To Be or Not to Be: Bayesian Survival Pruning for Multi-Agent System

Anonymous ACL submission

Abstract

To address the challenges of communication topology redundancy and Newcomb's paradox in open-world multi-agent systems, we propose **Agent Bayesian Out(ABO)**, an uncertainty-aware framework. By modeling edge weights in communication graphs as probabilistic random variables, ABO innovatively integrates three key mechanisms: Bayesian edge weight representation, Gaussian process priors, and Markov Chain Monte Carlo-based graph sampling. This integration establishes a *dynamic uncertainty propagation model*, overcoming the limitations of conventional methods in stochastic modeling and open-scenario adaptation. Experiments demonstrate that ABO effectively identifies low-contribution nodes and redundant connections, achieving an average accuracy improvement of **1.8%–2.59%** on benchmarks. Notably, it exhibits enhanced transfer capability in open-domain QA tasks. Ablation studies confirm the synergistic effects of components through a *Bayesian-uncertainty sampling mechanism*, while its *zero additional inference overhead* provides a novel paradigm for distributed agent collaboration. The proposed framework outperforms existing *state-of-the-art* (SOTA) methods across multiple metrics and achieves cutting-edge performance.

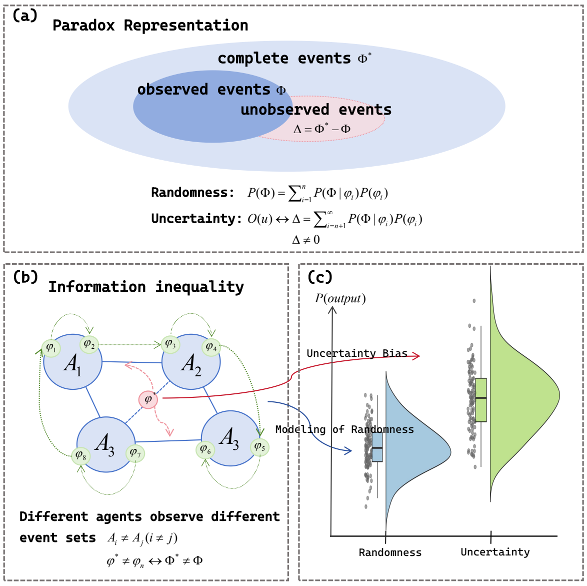

Figure 1: (a) The mathematical representation and interpretation of the Newcomb' paradox. (b) In Multi-Agent Systems, information asymmetry among agents restricts their cognitive and expressive capabilities, leading to uncertainty that is not captured within the scope of stochastic modeling. (c) Modeling uncertainty can, to some extent, compensate for the bias caused by considering only randomness.

1 Introduction

In recent years, multi-agent systems based on large language models have demonstrated significant potential in solving complex tasks (Yang et al., 2024a; Gu et al., 2024; Chen et al., 2024). By emulating the collaborative nature of human teams (Gao et al., 2023a; Yen et al.), these systems enable multiple agents to communicate and cooperate with each other, thereby substantially enhancing the efficiency and quality of task solutions (Islam et al., 2022; McClellan et al., 2024; Sarkar et al., 2022)

However, Multi-Agent Systems (MAS) are commonly confronted with the challenges of high token consumption and inefficiency (Dewes and Dimitrova, 2024; Zhang et al., 2025; Epperson et al., 2025). To address communication efficiency in MAS, existing work has explored various approaches, including dynamic pruning (You et al., 2024), information filtering (Mao et al., 2020), and cross-modal collaboration (Peigne-Lefebvre et al., 2025; Wu et al., 2025). These studies have provided a scalable theoretical framework for reducing communication overhead in MAS by breaking through the limitations of traditional methods. Despite these advancements, existing methods still face two major limitations: (1) The sampling probability values, measured from a stochastic perspective, fail to capture the inherent uncertainty of communication strength, resulting in insufficient sensitivity for redundancy detection (Tran et al., 2021); (2) The simplified assumptions regarding communication graph structures do not align well with the complex dependencies in open-world scenarios, potentially leading to suboptimal pruning decisions (Pezeshkpour et al., 2024).

Typically, the observed communication relationships in agent systems and the event sets of individual agents are assumed to be equivalent to the true complete event set. Consequently, the observed events can be modeled stochastically based on Bayes' theorem (Tran et al., 2025). However, in open and complex systems, factors such as individual agents' cognitive limitations and information asymmetry mean that the observed event set may not be the true complete event set. Instead, the observed event set may be a subset of the complete event set, leading to scenarios akin to the **Newcomb' paradox**. This unobserved portion of the event set cannot be modeled by existing stochastic methods, necessitating the introduction of event uncertainty to compensate for the insufficiencies of stochasticity, as shown in Figure 1 . Therefore, we propose **Agent Bayesian Out(ABO)**, which effectively addresses the Newcomb' paradox that multi-agent systems may encounter when dealing with open-world complex systems.

We introduce ABO, a multi-agent system (MAS) collaborative reasoning framework based on uncertainty perception, which effectively circumvents the occurrence of the Newcomb' paradox, thereby enhancing decision-making expectations. The edge weights in the communication graph are inherently probabilistic random variables, and their uncertainty characteristics reflect the confidence level of a specific communication path's contribution to the task. By explicitly modeling this uncertainty, the system can more accurately identify and eliminate redundant elements during multiple rounds of collaboration. The contributions of this study can be summarized as follows:

- We introduce uncertainty quantification into multi-agent communication optimization, proposing a probabilistic communication graph optimization method based on uncertainty and an accurate inference approach.
- We efficiently extend the optimization of multi-agent systems to complex open-world systems.
- Without incurring additional inference overhead, ABO outperforms existing baselines in multiple scenario-based reasoning tasks.

