COGNAC: Cooperative Graph-based Networked Agent Challenges for Multi-Agent Reinforcement Learning

Jules Sintes, Ana Bušić

Inria, DIENS, École Normale Supérieure, PSL University, Paris, France

A Reinforcement Learning Environments Suite

Many large-scale real-world systems exhibit an inherent **graph structure**.

COGNAC is a collection of fully cooperative multi-agent environments with network structures, enabling the development and evaluation of scalable, decentralized, and frugal RL algorithms.

Our contributions:

- A **flexible Python package** for graph-structured MARL tasks with a Gym-like interface.
- The **first standardized open-source implementations** of classical graph-based MARL problems.
- A set of **baseline experiments with MARL algorithms**, showcasing scalability challenges and opportunities for future method development.

🎯 *Identify special structures that make problems theoretically solvable.*

Multi-agent Reinforcement Learning (MARL)

The standard framework to describe MARL task is Decentralized Partially Observable Markov Decision Processes (Dec-POMDP):

$$\langle \underbrace{\mathcal{N}}_{\text{Agents}},\ \underbrace{\mathcal{S}}_{\text{State}},\ \underbrace{\mathcal{A}}_{\text{Action}},\ \underbrace{\mathcal{P}}_{\text{Transition Kernel}},\ \underbrace{\mathcal{Z}}_{\text{Obs.}},\ \underbrace{\mathcal{O}}_{\text{Obs. Kernel}},\ \underbrace{\mathcal{R}}_{\text{Reward}},\ \underbrace{\gamma}_{\text{Discount Factor}} \rangle$$

Theorem [Oliehoek, 2016]**:** In the general case, finding an optimal solution for a finite-horizon DecPOMDP is known to be **NEXP-complete** for $N \geq 2$ (i.e., two agents or more).

Parallel environment All agents act and observe simultaneously.
Fully cooperative tasks Agents maximize a common global objective.
Partially observable Agents only observe the environment locally

Available environments

Environment	Modular size	Graph Agnostic	Joint State Space	Joint Act. Space
Firefighting Graph (1D)	✅	❌	θ^N	2^N
Firefighting Graph (2D)	✅	❌	$\theta^{N \times M}$	4^N
Binary Consensus	✅	✅	2^N	2^N
SysAdmin Network	✅	✅	9^N	2^N
Multi-commodity Flow	✅	❌	$\rho_{max}^{k \times E}$	$\rho_{max}^{k \times E}$

Table 1: Summary of available environments with main features, and state/action space sizes

✅ **Graph-native API:** define any graph topology for cooperative problems.
🧩 **Simple, extensible environments:** start small, scale big.
🔌 **Baseline integrations:** compatible with standard MARL algorithms.

