InforMARL: Scalable Multi-Agent Reinforcement Learning through Intelligent Information Aggregation

Siddharth Nayak[1], Kenneth Choi[1], Wenqi Ding[1], Sydney Dolan[1],
Karthik Gopalakrishnan[2], Hamsa Balakrishnan[1]

[1]Massachusetts Institute of Technology

[2]Stanford University

{sidnayak, kenchoi, wenqi2, sydneyd, hamsa}@mit.edu

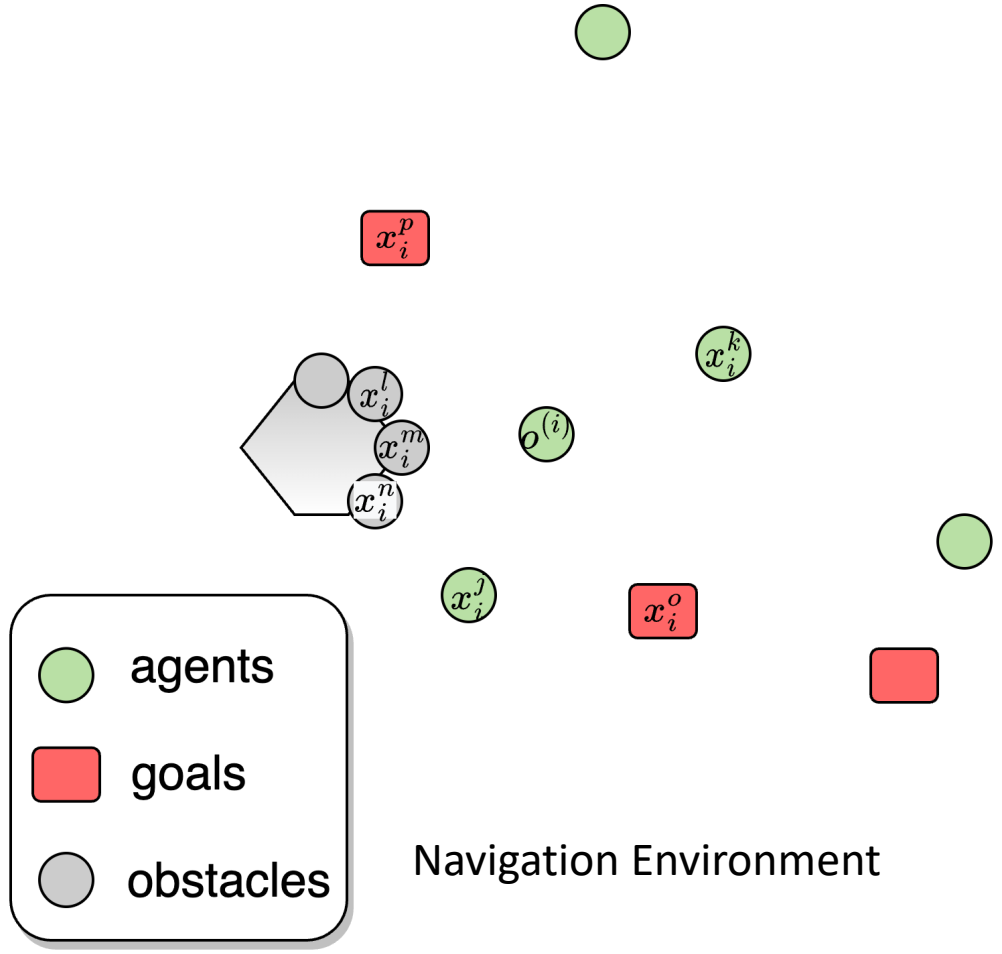
kgopalakrishnan@stanford.edu

Background and Motivation

Credit: U.S. Naval Institute

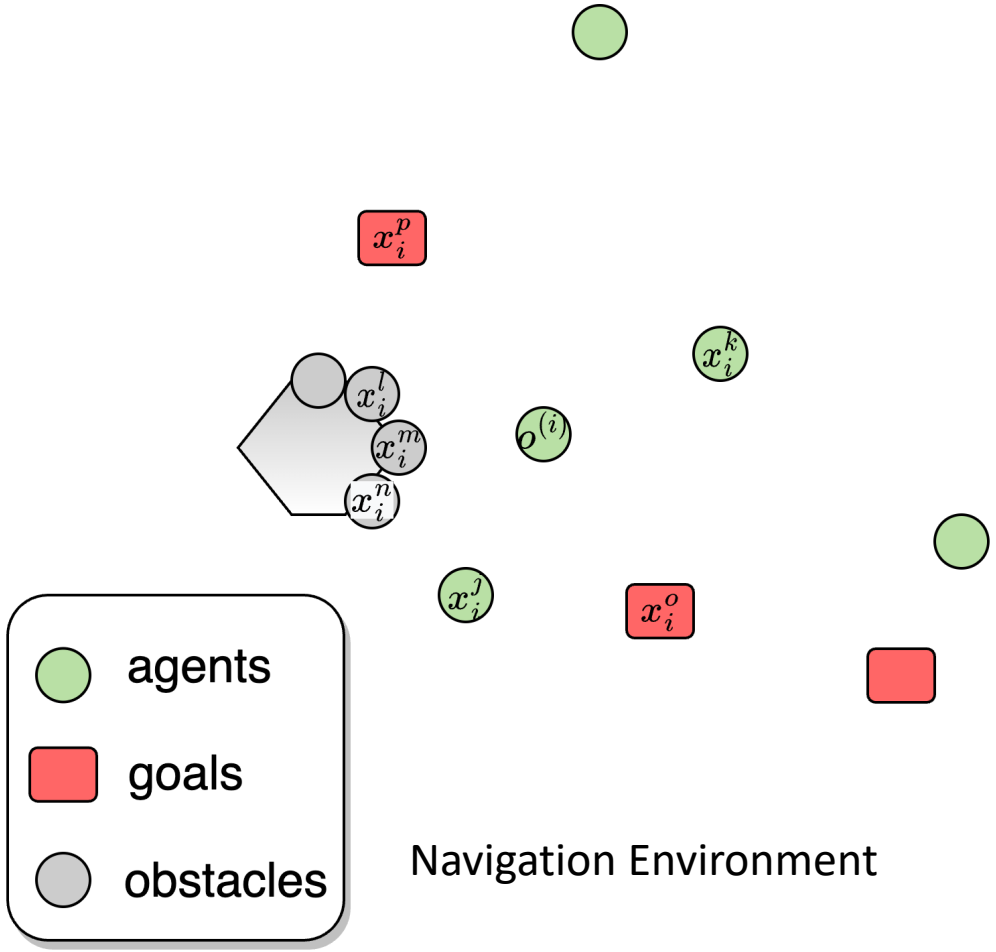
Credit: The Robot Report

Background and Motivation



Navigation Environment

Background and Motivation

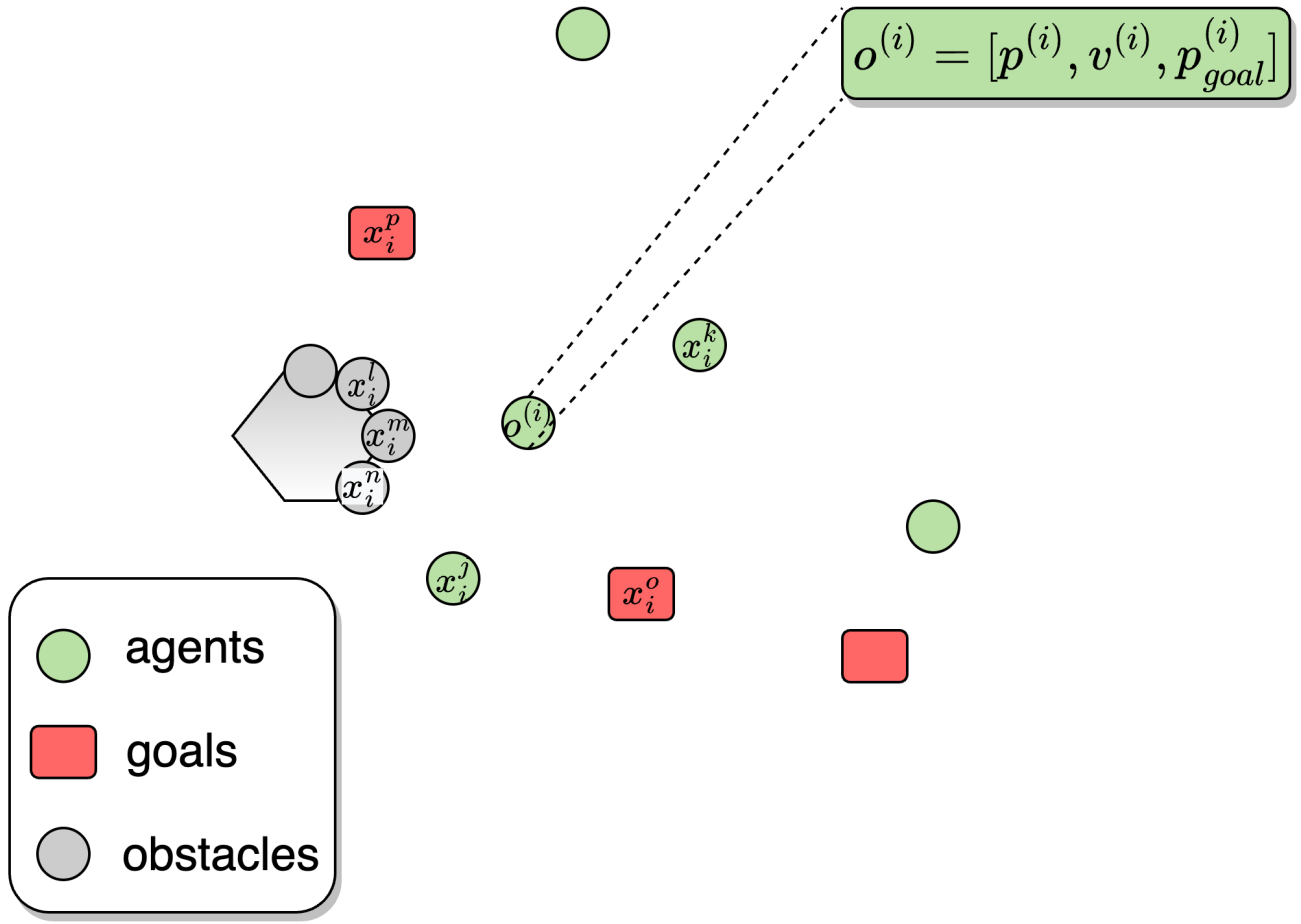


Navigation Environment

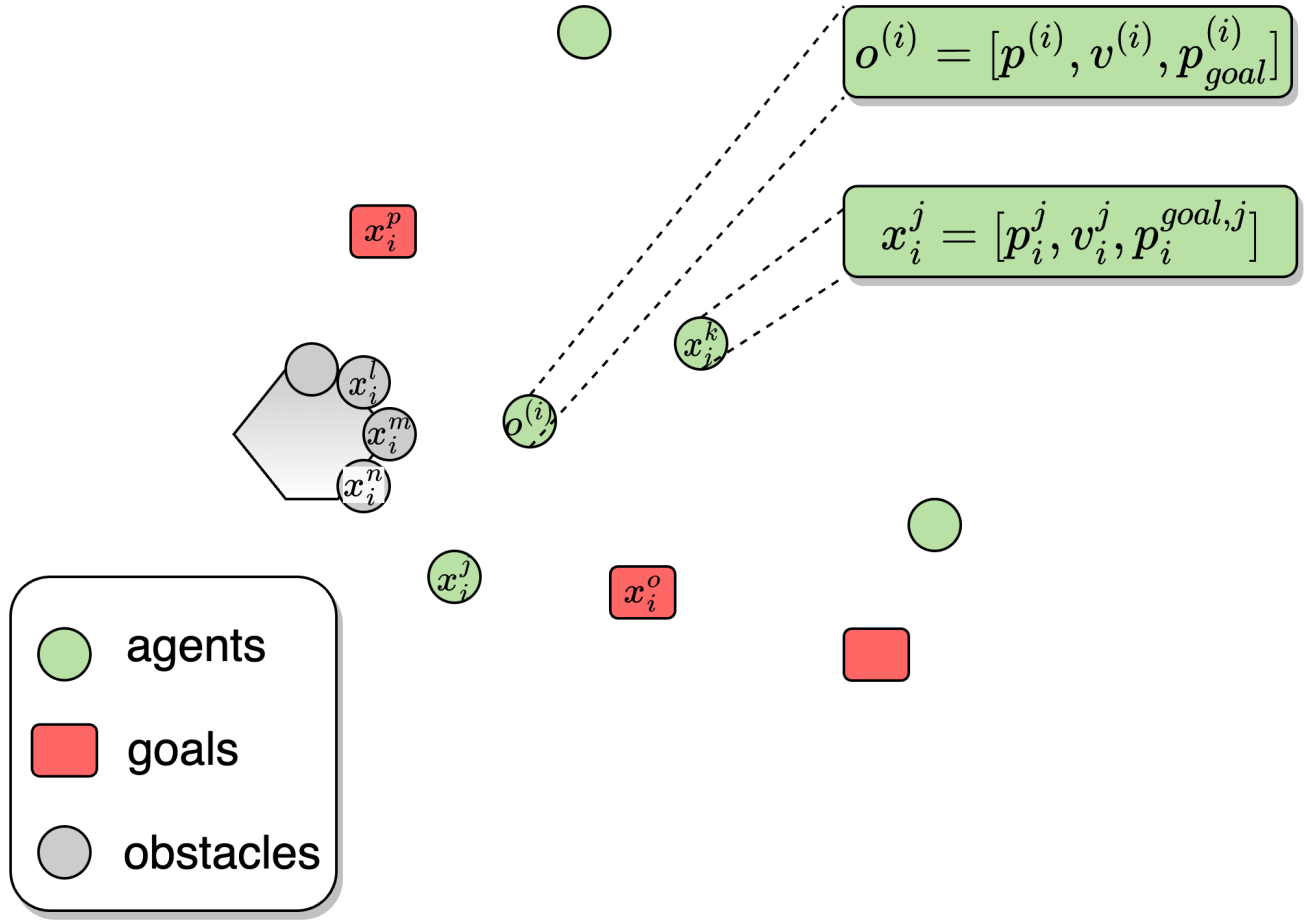
Key features expected from MARL algorithms:

- Decentralized execution
- Scalability
- Efficiency in sample complexity to train