2 Related Work

2.1 Dynamic Graph Structure Modeling and Uncertainty Quantification in MAS

Dynamic graph structure modeling of MAS is a cornerstone for enhancing efficiency (Ban et al., 2025; Wang et al., 2025; Dong et al., 2025). Traditional approaches typically rely on the Markov Decision Process (MDP) framework, describing agent interactions through states, actions, and transition probabilities (Zhang et al., 2024b). However, real-world scenarios often involve uncertainties in system topology (Shan et al., 2020a,b), necessitating the introduction of graph-theoretic methods such as zero-forcing set coloring conditions (Gomes et al., 2024), combined with probabilistic density functions and fuzzy set theory to quantify the impact of parametric uncertainties on dynamic behaviors. For instance, a topological modeling method based on ternary symbolic matrices has been proposed, which verifies the strong structural controllability of the system through coloring conditions and designs polynomial-complexity algorithms for selecting the minimum leader set, ensuring system robustness in dynamic changes (Chen and Li, 2024). Additionally, proxy modeling methods have been employed to approximate complex dynamic models, analyzing the propagation of uncertainties through Monte Carlo simulations to reduce computational costs and improve prediction accuracy.

2.2 Synergistic Optimization Mechanism of Uncertainty and Communication Graph

The synergistic optimization mechanism of dynamic communication topology and uncertainty is crucial for the efficient operation of MAS (Castro et al., 2022). Traditional static topologies are ill-suited to real-time environmental fluctuations (Teh et al., 2022), thus requiring optimization through intelligent algorithms and distributed strategies (Zhang et al., 2024a; Duan et al., 2024). In terms of uncertainty compensation, Multi-Agent Reinforcement Learning (MARL) enables adaptive adjustments through empirical replanning strategies (Utke et al., 2025; Zhou et al., 2025). Building on this foundation, AgentDropout (Wang et al., 2025) outperforms various types of multi-agent systems

(MAS) and AgentPrune in terms of performance and token efficiency. Additionally, a method for pruning inter-agent communication edges has been introduced (Dong et al., 2025), which defines communication redundancy and incorporates the pruning approach of AgentPrune into MAS.

2.3 LLM-Enabled Breakthroughs in Open-World Agent Modeling

Large Language Models (LLMs) have revolutionized the cognitive capabilities and collaborative paradigms of open-world agents (Yang et al., 2024b; Yu et al., 2024). Based on the role-memory-planning-action framework, agents enhance their generalization abilities through prompt engineering and self-evolution mechanisms (Sanwal, 2025; Lee and Tiwari, 2024). These agents not only overcome the limitations of traditional rule-based methods but also demonstrate human-like autonomous behaviors and the potential for emergent collective intelligence in diverse scenarios such as social simulations and urban system modeling, paving new pathways for constructing high-fidelity, scalable open-world simulation systems (Gao et al., 2023b). ToolEmu (Ruan et al., 2024) simulates high-risk tool execution environments through language models, creating dynamic sandboxes to detect long-tail risks and validating the breakthrough of semantic-level simulation over physical constraints. TrustAgent (Hua et al., 2024) pioneers a constitution-driven framework, internalizing safety rules into the decision-making process to reduce risky behaviors while improving task completion rates. Moreover, systems like Math Agents enhance mathematical reasoning accuracy through code self-verification, marking a transition from perception-based execution to autonomous cognition in agents (Cheng et al., 2024).

3 Methodology

This section provides a detailed exposition of Agent Bayesian Out, a novel graph communication framework designed to enhance the communication efficiency and task performance of MAS. Drawing inspiration from the core concepts of AgentDropout, we introduce uncertainty-aware modeling and more accurate statistical sampling techniques, as illustrated in Figure 2. Specifically, by leveraging the perspective of uncertainty, our framework can precisely identify and remove agent nodes and communication edges that contribute less to the current task. Without incurring additional computational overhead, this approach effectively utilizes statistical methods to achieve higher accuracy and demonstrates transferable representation capabilities in complex open-world systems.

3.1 Motivation

We can model the communication optimization problem in MAS within the framework of uncertainty theory, which first requires understanding the relationship between the observed event set and the complete event set. Suppose there exist Φ and Φ^*, where Φ denotes the set of events observed under true conditions, and Φ^* represents the true complete set of events. Thus, for all $\phi_i \in \Phi^*$, where ϕ_i represents the i-th event and $i = 1, 2, \ldots, n$ with $n \geq k$, we define $\Phi = \{\phi_1, \phi_2, \ldots, \phi_n | n \geq k\}$. If $k \to \infty$, then $\lim |\Phi^* - \Phi| \leq \epsilon(\to 0)$, implying that $\Phi \Leftrightarrow \Phi^*$, i.e., the observed event set is equivalent to the complete event set. Otherwise, $\Phi \subseteq \Phi^*$, meaning the observed event set is a subset of the complete event set.