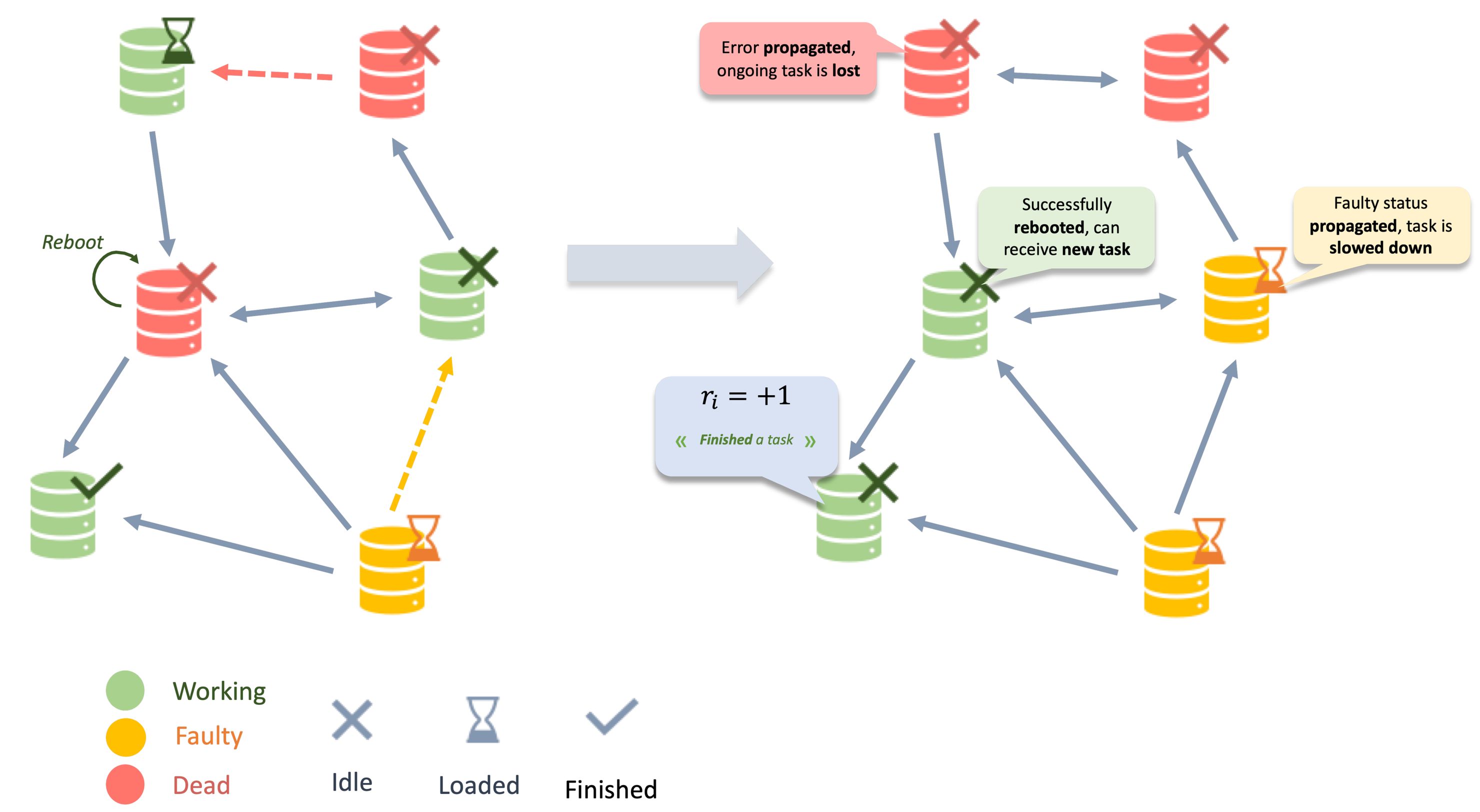

Figure 1: Illustration of the SysAdmin Network problem [Guestrin, 2002]. Example of an instance of the environment. Each node represents a computer managed by an agent; failures propagate through network links. Agents must coordinate to keep the system operational.

Benchmark example

We test the most-popular MARL algorithms on two environments:

- **SysAdmin:** agents manage servers, balancing task processing and reboots while failures propagate through the graph.
- **Binary Consensus:** agents update their target vote through state transitions that are stochastically influenced by their neighbors' actions.