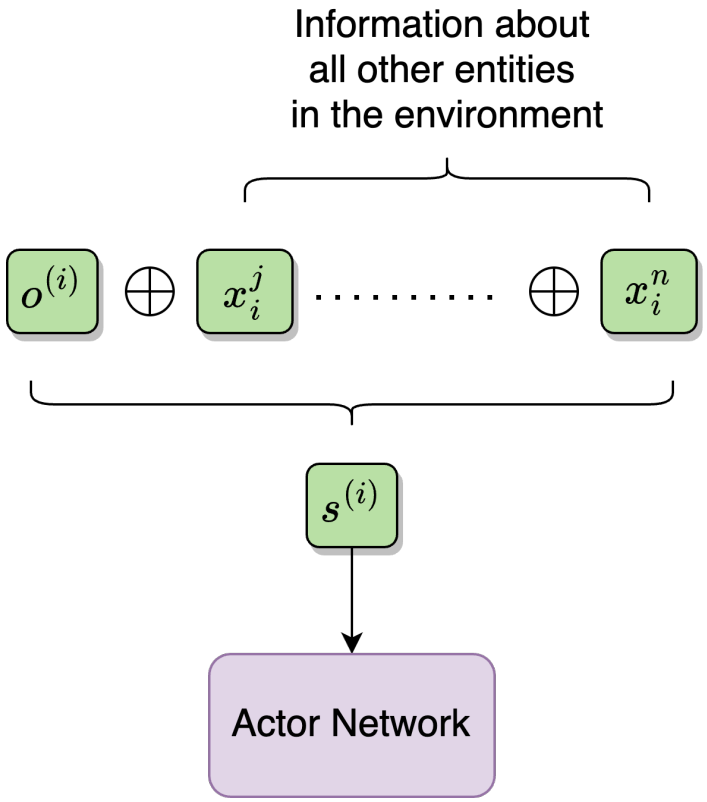
Background and Motivation



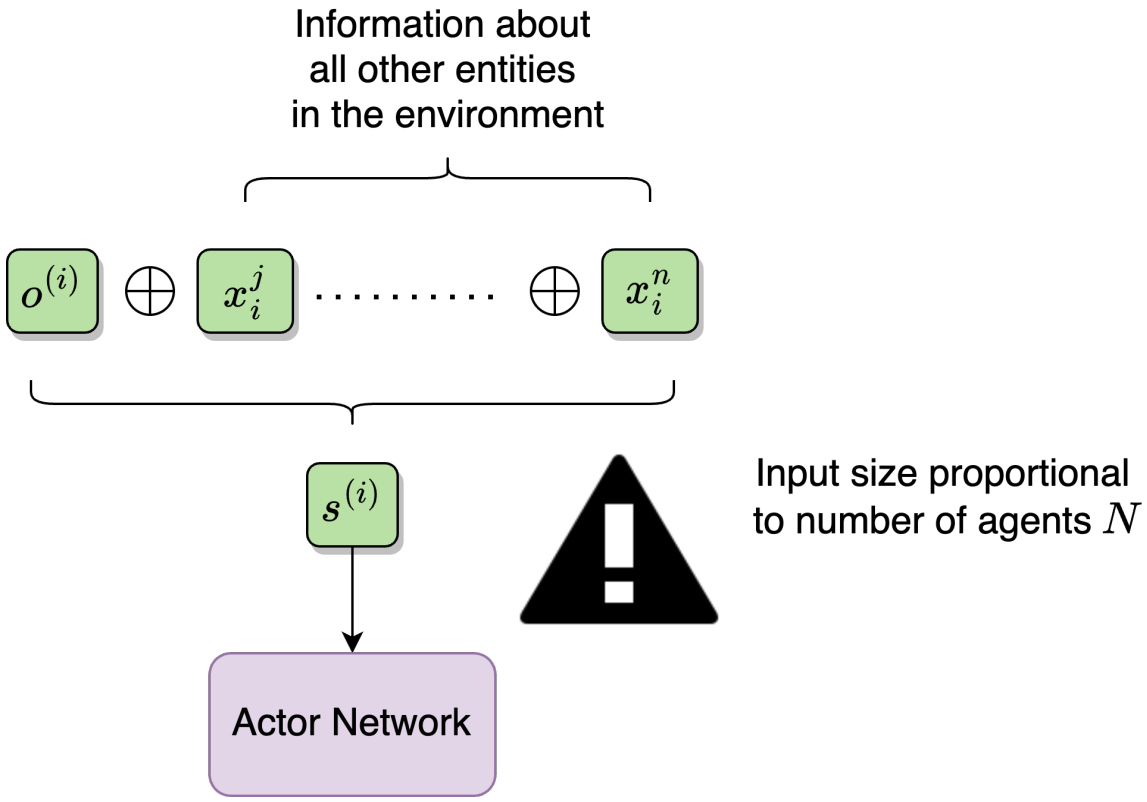
Background and Motivation



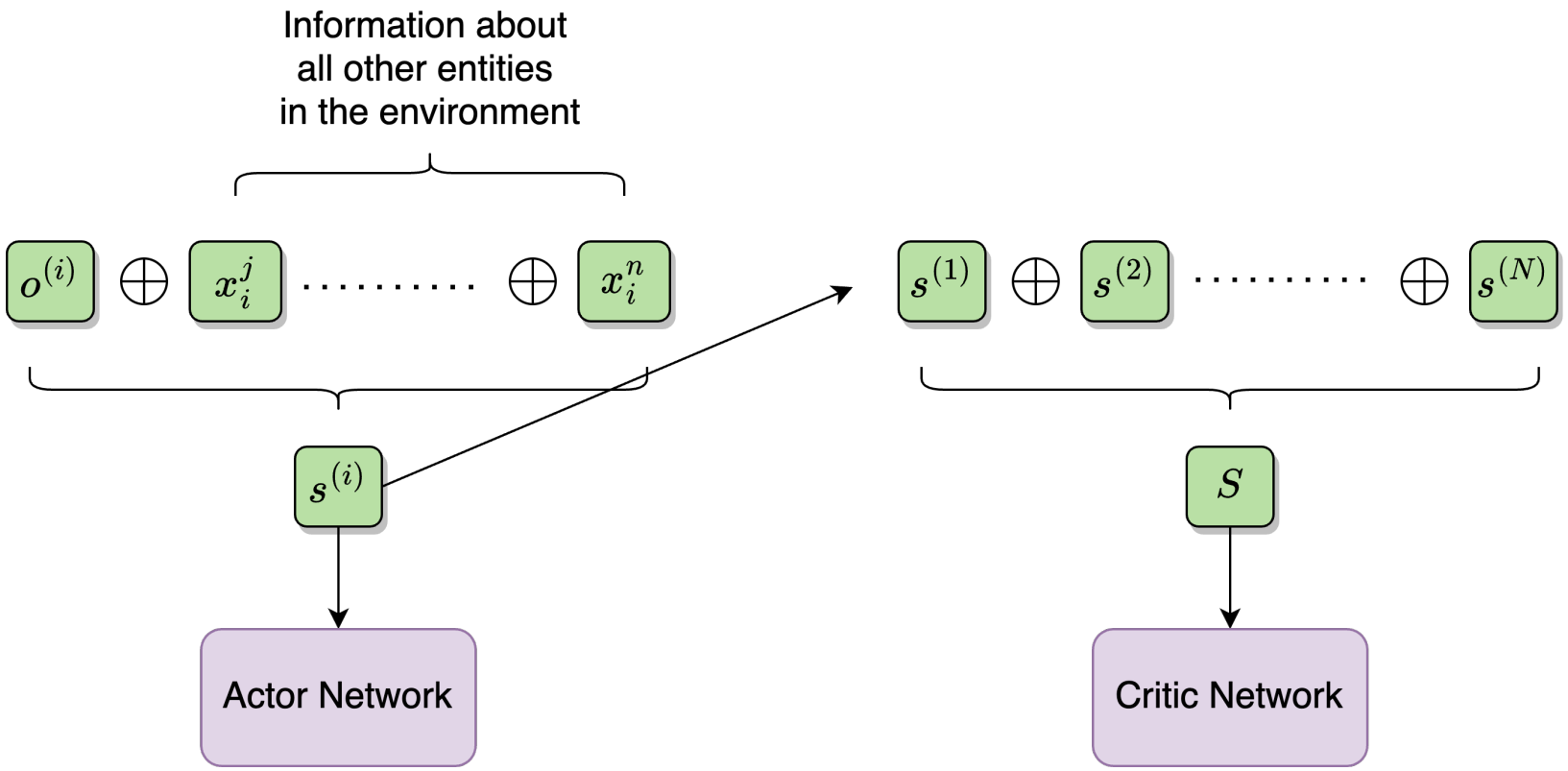
Background and Motivation



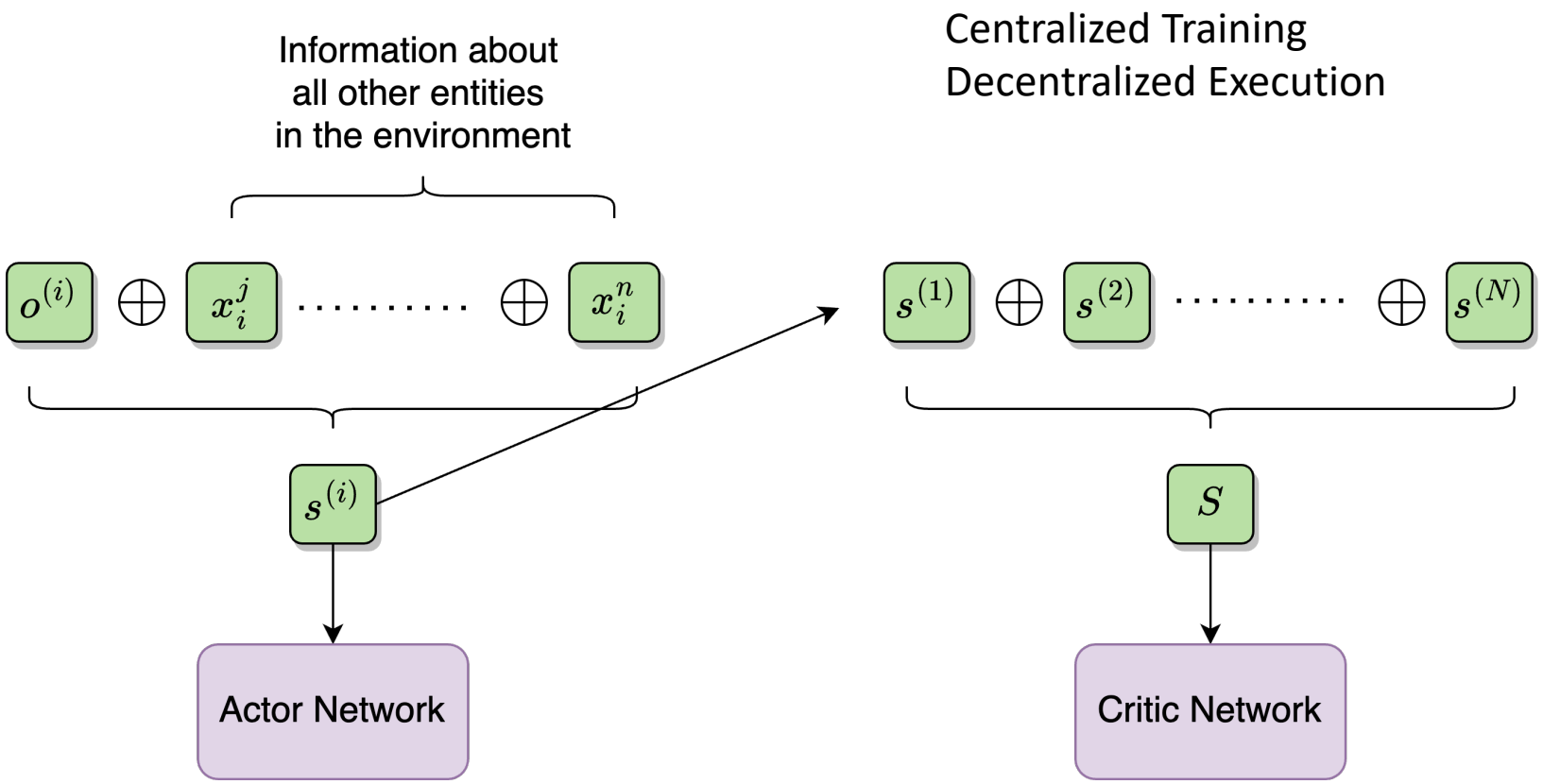
Background and Motivation



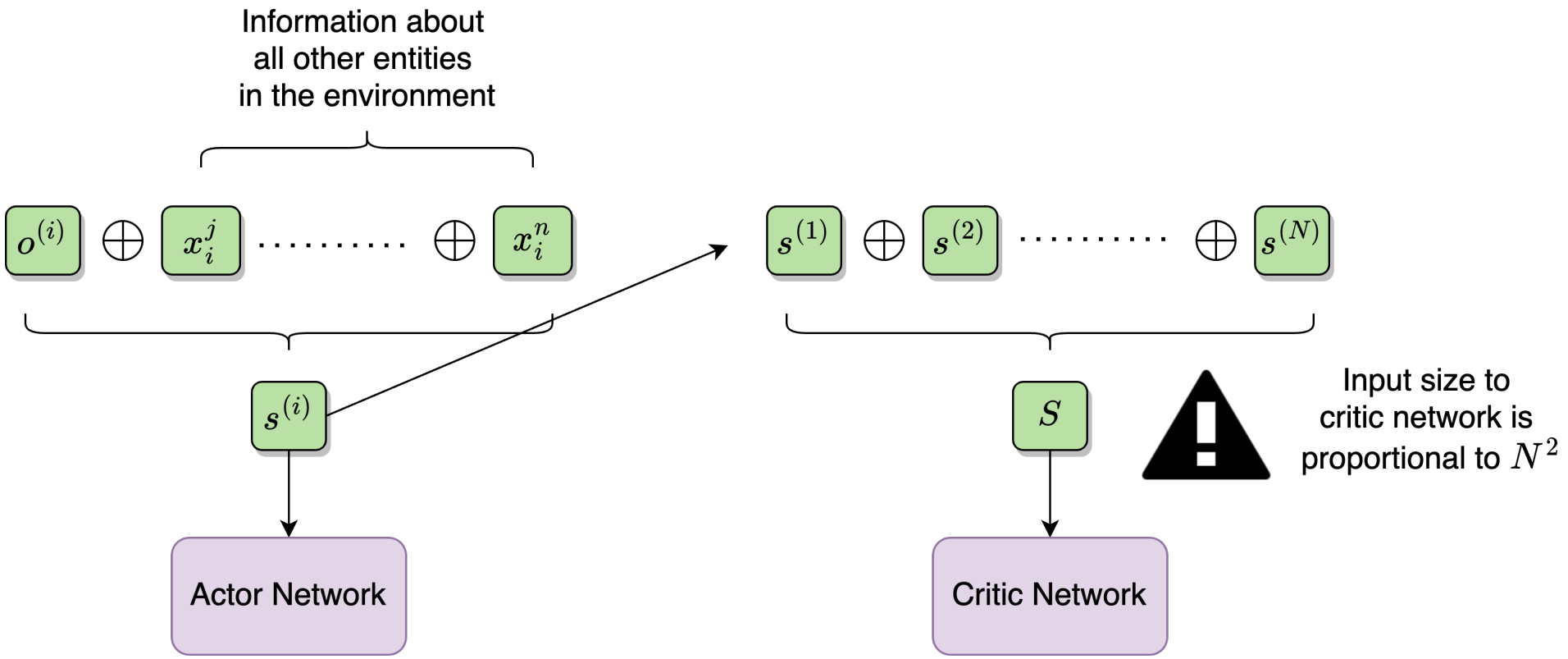
Background and Motivation