The traditional total probability formula solves for the true complete event set Φ^*, given by $P(\Phi) = \lim_{N\to\infty} \sum_{i=1}^{N} P(\Phi|\phi_i)P(\phi_i)$. However, when agents in MAS perform inference, they can only observe Φ, resulting in an actual probability of $\hat{P}(\Phi) = \sum_{i=1}^{n} P(\Phi|\phi_i)P(\phi_i)$. Consequently, there remains a probability deviation of $\Delta = \lim_{m\to\infty} \sum_{n+1}^{m} P(\Phi|\phi_i)P(\phi_i)$. In most deterministic systems, $\Delta \to \delta(\to 0)$ and can thus be modeled deterministically. However, in complex systems, $\delta \nrightarrow 0$, leading to an estimation bias $E(\Delta)$ in stochastic modeling. When the complexity $O(u) \to \infty$, there is a relationship $u \leftrightarrow \Delta$, where $O(\cdot)$ denotes the complexity function and u represents the uncertainty metric. Introducing the uncertainty metric u can mitigate the inference bias of MAS in complex systems.

Existing MAS optimization methods such as AgentPrune and AgentDropout have limitations from uniform pruning and scalar uncertainty modeling, along with oversimplified communication assumptions. ABO addresses these issues by integrating stochasticity and uncertainty modeling, which enables robust task inference in complex systems. It attains precise uncertainty-aware optimization, cuts redundancy and fits dynamic communication patterns.

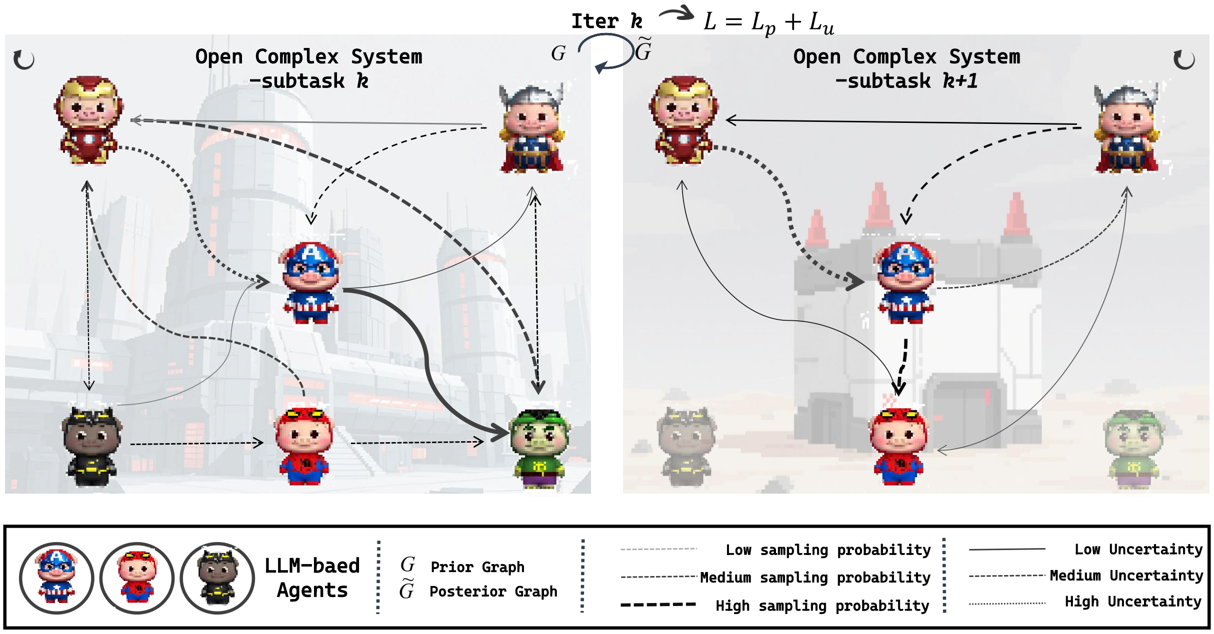

Figure 2: Overall is a Pipeline of ABO.In the same open - world system, agents have different initial information distributions (e.g., left and right agents have inconsistent scenarios). In various sub - tasks, the information agents access is unequal. This causes communication uncertainty beyond randomness. When updating communication graph (both inter - and intra - agent), constraints are imposed on both the randomness - related benefit L_p and the uncertainty - related benefit L_u.

3.2 Problem Formulation and Preliminaries

Following the initialization approach of Agent-Dropout, we model multi-agent communication as a weighted directed graph $\tilde{G} = (V, E)$, where V is the set of agent nodes and E is the set of communication edges. The adjacency matrix set is denoted as $\tilde{A} = \tilde{A}_{\text{intra}} \cup \tilde{A}_{\text{inter}}$, where $\tilde{A}_{\text{intra}} = \bigcup_t \tilde{A}^{(t)}_{\text{intra}}$ contains the intra-round adjacency matrices for each communication round t, and $\tilde{A}_{\text{inter}} = \bigcup_t \tilde{A}^{(t)}_{\text{inter}}$ contains the inter-round adjacency matrices connecting agents across different rounds. N represents the total number of agents.

A core aspect of ABO is the probabilistic modeling of edge weights. For any edge $(v_i, v_j) \in E$ in graph $\tilde{G}$, its weight $\tilde{A}^{(t)}_{ij}$ in round t is treated as a trainable random variable and follows a Gaussian process prior:

$$\tilde{A}^{(t)}_{ij} \sim \mathcal{N}(\mu^{(t)}_{ij}, (\sigma^{(t)}_{ij})^2) \tag{1}$$

Here, $\mu^{(t)}_{ij} \in [0, 1]$ represents the expected communication strength (i.e., the mean probability of edge existence), while $(\sigma^{(t)}_{ij})^2$ quantifies the uncertainty of the edge weight. In the subsequent inference phase, the actual task-executing communication graph G will be sampled from this learned probability distribution, ensuring that it is a Directed Acyclic Graph (DAG).