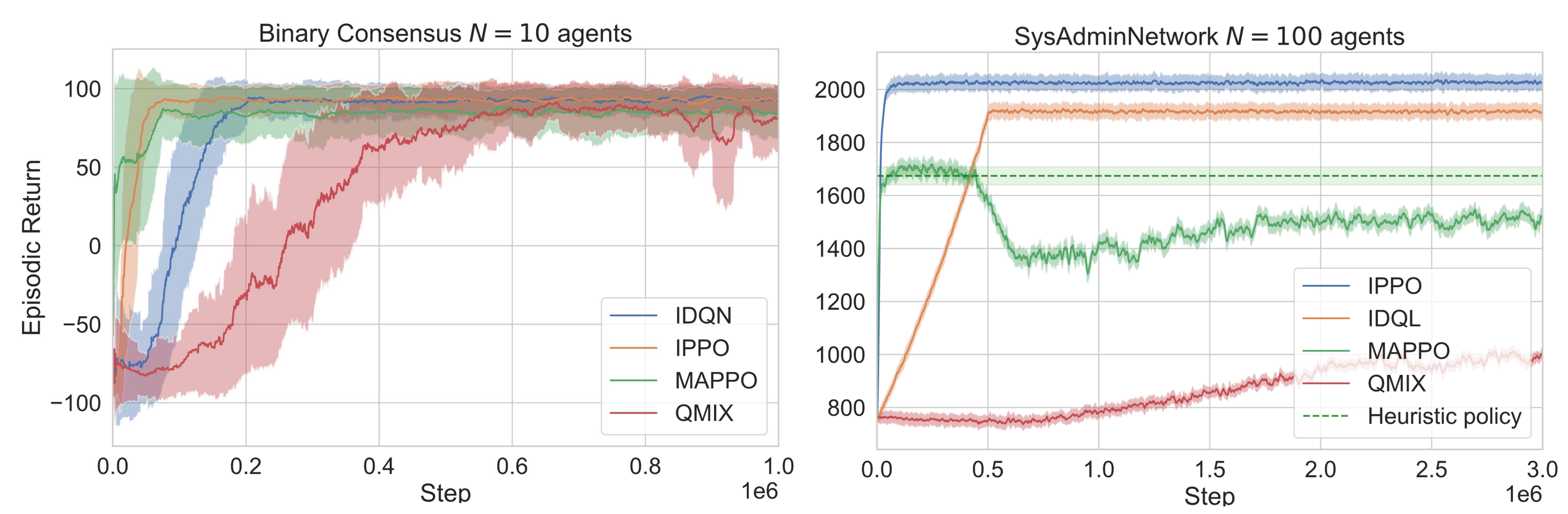

Figure 2: Training comparison on COGNAC's environment with various size and heuristic baselines.

- IPPO consistently learns robust policies across all environments.
- Standard Centralized Training Decentralized Execution (CTDE) methods fail to generalize effectively to larger systems.

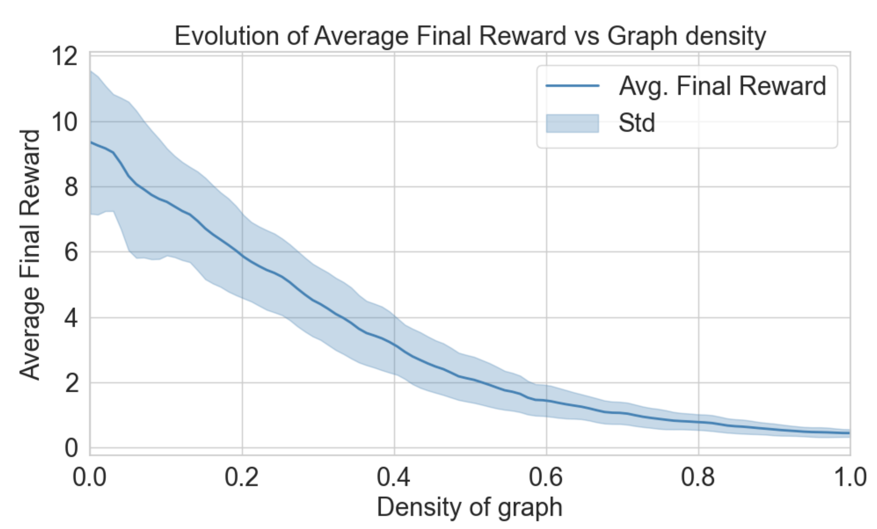

Figure 3: Average final reward achieved by standard IDQL training as a function of the density of Erdős–Rényi graphs with N=10 agents (Binary Consensus).

How can we use the underlying graph-structure?

- InforMARL algorithm [Nayak, 2023] with information aggregator based on Graph-Neural Networks.
- Policy sharing mechanism on standard MAPPO for homogeneous agents.