Background and Motivation



Background and Motivation

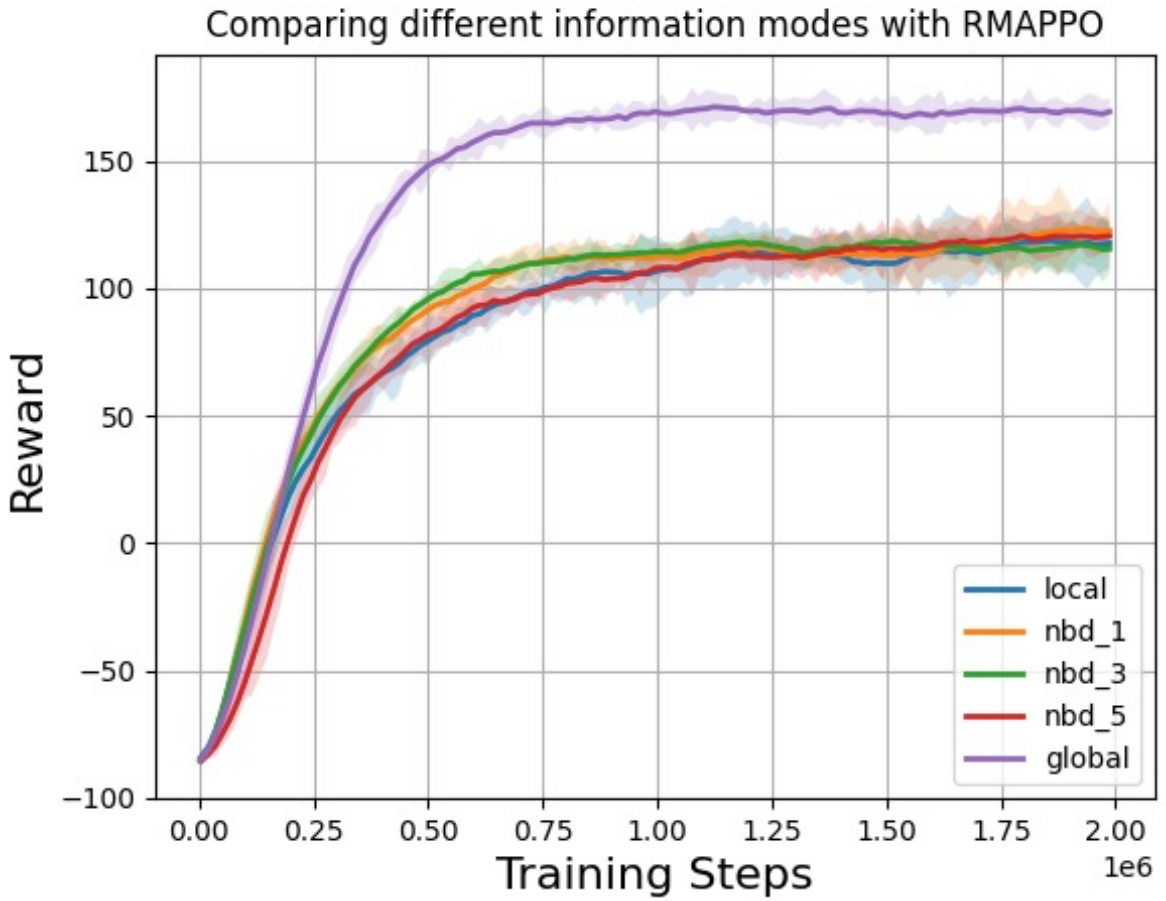


Motivating Experiment

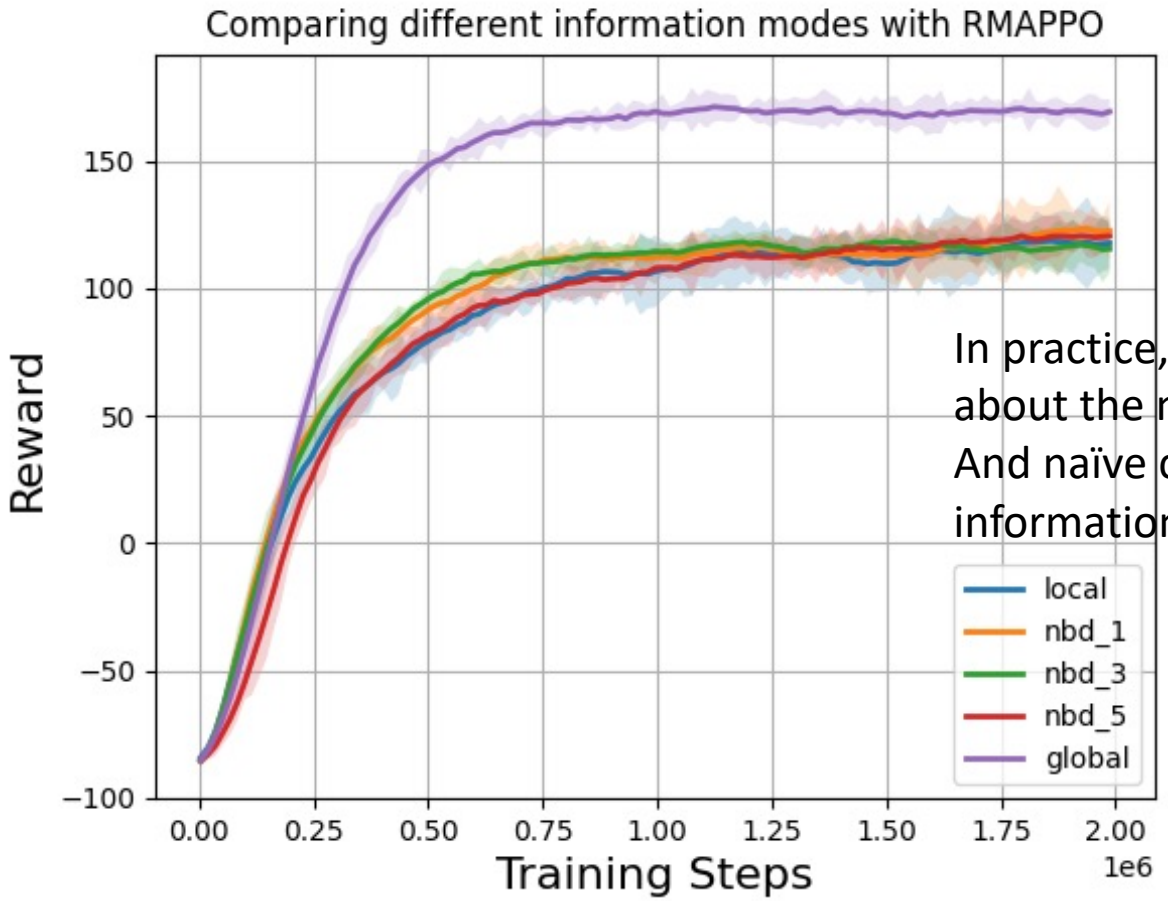
Vary the amount of information included in observations for actor

- Local Information Mode: $o^{(i)}$
- Global Information Mode: $o^{(i)} \oplus x_i^j \cdots\cdots \oplus x_i^n$
