We scale multi-agent coordination on graphs with a notion of risk and support using Reinforcement Learning.

**Motivation**
Multiple agents traversing an Environment Graph (EG) with risk edges can be given support from specific nodes.
- How to reduce overall team cost by taking coordinated actions?
- How to scale up coordination for increasing number of agents and complex graphs?

**Problem Formulation**
Formulate it as MDP for Single Environment Graph (Single EG):
- **State Space (S):** Joint state using N agents positions in one hot encoding.
- **Action Space (A):** Joint action of N agents indicating the nodes the agents traverse to.

Formulate it as MDP for Multiple Environment Graphs (Multiple EGs):
- **State Space (S):** Joint state as combination of agents’ positions as one-hot encoding, graph connectivity, and supporting mechanism.
- **Action Space (A):** Joint action of N agents indicating the nodes the agents traverse to.

Experimentally, we find that RL can solve complex graph problems with more agents with near optimality guarantee.

**Methods & Implementations**
Masked Q-Learning:
- State, Q-table, Q-values, Masked Q-values
- Invalid Action Masking

Masked PPO:
- State, Neural Network, Action Logits, Masked Action Logits
- Invalid Action Masking

Invalid Action Masking:
- Action Level Masking in Q-Learning:
  \[ \alpha = \arg \max_a (Q(s, a') \times M(s, a')) \]
  with prob. \(1 - \epsilon\) random action \(a'\), s.t. \(M(s, a') = 1\), with prob. \(\epsilon\).

- Logit Level Masking in PPO:
  \[ \pi(a | s) = \frac{\exp(\log \pi(a' | s))}{\sum_{a' \in A} \exp(\log \pi(a' | s))} \]

**Reward Shaping**
- **Goal Reward:** \[ r_g = \begin{cases} +10, & \text{if all agents arrive at goal(s)} \\ -0.1, & \text{otherwise} \end{cases} \]
- **Movement Reward:** \[ r_m = -\sum_{\text{all agents}} -C \left( \| \mathbf{x}_i - \mathbf{x}_j \| ight) \]
- **Coordination Reward:** \[ r_c = \alpha \times CC - \beta \times RC \]
- **Final Reward:** \[ r_f = w_1 r_g + w_2 r_m + w_3 r_c \]