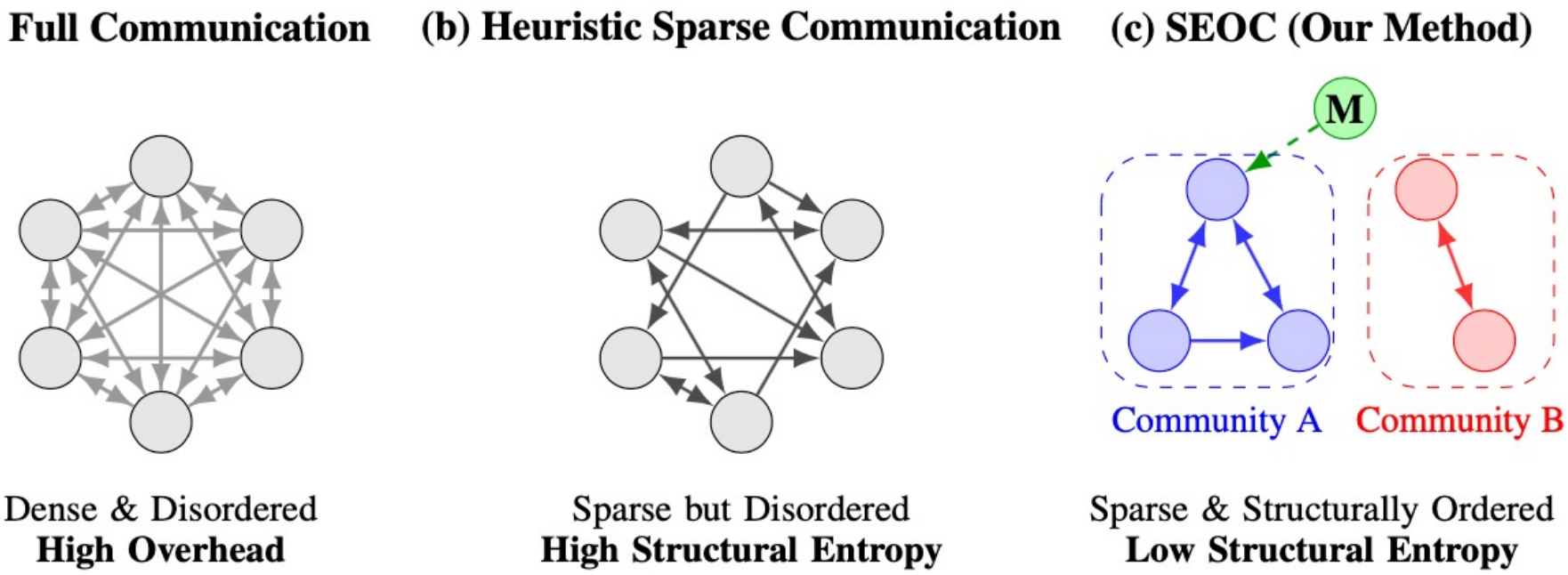
NEURAL INFORMATION PROCESSING SYSTEMS

Entropy Optimized Communication for Multi-Agent Reinforcement Learning

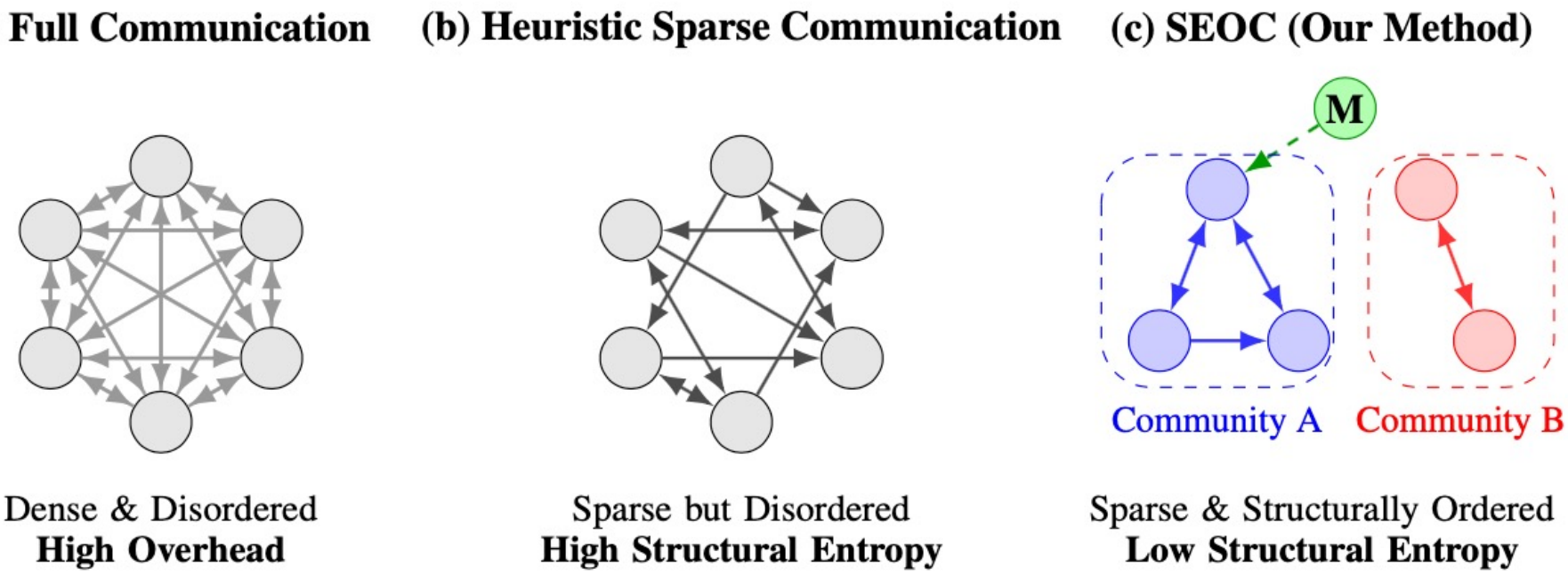
Wei Du, Benyu Wu, Wei Guo, Zhongmin Yan,
Guoxian Yu, **Lizhen Cui***

Motivation



- (a) Full communication protocol is dense and suffers from high overhead.
- (b) Typical sparse communication methods prune edges but often result in a structurally disordered graph with high structural entropy.

Motivation



- (c) SEOC explicitly optimizes for low structural entropy, leading to a sparse graph with clear, interpretable community structures (Community A and B) and targeted inter-community links. (e.g., a Medivac agent 'M' communicating with specific agents in a community in SMAC scenario)

Method

Structural Optimization (Who to talk to)

- Computes pairwise logits for all agents:

$$s_{ij} = f_{\theta_{\text{edge}}}([h_i, h_j])$$

- Samples sparse topology via Gumbel-Softmax:

$$A_{ij} \sim \text{Gumbel} - \text{Softmax}(s_{ij})$$

- Minimizes entropy to discover communities:

$$\mathcal{L}_{\text{SE}} = \lambda_{\text{inter}}\mathcal{L}_{\text{inter}} + \lambda_{\text{intra}}\mathcal{L}_{\text{intra}}$$

Content Optimization (What to say)

- Formulate Infor Bottleneck objective:

$$\max I\big(m_i;\ V(s)\big) - \beta I(m_i;\ h_i)$$

- Sample messages via reparameterization:

$$m_i \sim \mathcal{N}\big(\mu_\psi(h_i), \sigma_\psi^2(h_i) I\big)$$

- Optimize the variational upper bound:

$$\mathcal{L}_{\mathrm{XIB}} = \beta D_{\mathrm{KL}} - \mathbb{E}\big[\log q_\xi(V(s)|m_i)\big]$$