- Neighborhood Information Mode: $o^{(i)} \oplus x_i^j \cdots\cdots \oplus x_i^n$

Motivating Experiment

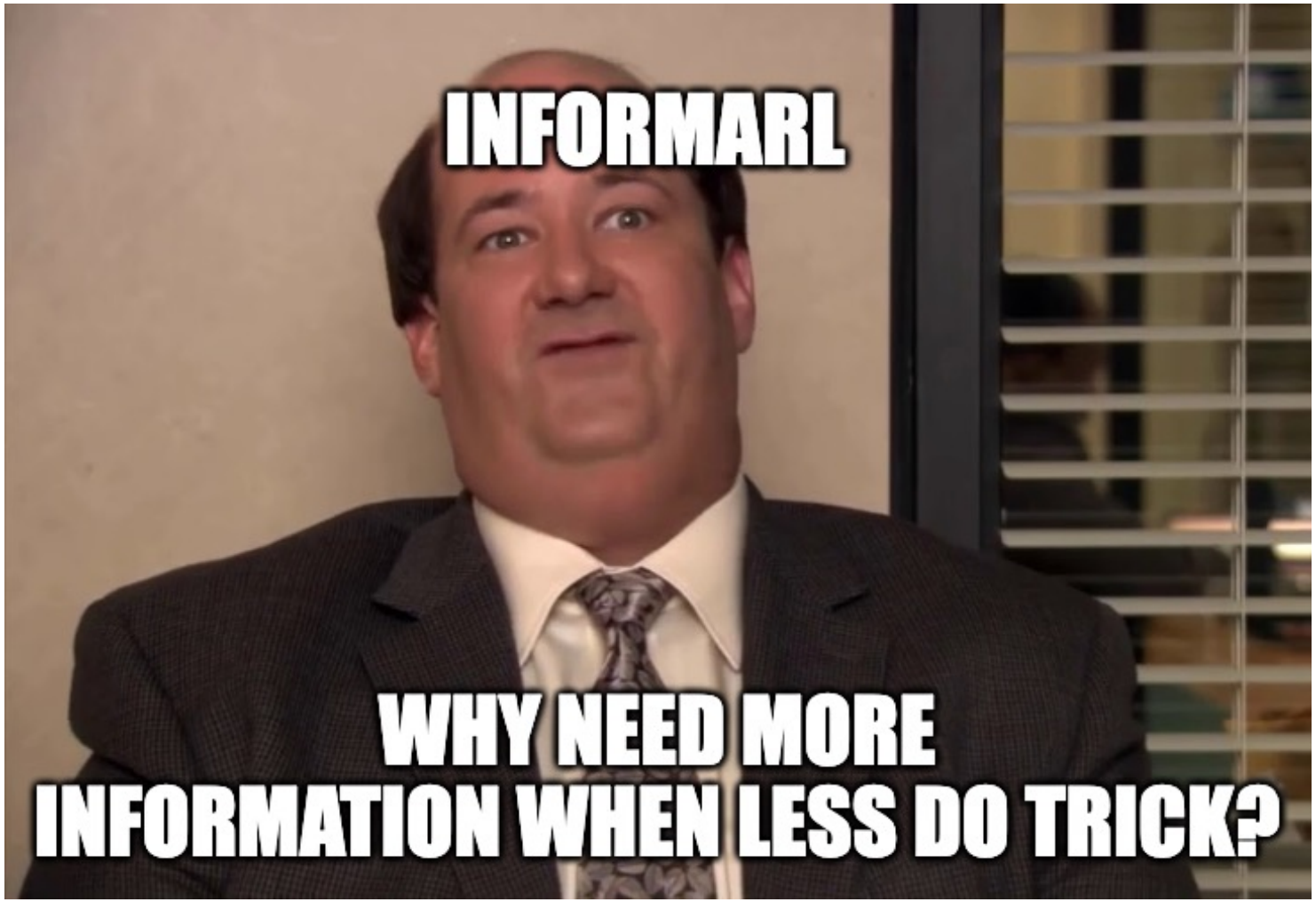


Motivating Experiment

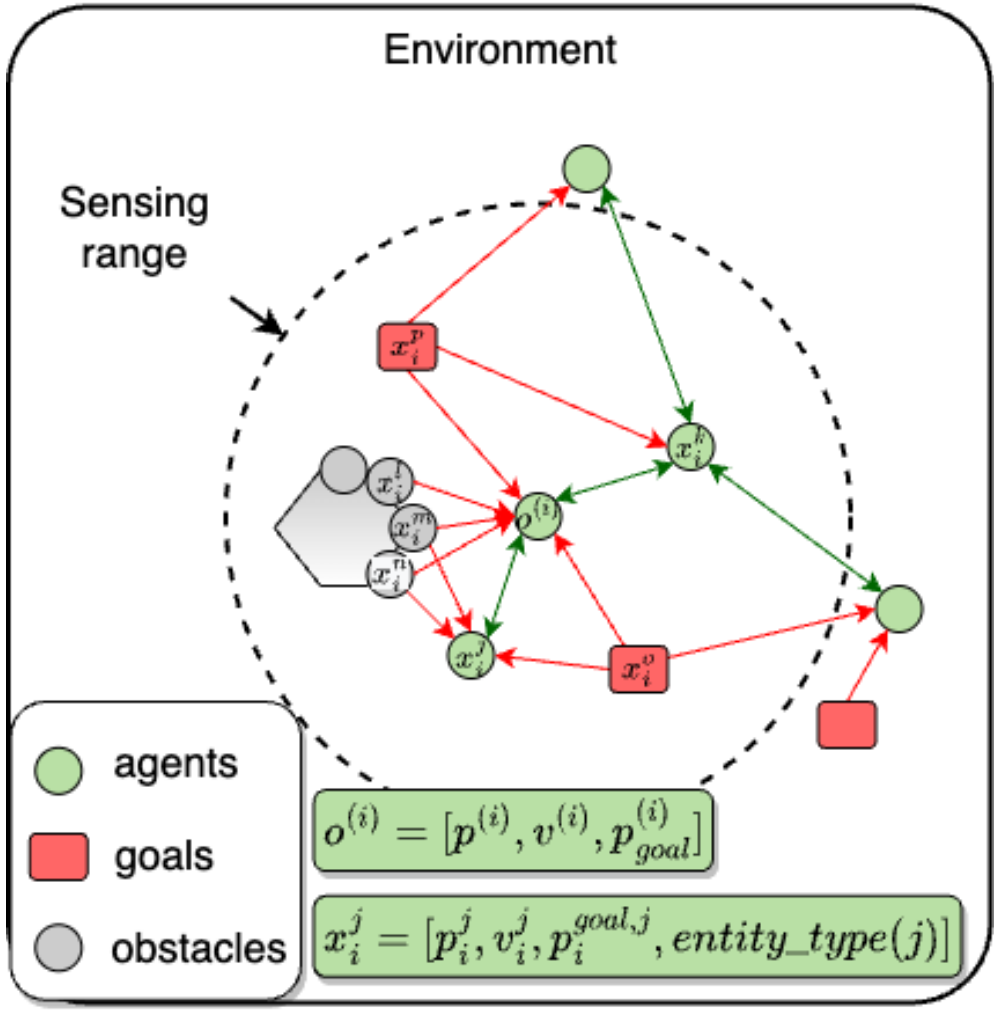


In practice, we just have local information about the neighborhood
And naïve concatenation of neighborhood information doesn't work