3.3 To Sample or Not to Sample

Initialization and Objective: Parameters (mean/variance) for intra-round ($\tilde{A}_{\text{intra}}$) and inter-round ($\tilde{A}_{\text{inter}}$) adjacency matrices are initialized with Bayesian priors $\mathcal{N}(0.5, 1.0)$, balancing initial communication strength and uncertainty. Optimization during node deactivation aims to maximize MAS task performance by evolving intra-round structure:

$$\arg\max_{\tilde{A}_{\text{intra}}} \mathbb{E}_{G \sim p(G|\tilde{A}_{\text{intra}})}[\mu(G)] \tag{2}$$

Communication graph sampling is modeled as a stochastic policy, leveraging score function estimators to derive unbiased gradients, enabling black-box optimization of non-differentiable objectives. Uncertainty constraints refine the optimization process.

MCMC-based Graph Sampling and Gradient Update: We leverage MCMC to sample complex graph distributions, initializing from a DAG. Each iteration proposes a new candidate DAG G' via edge operations (e.g., random edge addition/deletion) guided by current edge parameters

$(\mu_{ij}^{(t)}, (\sigma_{ij}^{(t)})^2)$. The Metropolis-Hastings acceptance probability α is computed as:

$$\alpha(G'|G_{\text{current}}) = \min\left(1, \frac{\mu(G') \cdot p(G'|\tilde{A}_{\text{intra}})}{\mu(G_{\text{current}}) \cdot p(G_{\text{current}}|\tilde{A}_{\text{intra}})}\right) \quad (3)$$

Here, G_{current} denotes the current graph structure, and $p(G|\tilde{A}_{\text{intra}})$ represents the probability of sampling graph G given the parameters of the intra-round adjacency matrix. G' is accepted or rejected based on α. By collecting M MCMC samples $\{G_m\}_{m=1}^M$, we estimate the gradient with respect to the parameters of $\tilde{A}_{\text{intra}}$ (mainly the mean $\mu_{ij}^{(t)}$):

$$\nabla_{\tilde{A}_{\text{intra}}} \mathbb{E}[\mu(G)] \approx \frac{1}{M} \sum_{m=1}^{M} \mu(G_m) \cdot \nabla_{\tilde{A}_{\text{intra}}} \log p(G_m|\tilde{A}_{\text{intra}}) \quad (4)$$

where

$$p(G_m|\tilde{A}_{\text{intra}}) = \prod_t \prod_{(v_i,v_j)\in E_{m,\text{intra}}^{(t)}} \tilde{A}_{\text{intra}}^{(t)}[i,j] \quad (5)$$

with $\tilde{A}_{\text{intra}}^{(t)}[i,j]$ representing the probability of edge (v_i, v_j) existing in round t (i.e., its mean $\mu_{ij}^{(t)}$). The mean parameters of $\tilde{A}_{\text{intra}}$ are updated using gradient descent based on the computed gradient.

Node Sampling: After optimizing the adjacency matrix parameters, for each intra-round communication graph $G^{(t)}$, we calculate the sum of the weighted in-degrees and out-degrees of each node based on the optimized edge weight means. Setting the node deactivation rate to α_N, we accept nodes that rank in the top α_N proportion in terms of sampling contribution.

3.4 To Connect or Not to Connect

Further confirmation of the remaining agent communication edges is carried out.

Sparse Variational Gaussian Process: After complete uncertain node sampling, we reinitialize the parameters of the intra-round and inter-round adjacency matrices $\tilde{A}_{\text{intra}}^{(t)}$ and $\tilde{A}_{\text{inter}}^{(t)}$. Considering the computational complexity of directly performing posterior inference on Gaussian processes, we employ sparse variational inference, introducing K inducing points $u = \{u_k\}_{k=1}^K$. The variational distribution $q(\tilde{A})$ is assumed to factorize over the edge weights in the form:

$$q(\tilde{A}) = \prod_{t,i,j} N(\tilde{A}_{ij}^{(t)}, s_{ij}^{(t)}) \prod_k q(u_k) \quad (6)$$

$$\begin{aligned}\text{ELBO} = {} & \mathbb{E}_{q(\tilde{A})}[\log p(D|\tilde{A})] \\ & - \text{KL}(q(\tilde{A}) \| p(\tilde{A}))\end{aligned} \quad (7)$$

where D represents the observed data (e.g., task feedback), KL denotes the Kullback-Leibler divergence. $p(D|\tilde{A})$ is the likelihood term, and $p(\tilde{A})$ is the prior term composed of the GP prior and the inducing point prior $p(u)$.

Constrained Optimization Objective: The optimization objective during the edge deactivation phase aims to balance task performance, communication efficiency (sparsity), uncertainty control, and model accuracy:

$$\begin{aligned} L = {} & \mathbb{E}_{G\sim q(G|\tilde{A})}[\mu(G)] + \lambda_1 \cdot \text{ELBO} \\ & - \lambda_2 \sum_{t=1}^{T} \left\| \text{mean}\left(\tilde{A}_{\text{intra}}^{(t)}\right) \right\|_* \\ & - \lambda_3 \sum_{t=2}^{T} \left\| \text{mean}\left(\tilde{A}_{\text{inter}}^{(t)}\right) \right\|_* \\ & - \lambda_4 \sum_{t,i,j} s_{ij}^{(t)} \end{aligned} \quad (8)$$

We maximize the objective L, which integrates expected task performance, an ELBO term for data fidelity, nuclear norm regularization ($||\cdot||_*$, weighted by λ_2, λ_3) on adjacency matrix means to enforce sparsity, and a variational variance penalty (λ_1, λ_4) to stabilize communication uncertainty. Parameter updates are driven by hybrid MCMC sampling, policy gradients, and variational inference outcomes.