GNN Aggregation (How to process)

- Project node features for messaging:

$$\hat{m}_i^{(l)} = \eta^{(l)}(h_i^{(l)})$$

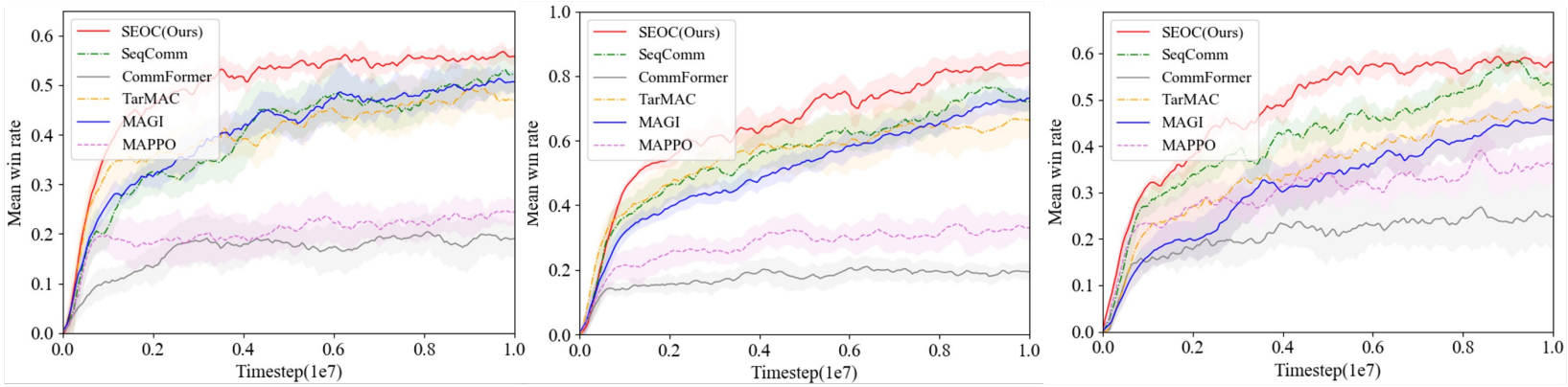
- Aggregate over sparse active edges:

$$\bar{m}_i^{(l)} = \sum_j A_{ij} \cdot \hat{m}_j^{(l)}$$

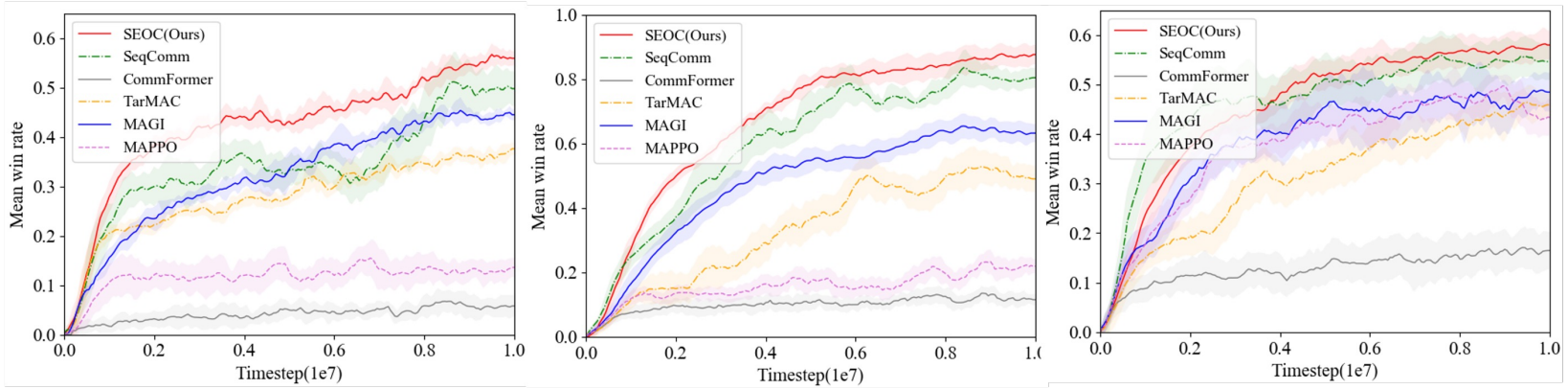
- Update node states for next layer:

$$h_i^{(l)} = \phi^{(l)}\left(h_i^{(l-1)} || \hat{m}_i^{(l)} || \bar{m}_i^{(l)}\right)$$

Experiments



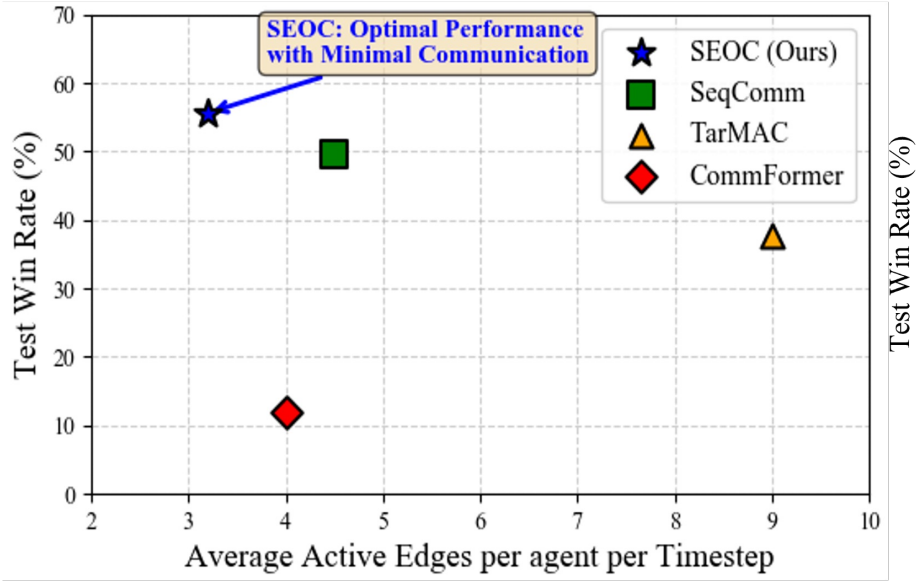
(a) terran 5 vs 5 (b) protoss 5 vs 5 (c) zerg 5 vs 5

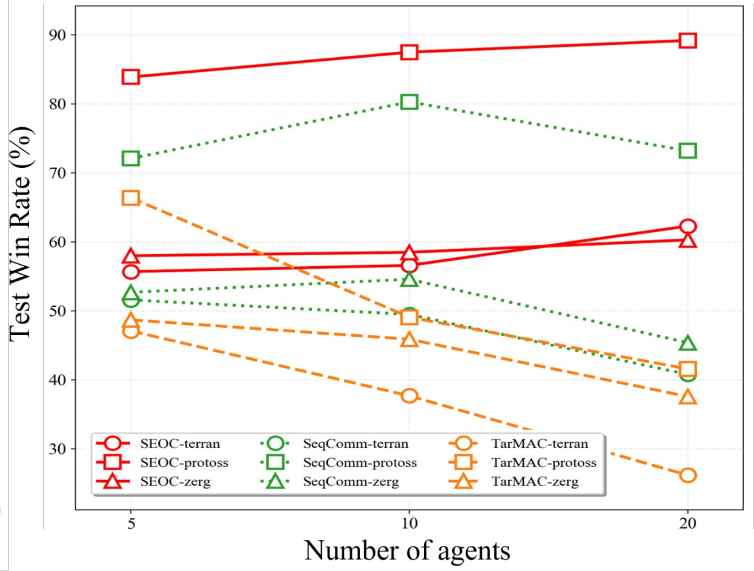


(d) terran 10 vs 10 (e) protoss 10 vs 10 (f) zerg 10 vs 10

- **Superior Overall Performance:** **SEOC achieves faster convergence, greater training stability, and higher final performance than all baselines across various maps in SMACv2.**