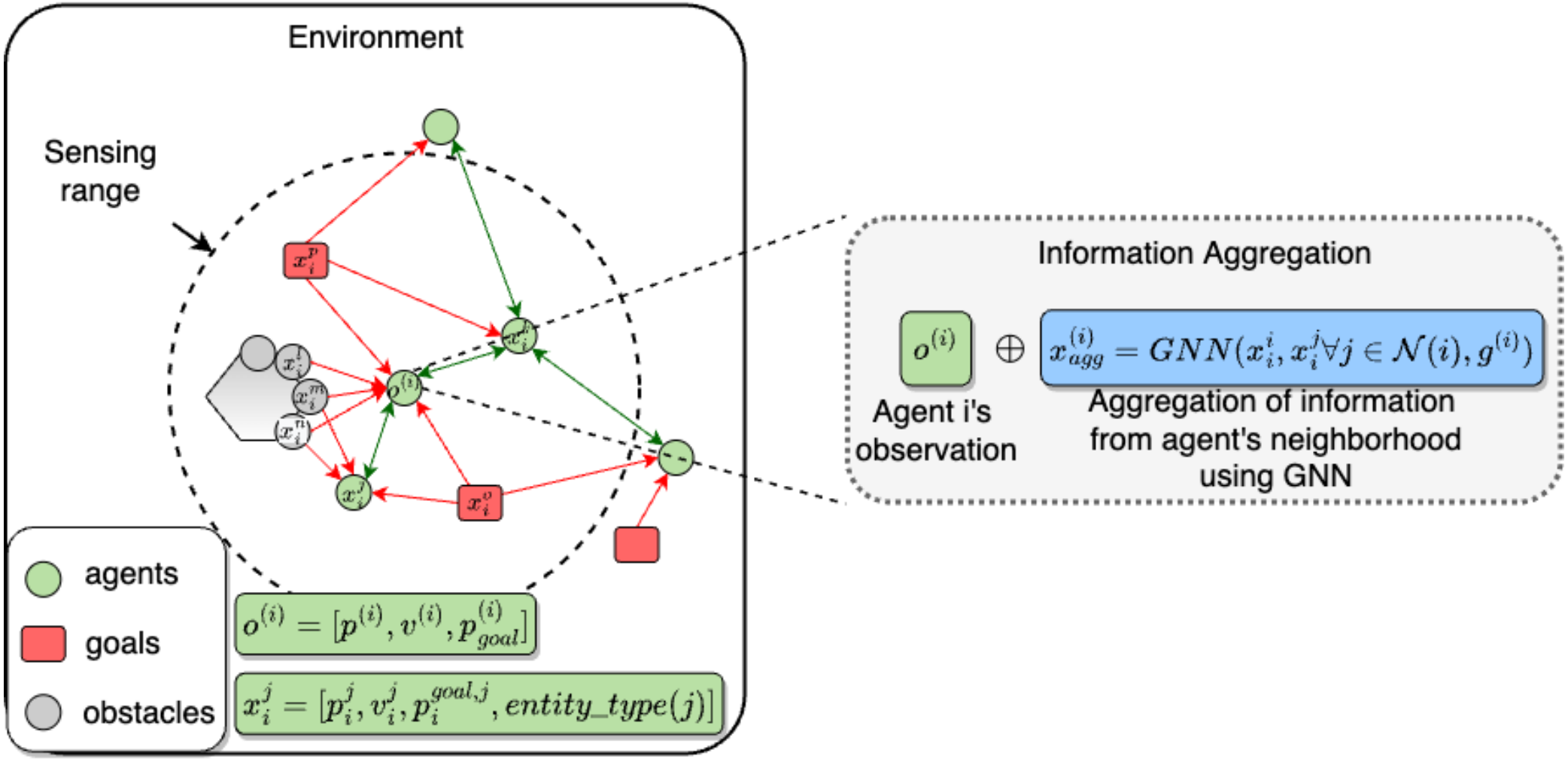
Motivating Experiment



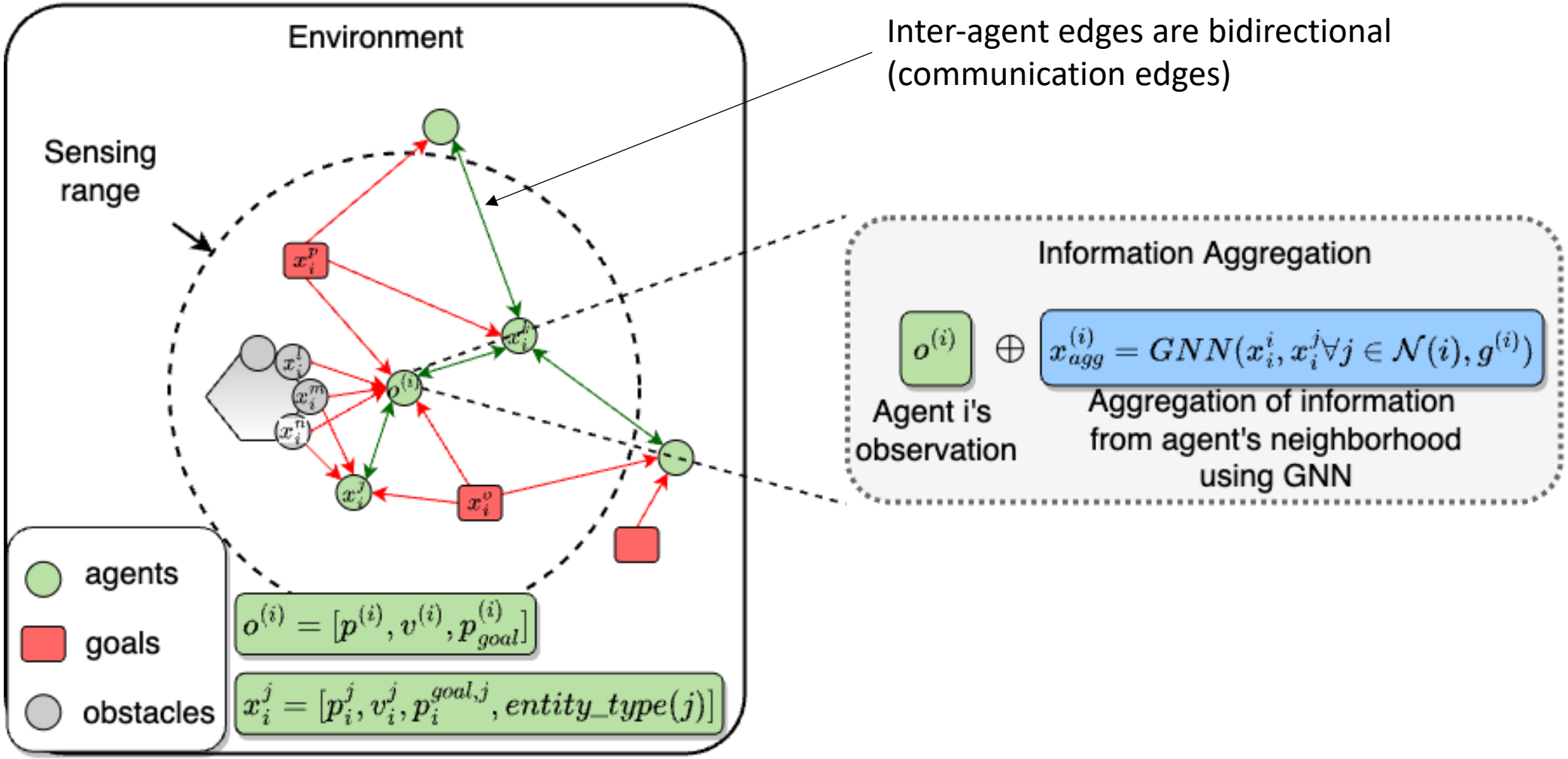
Method


Environment
Sensing range
agents
goals
obstacles
$o^{(i)} = [p^{(i)}, v^{(i)}, p^{(i)}_{goal}]$
$x^j_i = [p^j_i, v^j_i, p^{goal,j}_i, entity\_type(j)]$

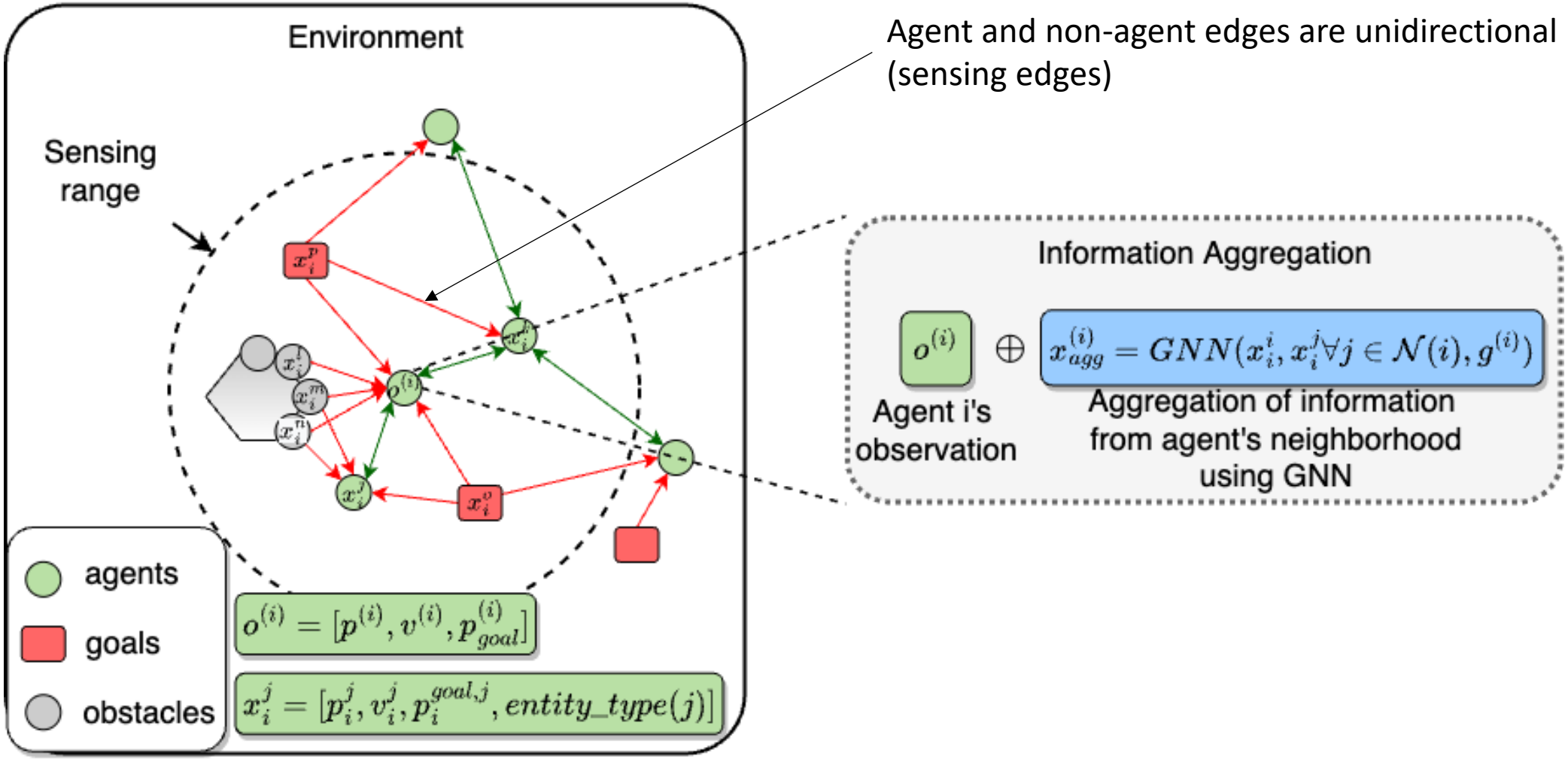
Method


Environment
Sensing range
agents
goals
obstacles
$o^{(i)} = [p^{(i)}, v^{(i)}, p^{(i)}_{goal}]$
$x^j_i = [p^j_i, v^j_i, p^{goal,j}_i, entity\_type(j)]$
Information Aggregation
$o^{(i)}$
$\oplus$
$x^{(i)}_{agg} = GNN(x^i_i, x^j_i \forall j \in \mathcal{N}(i), g^{(i)})$
Agent i's observation
Aggregation of information from agent's neighborhood using GNN