Edge Sampling Mechanism: After optimization, consistent with AgentDropout, we retain the top $1 - \beta_E$ proportion of edges with the highest weight values based on the learned edge weight means $\mu_{ij}^{(t)}$, where β_E is the edge deactivation rate.

3.5 Transfer in Open-World Complex Systems

To further address the stronger uncertainty challenges in open-world complex systems, we extend the original binary correctness judgment to an expectation optimization based on semantic evaluation metrics. Let $\mathcal{M}(\mathbf{y}, y^*)$ denote the normalized mean of evaluation metrics (BLEU-4, ROUGE-1, ROUGE-2, and ROUGE-L) between the generated answer $\mathbf{y}$ and the reference answer y^*, and define the indicator function as:

$$s(\mathbf{y}) = \sigma\left(\lambda \cdot \left(\mathcal{M}(\mathbf{y}, y^*) - \tau\right)\right) \tag{9}$$

where $\sigma(\cdot)$ is the sigmoid function, τ is the decision threshold, and λ is the temperature coefficient. This function maps the continuous evaluation metric $\mathcal{M} \in [0, 1]$ to a soft probability $s(\mathbf{y}) \in (0, 1)$ for binary decision-making.

Maintaining the Bayesian framework of Agent Bayesian Out, the loss function is redefined as the expectation of the evaluation metric:

$$\begin{aligned} \mathcal{L} = &-\mathbb{E}_{G \sim q(G|\tilde{A})}\left[\mathbb{E}_{\mathbf{y} \sim p(\mathbf{y}|q,G)} s(\mathbf{y})\right] \\ &+ \lambda \cdot \mathrm{KL}(q(G|\tilde{A}) \| p(G)) \\ &- \lambda_1 \sum_{t=1}^{T} \left\| \mathrm{mean}(\tilde{A}^{(t)}_{\text{intra}}) \right\|_* \\ &- \lambda_2 \sum_{t=2}^{T} \left\| \mathrm{mean}(\tilde{A}^{(t)}_{\text{inter}}) \right\|_* \end{aligned} \tag{10}$$

The gradient computation employs a score function estimator:

$$\begin{aligned} \nabla \mathcal{L} \approx &-\frac{1}{M} \sum_{m=1}^{M} s(\mathbf{y}_m) \nabla \log q(G_m|\tilde{A}) \\ &+ \lambda \nabla \mathrm{KL}(q \| p) \end{aligned} \tag{11}$$

where $\{\mathbf{y}_m, G_m\}_{m=1}^{M}$ are the jointly sampled samples. This design fully inherits the ABO mechanism, replacing the 0-1 correctness labels with differentiable semantic evaluation expectations to achieve smooth transfer to open-ended question-answering scenarios.

3.6 Graph Sampling for Inference

Post-training, communication graphs are sampled from the learned probabilistic adjacency parameters $(\mu_{ij}, \sigma^2_{ij})$ using MCMC methods, ensuring DAG constraints. Multiple samples are aggregated via inference strategies for task execution. Implementation details are provided in the A.1.

4 Experiments

We validated the ABO framework's core mechanisms theoretically in simulations and compared its performance with existing methods on multiple benchmarks, using three different-scale LLMs.

4.1 Experimental Setup

We evaluated our approach on five benchmark datasets, including MMLU (Hendrycks et al., 2020), HumanEval (Peng et al., 2024), GSM8K (Cobbe et al., 2021), AQUA (Wang et al., 2017), and SVAMP (Patel et al., 2021), against multiple comparison methods (Vanilla, CoT, MAS, AgentPrune, and AgentDropout). All experiments maintained consistent agent configurations and random seeds to ensure fair comparisons.

4.2 Validation of Theoretical Effectiveness of Core Mechanisms

To validate the theoretical advantages of ABO's core design, we constructed a synthetic dataset containing 100 samples: feature vectors were sampled from a multivariate Gaussian distribution with zero mean and identity covariance, and labels were randomly generated binary classification labels. Table 1 presents the average accuracy and variance of each variant across 10 independent runs.

Model Variant	Average Accuracy	Accuracy Variance ($\times 10^{-4}$)
AgentDropout	49.1	6.322
MCMC	49.4	11.378
Bayesian	50.1	6.767
GP	50.1	9.433
GP + MCMC	52.1	6.767
Bayesian + GP	51.9	29.656
Bayesian + MCMC	52.1	19.656
ABO	**58.8**	**4.622**

Table 1: Ablation Study on Core Components: Accuracy (%)

ABO's full setup hits 58.8% accuracy, a 9.7-point gain over the baseline with the least variance. Its individual parts only add 0.3 - 1.0 percentage points, but the whole framework brings major improvements, proving the three core parts work well together.

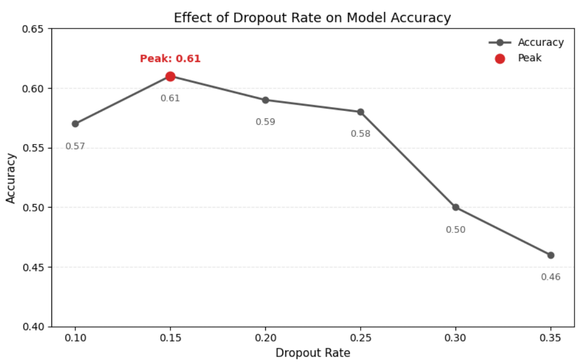

Figure 3: Effect of Dropout Rate on Model Accuracy

Figure 3 shows that the model achieves optimal performance at a dropout rate of 0.15 (61%), which is adopted in subsequent experiments. Adjacency matrix analysis reveals that ABO learns modular graph structures (detailed in Appendix A.3), providing a theoretical foundation for practical applications.