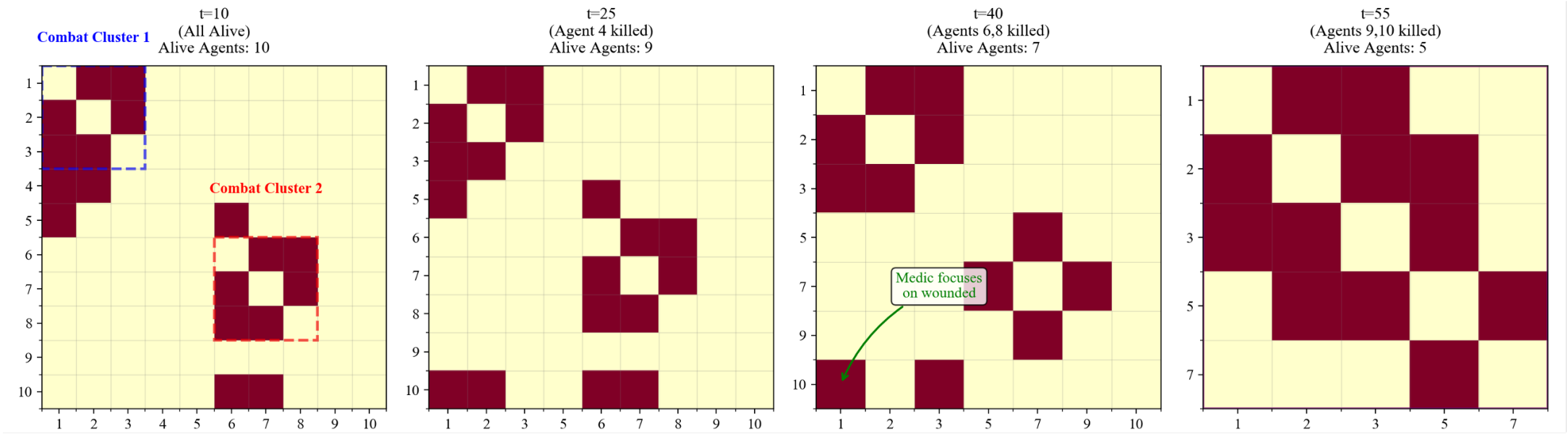
Experiments





- **Communication Efficiency:** SEOC achieves the highest task win rate with significantly fewer active edges compared to all baselines.
- **Scalability**: SEOC exhibits robust scalability, maintaining high and stable win rates as the number of agents increases from 5 to 20.

Experiments



- **Interpretability :** **The learned communication matrices reveal highly interpretable community structures, naturally forming task-specific coordination patterns such as Medivacs selectively linking to wounded units.**

Takeaways

- **Principled Topology Design:** First to introduce Structural Entropy as a differentiable objective for MARL communication, replacing heuristic pruning with a theoretically grounded criterion to discover sparse, modular topologies.
- **Dual-Level Joint Optimization:** Co-designs topology structure and message semantics in a unified framework, ensuring agents communicate with the right partners and exchange only task-relevant information.
- **Emergent Interpretability:** Learns highly interpretable graphs that align with functional roles (e.g., Medivacs connecting to wounded units), quantitatively confirmed by high NMI and Modularity metrics.

SHANDONG UNIVERSITY

C-FAIR
Joint SDU-NTU Centre for Artificial Intelligence Research (C-FAIR)

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

NEURAL INFORMATION PROCESSING SYSTEMS

Thank you for your attention

Wei Du, Benyu Wu, Wei Guo, Zhongmin Yan,
Guoxian Yu, **Lizhen Cui***