Method


Environment
Sensing range
agents
goals
obstacles
$o^{(i)} = [p^{(i)}, v^{(i)}, p^{(i)}_{goal}]$
$x^j_i = [p^j_i, v^j_i, p^{goal,j}_i, entity\_type(j)]$
Inter-agent edges are bidirectional (communication edges)
Information Aggregation
$o^{(i)}$
$\oplus$
$x^{(i)}_{agg} = GNN(x^i_i, x^j_i \forall j \in \mathcal{N}(i), g^{(i)})$
Agent i's observation
Aggregation of information from agent's neighborhood using GNN

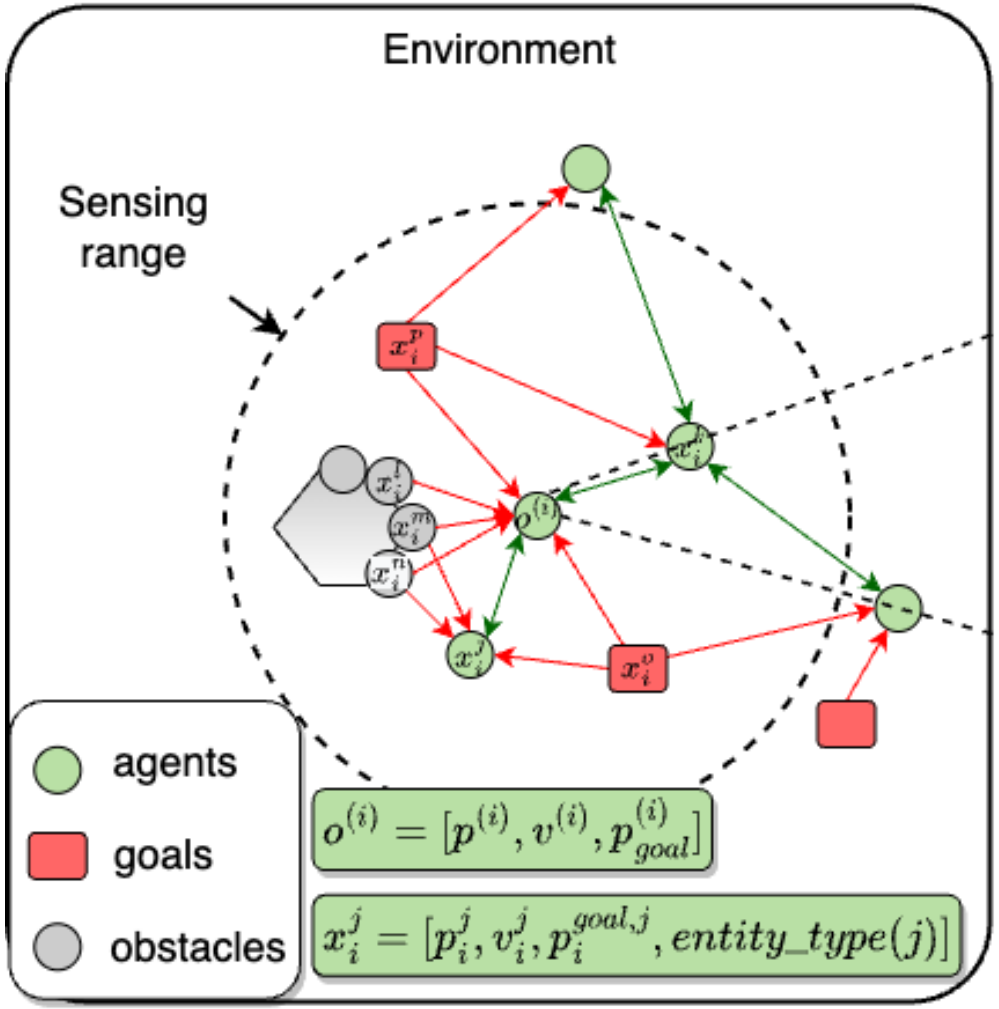
Method

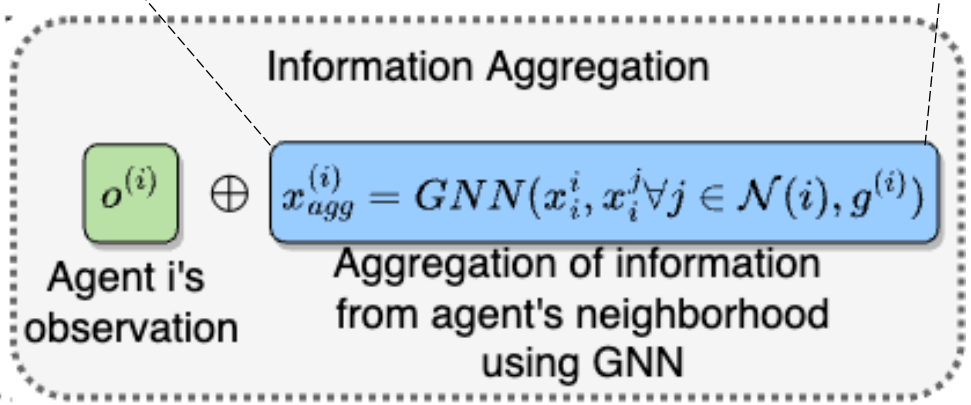

Environment
Sensing range
agents
goals
obstacles
$o^{(i)} = [p^{(i)}, v^{(i)}, p^{(i)}_{goal}]$
$x^j_i = [p^j_i, v^j_i, p^{goal,j}_i, entity\_type(j)]$
Agent and non-agent edges are unidirectional (sensing edges)
Information Aggregation
$o^{(i)}$
$\oplus$
$x^{(i)}_{agg} = GNN(x^i_i, x^j_i \forall j \in \mathcal{N}(i), g^{(i)})$
Agent i's observation
Aggregation of information from agent's neighborhood using GNN

Method

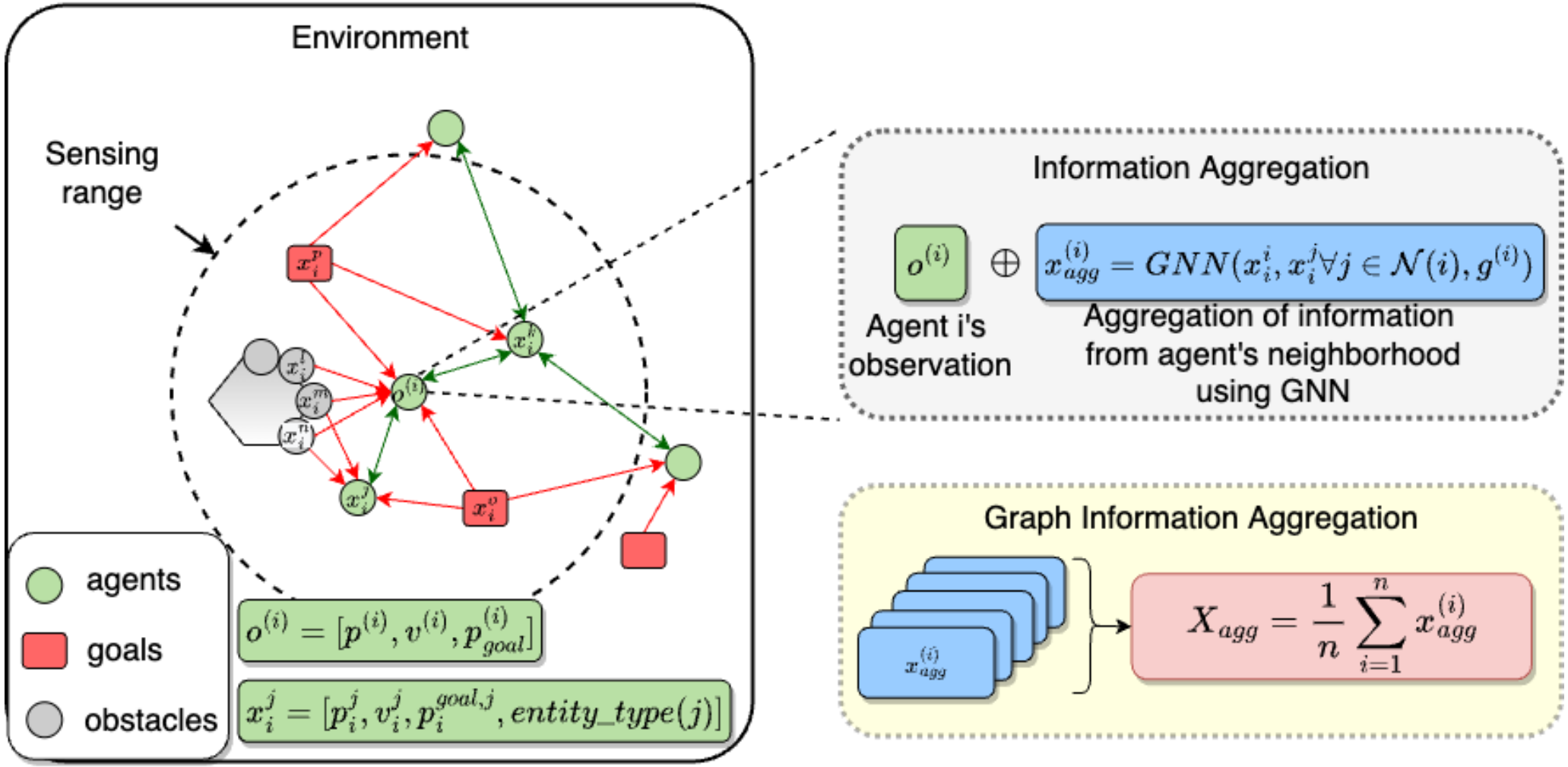
$$x_i^{'} = W_1 \cdot x_i + \sum_{j=\mathcal{N}(i)} \alpha_{i,j} W_2 \cdot x_j$$

$$\alpha_{i,j} = \text{softmax}\left(\frac{(W_3 \cdot x_i)^T (W_4 \cdot x_j + W_5 \cdot e_{i,j})}{\sqrt{c}} \right)$$