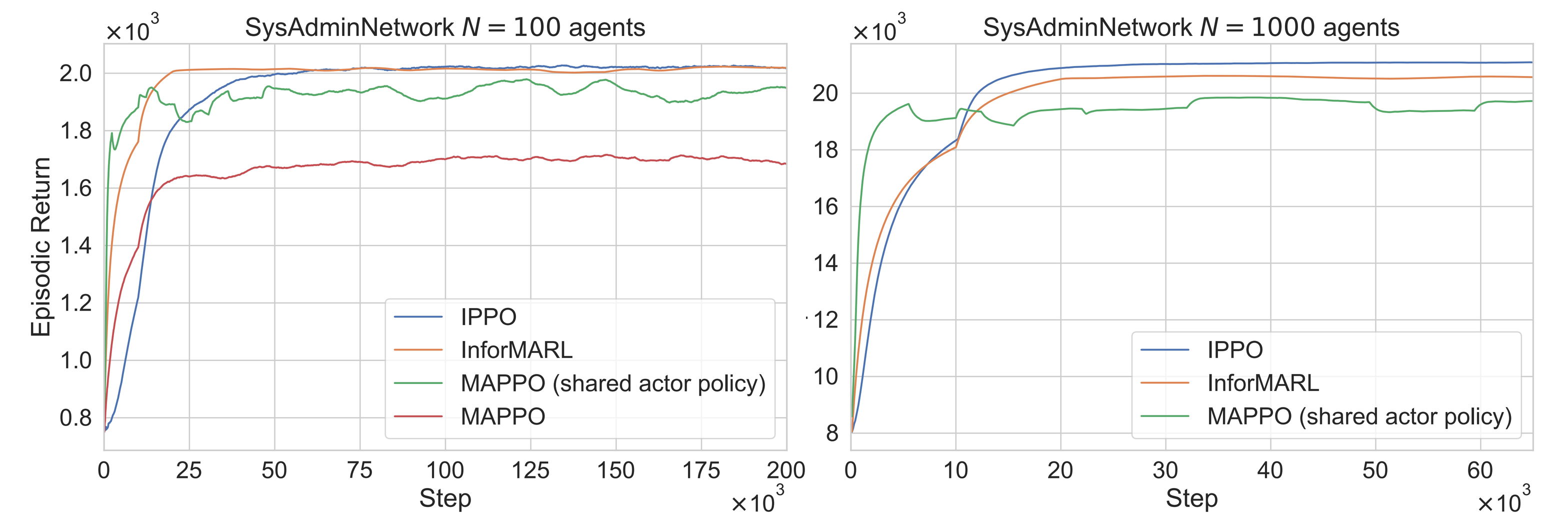

Figure 4: Comparison of PPO-based methods on the SysAdmin Network problem.

- **Actor policy-sharing on MAPPO** stabilizes training while significantly increasing computation speed.
- **InforMARL** effectively uses the graph structure at scale with increased learning speed.
- Requires **full-state knowledge,** network structure, and additional **information sharing** between agents as opposed to **fully decentralized IPPO.**

Conclusion

We believe in COGNAC's relevance as a bridge between theoretical MARL research and practical applications in distributed network systems. We hope it will foster future work in structure-exploiting scalable MARL algorithms.

References

Oliehoek, F. A., & Amato, C. (2016). *A concise introduction to decentralized POMDPs* (Vol. 1). Cham, Switzerland: Springer International Publishing.
Guestrin, C., Lagoudakis, M., & Parr, R. (2002). Coordinated reinforcement learning. In *Proceedings of the 19th International Conference on Machine Learning (ICML)*.
Nayak, S., et al. (2023). Scalable multi-agent reinforcement learning through intelligent information aggregation. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*. PMLR.

Links

Main package repository: https://github.com/yojul/cognac
Benchmark experiments: https://github.com/yojul/cognac-benchmark-example