4.3 Benchmark Performance

4.3.1 Comparative Experiment

After validating the theoretical effectiveness of core mechanisms, we further evaluated ABO's practical performance on real benchmark datasets. We selected five widely used benchmark datasets for evaluation: MMLU, HumanEval, GSM8K, AQUA, and SVAMP. All experiments were conducted using three different scales of large language models: DeepSeekV3(DeepSeek-AI et al., 2025), Llama-3-70B-Instruct, and Llama-3-8B-Instruct(et al., 2024), to verify the generalizability of our approach. Table 2 presents detailed performance comparisons of ABO with existing methods across three different scales of LLMs on various benchmark datasets.

Results show that ABO achieves optimal performance across all model scales and datasets, with average accuracy improvements of 1.18%-2.59%. The largest improvements are observed on complex reasoning tasks (such as AQUA), with the most significant enhancement on smaller models (Llama-3-8B).

4.3.2 Ablation Study

Independent ABO analyses show each part boosts model performance: Bayesian modeling improves info reliability assessment, MCMC sampling enhances collaboration via better communication paths. The full ABO framework achieves top performance across models and benchmarks, effectively addressing communication redundancy and improving collaboration efficiency, intelligence, adaptability, and robustness. This confirms the effectiveness of our innovative designs. More case studies are shown in A.4.

4.3.3 Consumption Analysis

Figure 4 illustrates the comparison of token consumption between ABO and AgentDropout on the Llama-3-8B-Instruct model.

Token consumption analysis shows ABO saves 0.8% in completion tokens with a slight prompt phase increase (+1.3%), an acceptable trade-off for its accuracy gains, and has inference time overhead similar to AgentDropout. Communication graph analysis reveals ABO learns a more modular topology with clearer node roles in GSM8K (detailed analysis in A.5). ABO's robustness is verified through key hyperparameter impact analysis.

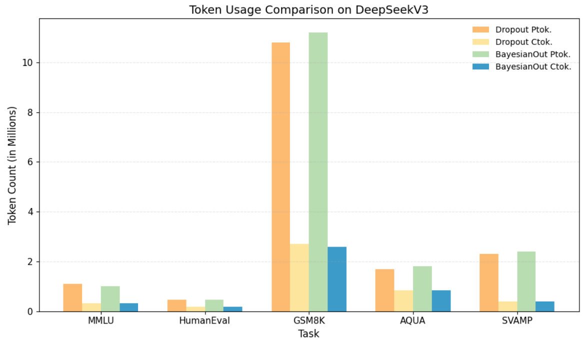

Figure 4: Token Usage Comparison on DeepSeekV3.

4.3.4 Further Analysis

To further explore the applicability of the method, we examined the kernel functions of different estimation processes, varying MCMC sampling steps, and the random dropout rate used by MAS during inference (for detailed experimental information, shown in A.6). Optimizing the choices of kernel function, MCMC sampling steps, and dropout rate can enhance the performance of Agent-BayesianOut. For practical use, an RBF + White kernel combination, 20 - 50 MCMC sampling steps, and a dropout rate of 0.15 or 0.25 are recommended for the best performance balance.

4.4 Open-World Question Answering Evaluation

To assess ABO's effectiveness in open - world scenarios, we experimented on PubMedQA, FinQA, and SafetyQA. These tasks demand coherent text responses. Evaluation used BLEU-4 (measuring n - gram precision match between generated text and reference answers, scored 0 - 100, higher is better)(Papineni et al., 2002), ROUGE-1 (word - level overlap and recall)(Lin and Hovy, 2003), ROUGE-2 (bigram - level matching for phrase similarity), and ROUGE-L (overall structural similarity via longest common subsequences).

Table 3 shows that ABO consistently achieves the highest scores across all evaluation metrics and datasets. Experimental results demonstrate that ABO significantly outperforms baseline methods across all open-ended question-answering datasets