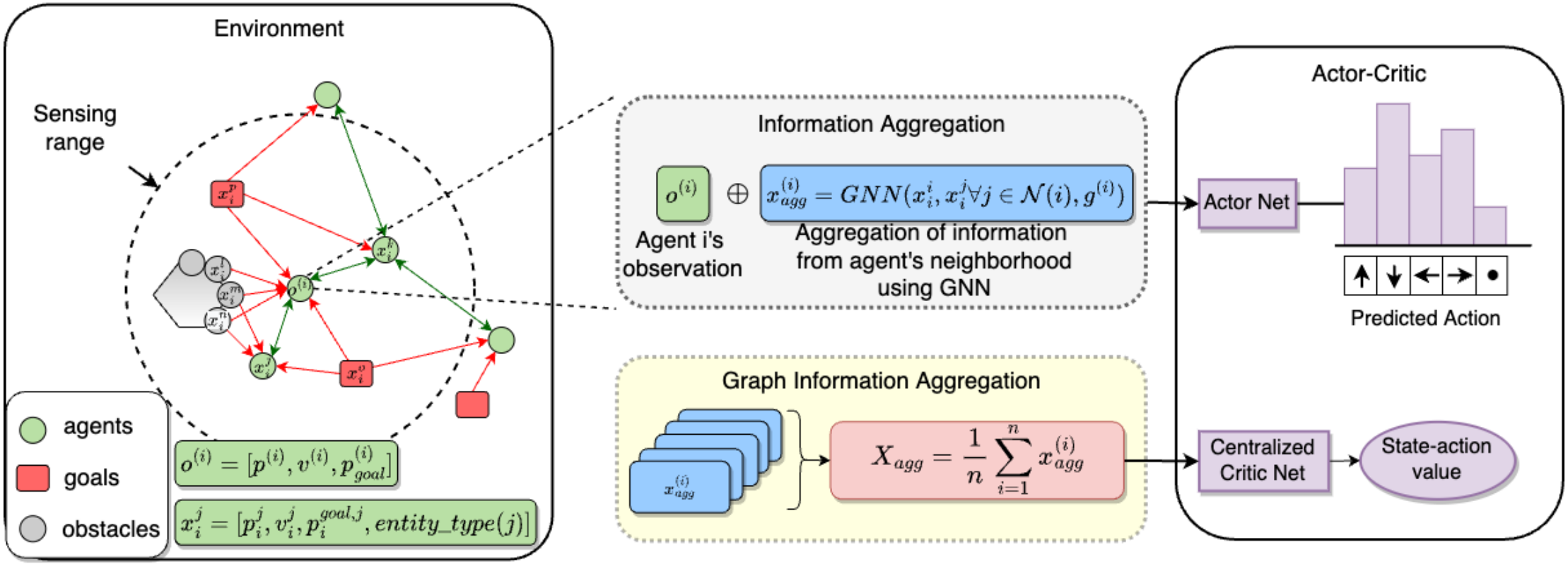




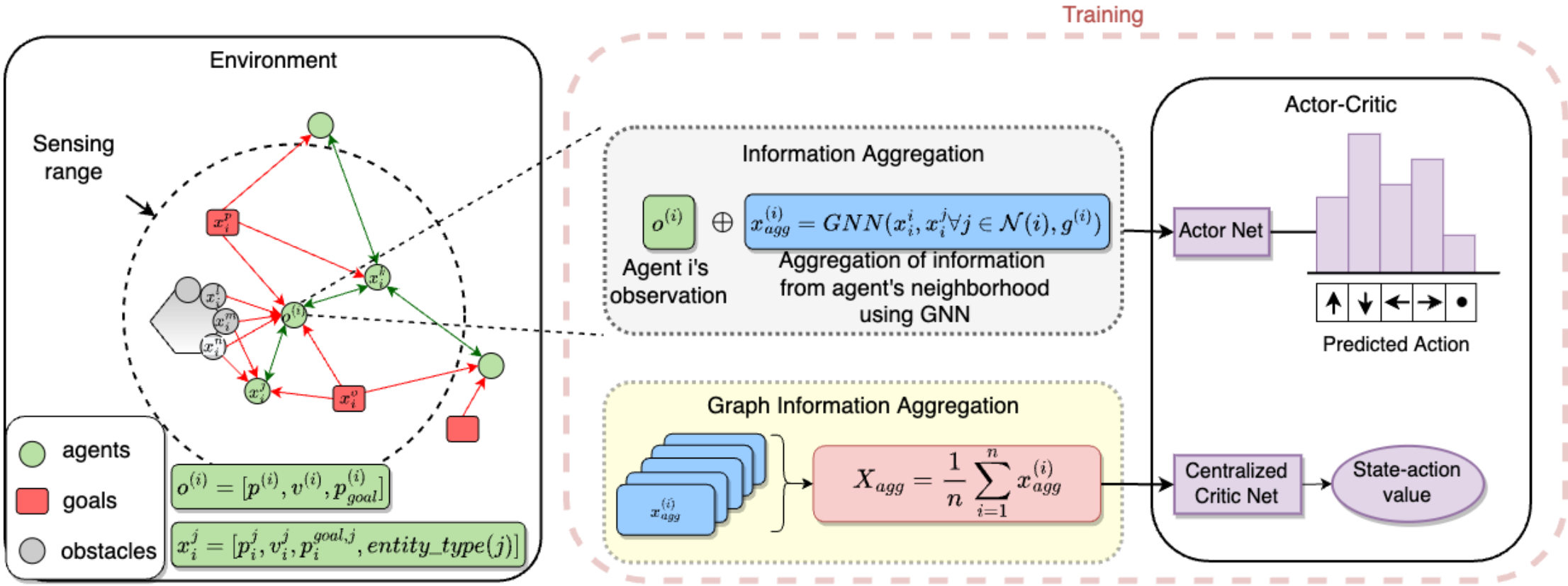
Method


Environment
Sensing range
agents
goals
obstacles
$o^{(i)} = [p^{(i)}, v^{(i)}, p^{(i)}_{goal}]$
$x^j_i = [p^j_i, v^j_i, p^{goal,j}_i, entity\_type(j)]$
Information Aggregation
$o^{(i)}$
Agent i's observation
$\oplus$
$x^{(i)}_{agg} = GNN(x^i_i, x^j_i \forall j \in \mathcal{N}(i), g^{(i)})$
Aggregation of information from agent's neighborhood using GNN
Graph Information Aggregation
$x^{(i)}_{agg}$
$X_{agg} = \frac{1}{n} \sum_{i=1}^{n} x^{(i)}_{agg}$

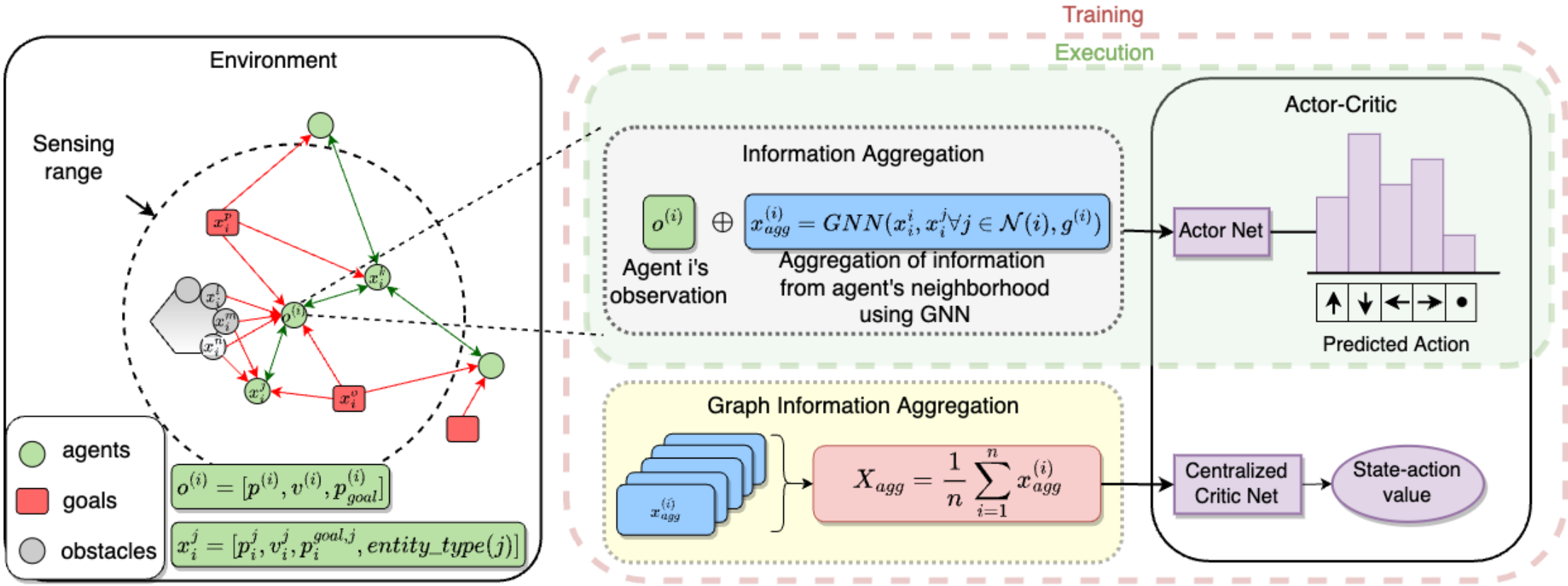
Method


Environment
Sensing range
agents
goals
obstacles
$o^{(i)} = [p^{(i)}, v^{(i)}, p^{(i)}_{goal}]$
$x^j_i = [p^j_i, v^j_i, p^{goal,j}_i, entity\_type(j)]$
Information Aggregation
$o^{(i)}$
$\oplus$
$x^{(i)}_{agg} = GNN(x^i_i, x^j_i \forall j \in \mathcal{N}(i), g^{(i)})$
Agent i's observation
Aggregation of information from agent's neighborhood using GNN
Graph Information Aggregation
$X_{agg} = \frac{1}{n}\sum_{i=1}^{n} x^{(i)}_{agg}$
Actor-Critic
Actor Net
Predicted Action
Centralized Critic Net
State-action value

