, Writing. **Alena Puchkova:** Concept, Methods, Writing. **Ajith K. Parlikad:** Concept, Methods, Review.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Ajith Parlikad reports financial support was provided by BT Group Plc. Ajith Parlikad reports financial support was provided by Engineering and Physical Sciences Research Council.

Data availability

Some data is confidential (e.g. metro-core network topology) other data available upon request.

References

Amin, M., et al., 2000. National infrastructures as complex interactive networks. In: Automation, Control, and Complexity: An Integrated Approach, Vol. 3. Citeseer, pp. 263–286.

Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks. science 286 (5439), 509–512.

Bukhsh, Z.A., Molegraaf, H., Jansen, N., 2023. A maintenance planning framework using online and offline deep reinforcement learning. Neural Comput. Appl. 1–12.

Busoniu, L., Babuska, R., De Schutter, B., 2010. Multi-agent reinforcement learning: An overview. In: Srinivasan, D., Jain, L.C. (Eds.), Innovations in Multi-agent Systems and Applications-1. Springer Berlin Heidelberg, pp. 183–221, chapter 7.

Cappart, Q., Chételat, D., Khalil, E., Lodi, A., Morris, C., Veličković, P., 2021. Combinatorial optimization and reasoning with graph neural networks. arXiv preprint arXiv:2102.09544.

Chen, J., Wang, Y., 2023. A deep reinforcement learning approach for maintenance planning of multi-component systems with complex structure. Neural Comput. Appl. 1–14.

Chu, J., Zhou, Z., Ding, X., Tian, Z., 2022. A life cycle oriented multi-objective optimal maintenance of water distribution: Model and application. Water Resour. Manag. 36 (11), 4161–4182.

Ercsey-Ravasz, M., Toroczkai, Z., 2010. Centrality scaling in large networks. Phys. Rev. Lett. 105 (3), 038701.

Foerster, J., Assael, I.A., De Freitas, N., Whiteson, S., 2016. Learning to communicate with deep multi-agent reinforcement learning. Adv. Neural Inf. Process. Syst. 29.

Frangopol, D.M., Liu, M., 2007. Bridge network maintenance optimization using stochastic dynamic programming. J. Struct. Eng. 133 (12), 1772–1782.

Han, C., Ma, T., Xu, G., Chen, S., Huang, R., 2022. Intelligent decision model of road maintenance based on improved weight random forest algorithm. Int. J. Pavement Eng. 23 (4), 985–997.

ISO (International Organization for Standardization), 2014. ISO 55001 asset management—Management systems: Requirements.

Kosinski, J., Nawrocki, P., Radziszowski, D., Zielinski, K., Zielinski, S., Przybylski, G., Wnek, P., 2008. SLA monitoring and management framework for telecommunication services. In: Fourth International Conference on Networking and Services (icns 2008). IEEE, pp. 170–175.

Kuhnle, A., Jakubik, J., Lanza, G., 2019. Reinforcement learning for opportunistic maintenance optimization. Prod. Eng. 13, 33–41.

Lee, K., Rucker, M., Scherer, W.T., Beling, P.A., Gerber, M.S., Kang, H., 2017. Agent-based model construction using inverse reinforcement learning. In: 2017 Winter Simulation Conference (WSC). IEEE, pp. 1264–1275.

Lei, X., Xia, Y., Deng, L., Sun, L., 2022. A deep reinforcement learning framework for life-cycle maintenance planning of regional deteriorating bridges using inspection data. Struct. Multidiscip. Optim. 65 (5), 149.

Leoni, L., BahooToroody, A., De Carlo, F., Paltrinieri, N., 2019. Developing a risk-based maintenance model for a Natural Gas Regulating and Metering Station using Bayesian Network. J. Loss Prev. Process Ind. 57, 17–24.

Li, F., Ma, L., Sun, Y., Mathew, J., 2014. Group maintenance scheduling: A case study for a pipeline network. Lect. Notes Mech. Eng. 9.

Liang, Z., Parlikad, A.K., 2020. Predictive group maintenance for multi-system multi-component networks. Reliab. Eng. Syst. Saf. 195 (October 2019), 106704.

Lidén, T., Joborn, M., 2017. An optimization model for integrated planning of railway traffic and network maintenance. Transp. Res. C 74, 327–347.

Márquez, A.C., Alberca, J.A.M., del Castillo, A.C., 2023. Simulating dynamic RUL based CBM scheduling. A case study in the railway sector. Comput. Ind. 148, 103914.

Matin, S.A.A., Mansouri, S.A., Bayat, M., Jordehi, A.R., Radmehr, P., 2022. A multi-objective bi-level optimization framework for dynamic maintenance planning of active distribution networks in the presence of energy storage systems. J. Energy Storage 52, 104762.