Model	Method	MMLU	HumanEval	GSM8K	AQUA	SVAMP	Average
DeepSeekV3	Vanilla	85.97	88.46	94.72	84.61	93.68	89.49
	CoT	86.31	89.24	95.45	85.46	93.92	90.08
	MAS	88.98	89.52	95.51	85.66	94.18	90.77
	AgentPrune	90.62	90.93	96.01	87.93	95.02	92.10
	AgentDropout	$\underline{90.89}$	$\underline{91.70}$	$\underline{96.17}$	$\underline{88.35}$	$\underline{95.80}$	$\underline{92.58}$
	+MCMC	$90.86\downarrow^{-0.03}$	$91.72\uparrow^{+0.02}$	$96.21\uparrow^{+0.04}$	$88.36\uparrow^{+0.01}$	$95.78\downarrow^{-0.02}$	$92.57\uparrow^{+0.01}$
	+Bayesian	$91.20\uparrow^{+0.31}$	$92.68\uparrow^{+0.98}$	$96.41\uparrow^{+0.24}$	$88.96\uparrow^{+0.61}$	$96.27\uparrow^{+0.47}$	$93.10\uparrow^{+0.52}$
	ABO	**91.86**$\uparrow^{+0.97}$	**93.10**$\uparrow^{+1.40}$	**96.67**$\uparrow^{+0.50}$	**90.21**$\uparrow^{+1.86}$	**96.98**$\uparrow^{+1.18}$	**93.76**$\uparrow^{+1.18}$
Llama-3-70B	Vanilla	83.12	85.76	90.81	82.06	89.89	86.33
	CoT	83.69	86.11	92.09	82.63	90.81	87.07
	MAS	86.37	87.83	93.74	83.81	91.49	88.65
	AgentPrune	87.80	$\underline{88.87}$	93.08	85.40	92.64	89.56
	AgentDropout	$\underline{88.69}$	88.63	$\underline{94.85}$	$\underline{86.30}$	$\underline{93.02}$	$\underline{89.90}$
	+MCMC	$86.72\downarrow^{-1.08}$	$88.71\downarrow^{-0.16}$	$94.73\downarrow^{-0.12}$	$86.52\uparrow^{+0.22}$	$93.29\uparrow^{+0.27}$	$89.99\uparrow^{+0.09}$
	+Bayesian	$88.66\uparrow^{+0.86}$	$89.70\uparrow^{+0.83}$	$95.01\uparrow^{+0.16}$	$86.95\uparrow^{+0.65}$	$94.22\uparrow^{+1.20}$	$90.91\uparrow^{+1.01}$
	ABO	**90.17**$\uparrow^{+2.38}$	**90.16**$\uparrow^{+1.29}$	**95.62**$\uparrow^{+0.77}$	**88.60**$\uparrow^{+2.30}$	**94.08**$\uparrow^{+1.06}$	**91.73**$\uparrow^{+1.83}$
Llama-3-8B	Vanilla	53.56	53.32	70.24	41.66	75.12	58.78
	CoT	56.84	54.18	70.46	43.76	76.26	60.30
	MAS	60.14	48.34	69.32	45.32	77.63	60.15
	AgentPrune	60.82	49.16	$\underline{71.49}$	47.26	78.86	61.52
	AgentDropout	$\underline{62.61}$	$\underline{55.68}$	71.01	$\underline{47.82}$	$\underline{79.24}$	$\underline{63.27}$
	+MCMC	$62.10\downarrow^{-0.51}$	$55.69\uparrow^{+0.01}$	$71.24\downarrow^{-0.25}$	$47.67\downarrow^{-0.15}$	$79.76\uparrow^{+0.52}$	$63.292\uparrow^{+0.02}$
	+Bayesian	$63.26\uparrow^{+0.65}$	$56.88\uparrow^{+1.20}$	$72.60\uparrow^{+1.11}$	$51.24\uparrow^{+3.42}$	$80.72\uparrow^{+1.48}$	$64.94\uparrow^{+1.67}$
	ABO	**63.98**$\uparrow^{+1.37}$	**57.84**$\uparrow^{+2.16}$	**72.78**$\uparrow^{+1.29}$	**52.59**$\uparrow^{+4.77}$	**82.11**$\uparrow^{+2.87}$	**65.86**$\uparrow^{+2.59}$

Table 2: Accuracy Comparison of Different Methods Across LLM Base Models (%)

Dataset	Model	BLEU-4	ROUGE-1	ROUGE-2	ROUGE-L
PubMedQA	Vanilla	42.4	53.1	37.9	51.6
	CoT	45.7	56.9	41.3	55.3
	MAS	48.2	59.8	44.6	58.2
	ABO	**51.9**	**63.4**	**48.7**	**62.4**
FinQA	Vanilla	39.0	49.5	33.2	47.8
	CoT	43.2	54.8	38.9	52.4
	MAS	46.5	58.3	42.2	55.9
	ABO	**49.9**	**61.6**	**46.8**	**59.4**
SafetyQA	Vanilla	47.6	58.2	43.2	56.9
	CoT	50.1	61.8	46.9	59.5
	MAS	52.9	64.4	49.6	62.2
	ABO	**55.6**	**67.9**	**53.2**	**65.4**

Table 3: Performance comparison on different models and datasets

and metrics, showcasing its versatility in handling both deterministic and uncertain problems. In the medical domain, ABO's ROUGE-2 score improves by 4.1% over MAS on PubMedQA (Jin et al., 2019), highlighting its advantage in processing specialized terminology and phrase combinations. ABO also maintains consistent performance advantages in the financial domain, as evidenced by its results on FinQA (Yang et al., 2023; Zhang et al., 2023a,b; Wang et al., 2023; Liu et al., 2023; Zhang et al., 2023c), and in the safety domain on SafetyQA(to be fully publicly available upon acceptance). These results validate ABO's effectiveness in professional domain open-question-answering tasks, providing support for its deployment in diverse application scenarios.

5 Conclusion

We introduce Agent Bayesian Out (ABO), a framework for optimizing multi-agent communication topology with uncertainty awareness. It combines Bayesian edge weights, Gaussian process priors, and MCMC sampling to eliminate low-contribution nodes/edges. ABO surpasses existing methods on benchmarks. Ablation studies highlight core component synergy, and hyperparameter analysis pinpoints an optimal dropout rate. ABO shows more gains on smaller models and complex tasks, proving effective communication optimization in resource-constrained settings. Importantly, it boosts performance without extra inference costs, offering a new paradigm for efficient multi-agent collaboration.