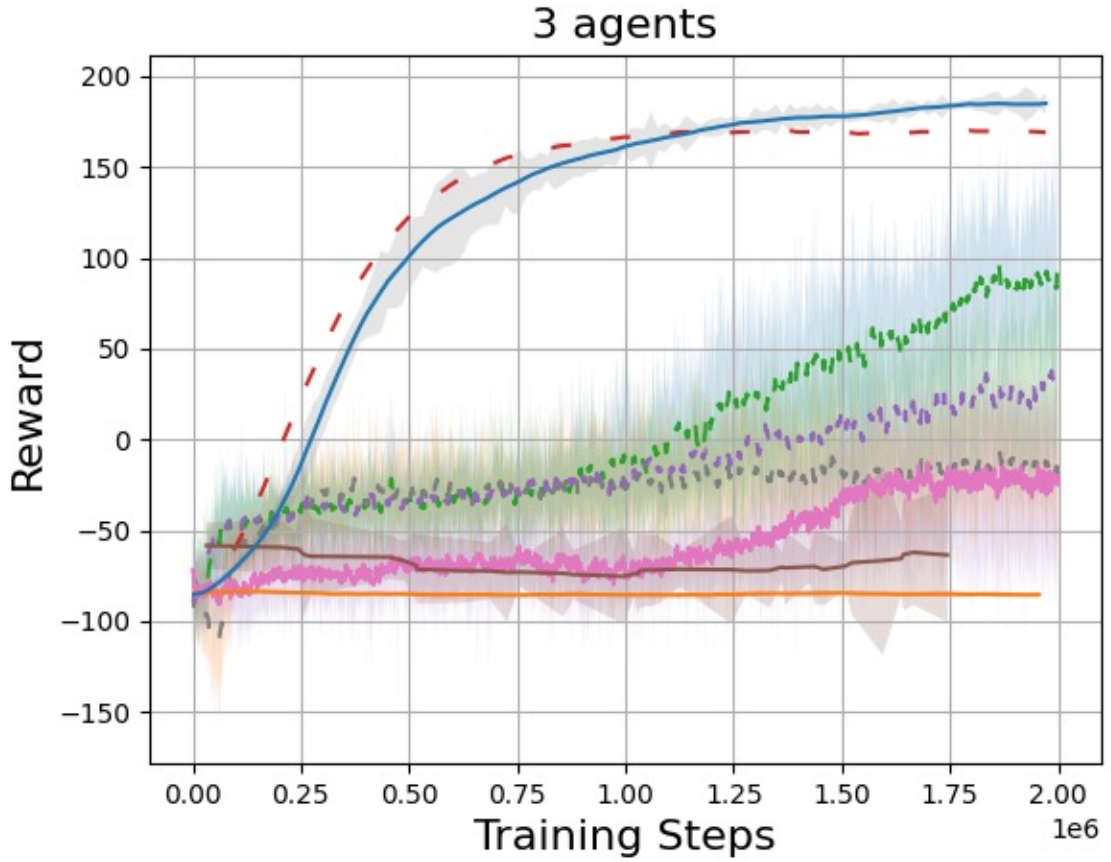
Method

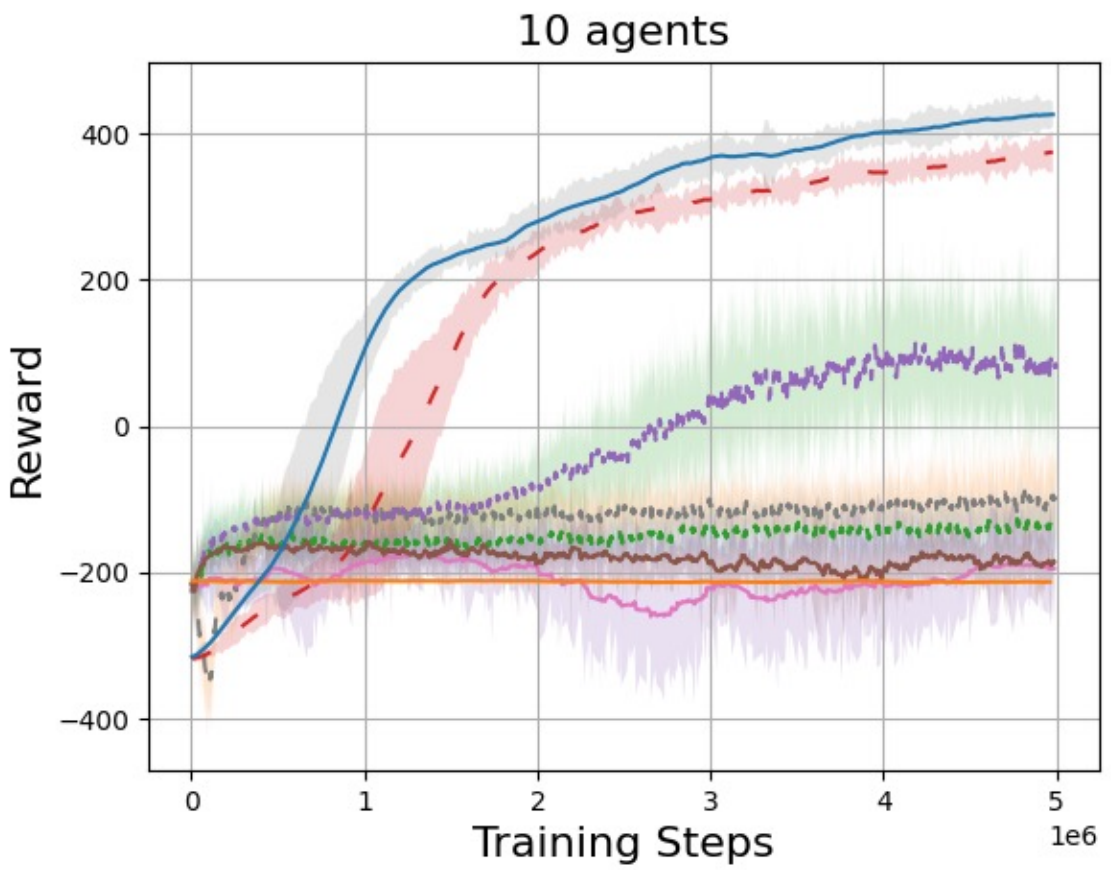

Training
Environment
Sensing range
agents
goals
obstacles
$o^{(i)} = [p^{(i)}, v^{(i)}, p^{(i)}_{goal}]$
$x^j_i = [p^j_i, v^j_i, p^{goal,j}_i, entity\_type(j)]$
Information Aggregation
$o^{(i)}$
$\oplus$
$x^{(i)}_{agg} = GNN(x^i_i, x^j_i \forall j \in \mathcal{N}(i), g^{(i)})$
Agent i's observation
Aggregation of information from agent's neighborhood using GNN
Graph Information Aggregation
$X_{agg} = \frac{1}{n}\sum_{i=1}^{n} x^{(i)}_{agg}$
Actor-Critic
Actor Net
Predicted Action
Centralized Critic Net
State-action value

Method


Training
Execution
Environment
Sensing range
agents
goals
obstacles
$o^{(i)} = [p^{(i)}, v^{(i)}, p^{(i)}_{goal}]$
$x^j_i = [p^j_i, v^j_i, p^{goal,j}_i, entity\_type(j)]$
Information Aggregation
$o^{(i)}$
$\oplus$
$x^{(i)}_{agg} = GNN(x^i_i, x^j_i \forall j \in \mathcal{N}(i), g^{(i)})$
Agent i's observation
Aggregation of information from agent's neighborhood using GNN
Graph Information Aggregation
$X_{agg} = \frac{1}{n}\sum_{i=1}^{n} x^{(i)}_{agg}$
Actor-Critic
Actor Net
Predicted Action
Centralized Critic Net
State-action value

Experiments: Sample complexity





Global Information RMATD3 RQMIX RVDN RMAPPO

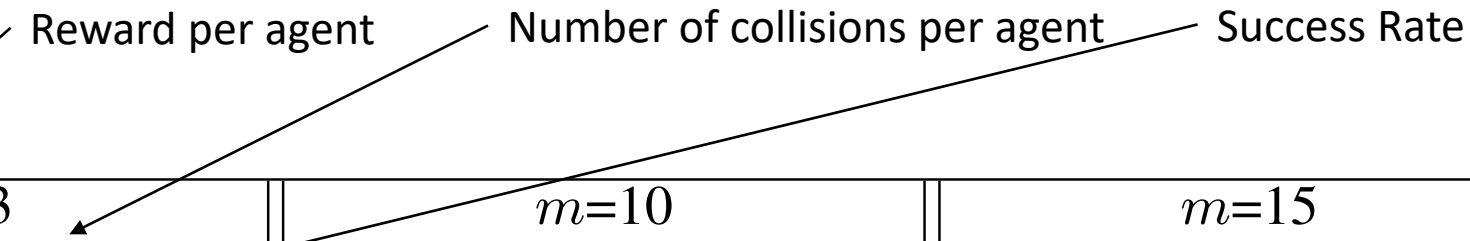
Local Information GPG (dynamic) DGN + ATOC EMP InforMARL

DINaMo

Experiments: Scalability

Train \ Test	m=3			m=10			m=15		
	R/m	# col/m	$S\%$	R/m	$\#col/m$	$S\%$	R/m	# col/m	$S\%$
$n = 3$	68.31	0.48	100	58.59	1.60	100	53.19	2.24	100
$n = 7$	58.30	0.61	100	53.25	1.43	99	46.39	2.31	99
$n = 10$	57.27	0.64	99	52.10	1.68	100	48.15	2.20	99

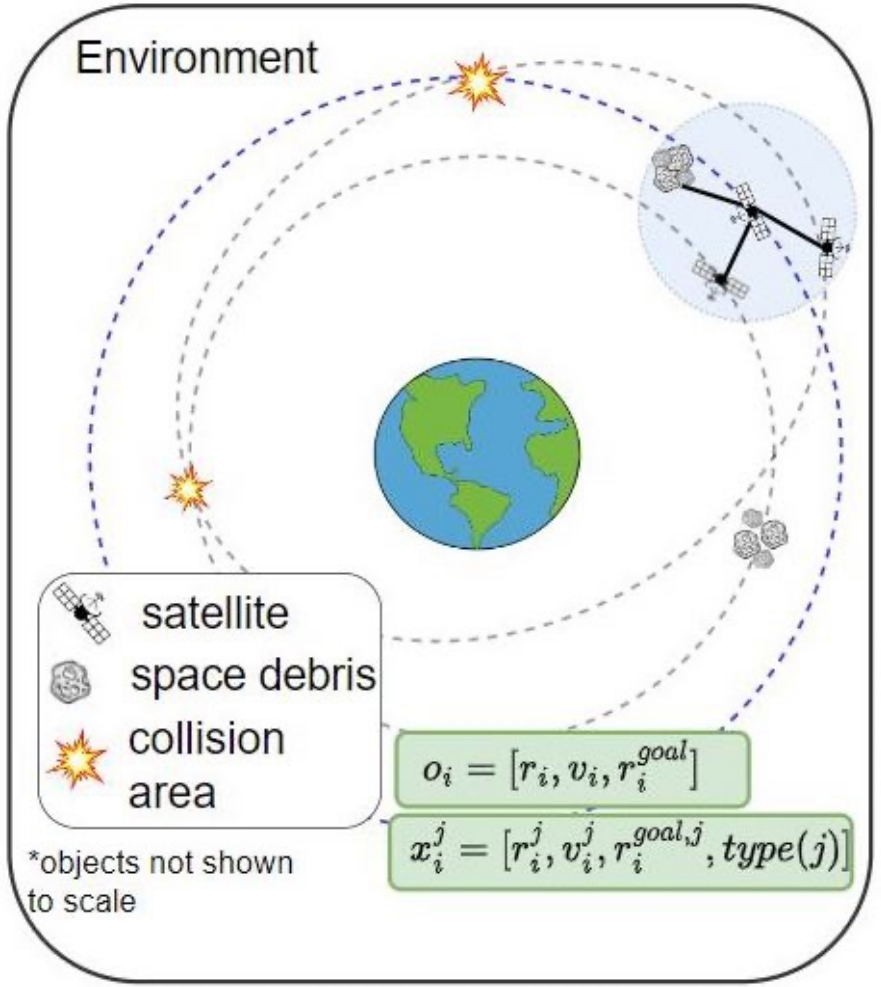
Experiments: Scalability



Train \ Test	m=3			m=10			m=15		
	R/m	# col/m	$S\%$	R/m	$\#col/m$	$S\%$	R/m	# col/m	$S\%$
$n = 3$	68.31	0.48	100	58.59	1.60	100	53.19	2.24	100
$n = 7$	58.30	0.61	100	53.25	1.43	99	46.39	2.31	99
$n = 10$	57.27	0.64	99	52.10	1.68	100	48.15	2.20	99

Satellite Environment

- Concurrent work in leveraging transfer learning for satellite environment
- More complex non-linear dynamics for all entities in the environment



Conclusions and Future Work

- Introduced a graph-based algorithm for scaling standard MARL algorithms to arbitrary scenarios.
- Uses just neighborhood information instead of global information required by previous methods.

Project Website