Meneses, S., Ferreira, A., 2012. New optimization model for road network maintenance management. Procedia - Soc. Behav. Sci. 54, 956–965.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al., 2015. Human-level control through deep reinforcement learning. nature 518 (7540), 529–533.

Mohammadi, R., He, Q., 2022. A deep reinforcement learning approach for rail renewal and maintenance planning. Reliab. Eng. Syst. Saf. 225, 108615.

Moradi, S., Vahidinasab, V., Kia, M., Dehghanian, P., 2019. A mathematical framework for reliability-centered maintenance in microgrids. Int. Trans. Electr. Energy Syst. 29 (1), e2691.

Newman, M., 2018. Networks. OUP Oxford.

Pastor-Satorras, R., Vespignani, A., 2004. Evolution and Structure of the Internet: A Statistical Physics Approach. Cambridge University Press.

Pérez Hernández, M., Puchkova, A., Parlikad, A., 2022. Maintenance strategies for networked assets. IFAC-PapersOnLine (19), 151–156.

Petchrompo, S., Parlikad, A.K., 2019. A review of asset management literature on multi-asset systems. Reliab. Eng. Syst. Saf. 181, 181–201.

Pinciroli, L., Baraldi, P., Zio, E., 2023. Maintenance optimization in Industry 4.0. Reliab. Eng. Syst. Saf. 109204.

Poór, P., Ženíšek, D., Basl, J., 2019. Historical overview of maintenance management strategies: Development from breakdown maintenance to predictive maintenance in accordance with four industrial revolutions. In: Proceedings of the International Conference on Industrial Engineering and Operations Management, Pilsen, Czech Republic. pp. 23–26.

Rocchetta, R., Bellani, L., Compare, M., Zio, E., Patelli, E., 2019. A reinforcement learning framework for optimal operation and maintenance of power grids. Appl. Energy 241, 291–301.

Rodriguez, M., Kubler, S., de Giorgio, A., Cordy, M., Robert, J., LeTraon, Y., 2022. Multi-agent deep reinforcement learning based predictive maintenance on parallel machines. Robot. Comput. Integr. Manuf. 78.

Scott, J., Carrington, P., 2011. The SAGE Handbook of SNA. SAGE Publications.

Stavdas, A., 2010. Core and Metro Networks. John Wiley & Sons.

Sutton, R.S., Barto, A.G., 2018. Reinforcement Learning: An Introduction. MIT Press.

Tampuu, A., Matiisen, T., Kodelja, D., Kuzovkin, I., Korjus, K., Aru, J., Aru, J., Vicente, R., 2017. Multiagent cooperation and competition with deep reinforcement learning. PLoS One 12 (4).

Tanenbaum, A.S., 2003. Computer Networks. Pearson Education India.

Thomas, J., Hernandez, M.P., Parlikad, A., Piechocki, R., 2021. Network maintenance planning via multi-agent reinforcement learning. In: Proc. for the 2021 IEEE Conference on Systems, Man, and Cybernetics (SMC). pp. 2289–2295.

Tian, J., other-contributors, 2020. ReinforcementLearning.jl: A reinforcement learning package for julia.

Vargas-Pérez, V.A., Mesejo, P., Chica, M., Cordón, O., 2023. Deep reinforcement learning in agent-based simulations for optimal media planning. Inf. Fusion 91, 644–664.

Vespignani, A., 2018. Twenty years of network science. Nat. News Views (558), 528.

Wang, Y., Limmer, S., Van Nguyen, D., Olhofer, M., Bäck, T., Emmerich, M., 2022. Optimizing the maintenance schedule for a vehicle fleet: a simulation-based case study. Eng. Optim. 54 (7), 1258.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world'networks. nature 393 (6684), 440–442.

Yao, L., Dong, Q., Jiang, J., Ni, F., 2020. Deep reinforcement learning for long-term pavement maintenance planning. Comput. Aided Civ. Infrastruct. Eng. (11).

Zhang, C., Gao, Y., Yang, L., Gao, Z., Qi, J., 2020. Joint optimization of train scheduling and maintenance planning in a railway network: A heuristic algorithm using Lagrangian relaxation. Transp. Res. B 134, 64.

Zhang, N., Si, W., 2020. Deep reinforcement learning for condition-based maintenance planning of multi-component systems under dependent competing risks. Reliab. Eng. Syst. Saf. 203 (June).

Zio, E., 2007. From complexity science to reliability efficiency: a new way of looking at complex network systems and critical infrastructures. Int. J. Crit. Infrastruct. 3 (3–4), 488–508.