6 Limitations

Despite its outstanding performance in multi-agent communication optimization, ABO still has several limitations that warrant attention. Although the inference stage has comparable overhead to existing methods, the MCMC sampling and variational inference in the training stage may impose significant computational burdens in large-scale agent systems. The current framework applies the same dropout rate to all agents, failing to fully account for the role differences and importance of different agents. Moreover, our experiments have mainly focused on language understanding, code generation, and mathematical reasoning, and the effectiveness of ABO in more extensive scenarios such as multimodal interaction or real-time decision-making remains to be verified. The uncertainty modeling based on Gaussian distribution may not be precise enough in some complex nonlinear or multimodal distribution scenarios. Additionally, in the face of adversarial scenarios or highly dynamic environments with cognitive inequality and incomplete information, the existing methods still need improvement. Addressing these limitations will help develop more efficient and adaptive multi-agent communication optimization frameworks.

References

Manlu Ban, Jianjun Li, Hongyong Deng, and Yanhong Wang. 2025. Dynamic event-triggered h leader-following consensus for nonlinear stochastic multi-agent systems with semi-markov switching topologies and application. *Journal of the Franklin Institute*, 362(1287):107596.

Spencer C Castro, P. Samuel Quinan, Helia Hosseinpour, and Lace Padilla. 2022. Examining effort in 1d uncertainty communication using individual differences in working memory and nasa-tlx. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):411–421. Epub 2021 Dec 29.

Junzhe Chen, Xuming Hu, Shuodi Liu, Shiyu Huang, Wei-Wei Tu, Zhaofeng He, and Lijie Wen. 2024. Llmarena: Assessing capabilities of large language models in dynamic multi-agent environments. *arXiv preprint arXiv:2402.16499*.

Qitian Chen and Shaoyuan Li. 2024. Strong structural controllability of multi-agent systems with uncertain connection topologies. *Journal of Systems Science & Complexity*. Received: 17 Oct 2023 / Revised: 12 Nov 2024.

Yuheng Cheng, Ceyao Zhang, Zhengwen Zhang, Xiangrui Meng, Sirui Hong, Wenhao Li, Zihao Wang, Zekai Wang, Feng Yin, Junhua Zhao, and Xiuqiang He. 2024. Exploring large language model based intelligent agents: Definitions, methods, and prospects. *arXiv preprint arXiv:2401.03428*. Submitted on 7 Jan 2024.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. 2021. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*. ArXiv:2110.14168v2 [cs.LG] for this version.

DeepSeek-AI, Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, Damai Dai, Daya Guo, Dejian Yang, Deli Chen, Dongjie Ji, Erhang Li, Fangyun Lin, Fucong Dai, and 81 others. 2025. Deepseek-v3 technical report. arXiv preprint arXiv:2412.19437v2, arXiv.

Rafael Dewes and Rayna Dimitrova. 2024. Contract-based design and verification of multi-agent systems with quantitative temporal requirements. *arXiv preprint arXiv:2412.13114*. Extended version of paper accepted at AAAI-25.

Anqi Dong, Karl H. Johansson, and Johan Karlsson. 2025. Task allocation for multi-agent systems via unequal-dimensional optimal transport. *arXiv preprint arXiv:2503.09369*. Submitted on 12 Mar 2025, 6 pages, 4 figures.

Qiqi Duan, Chang Shao, Guochen Zhou, Minghan Zhang, Qi Zhao, and Yuhui Shi. 2024. Distributed evolution strategies with multi-level learning for large-scale black-box optimization. *IEEE Transactions on Parallel and Distributed Systems*, 35(11):2087–2101.

Will Epperson, Gagan Bansal, Victor Dibia, Adam Fourney, Jack Gerrits, Erkang Zhu, and Saleema Amershi. 2025. Interactive debugging and steering of multi-agent ai systems. In *CHI Conference on Human Factors in Computing Systems (CHI 2025)*. Published at CHI 2025.

Aaron Grattafiori et al. 2024. The llama 3 herd of models. arXiv preprint arXiv:2407.21783v3, arXiv.

Chen Gao, Xiaochong Lan, Nian Li, Yuan Yuan, Jingtao Ding, Zhilun Zhou, Fengli Xu, and Yong Li. 2023a. Large language models empowered agent-based modeling and simulation: A survey and perspectives. *arXiv preprint arXiv:2312.11970*.

Chen Gao, Xiaochong Lan, Nian Li, Yuan Yuan, Jingtao Ding, Zhilun Zhou, Fengli Xu, and Yong Li. 2023b. Large language models empowered agent-based modeling and simulation: A survey and perspectives. *arXiv preprint arXiv:2312.11970*. Submitted on 19 Dec 2023.

Iago Pachêco Gomes, Cristiano Premebida, and Denis Fernando Wolf. 2024. Multi-agent interaction-aware behavior intention prediction using graph mixture of experts attention network on urban roads. *Expert Systems with Applications*. Preprint submitted October 22, 2024.

Xiangming Gu, Xiaosen Zheng, Tianyu Pang, Chao Du, Qian Liu, Ye Wang, Jing Jiang, and Min Lin. 2024. Agent smith: A single image can jailbreak one million multimodal LLM agents exponentially fast. In *Forty-first International Conference on Machine Learning*.

Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. 2020. Measuring massive multitask language understanding. *arXiv preprint arXiv:2009.03300*.

Wenyue Hua, Xianjun Yang, Mingyu Jin, Zelong Li, Wei Cheng, Ruixiang Tang, and Yongfeng Zhang. 2024. Trustagent: Towards safe and trustworthy llm-based agents. *arXiv preprint arXiv:2402.01586*. Submitted on 2 Feb 2024 (v1), last revised 3 Oct 2024 (v4).

Amirul Islam, Leila Musavian, and Nikolaos Thomos. 2022. Multi-agent deep reinforcement learning in vehicular occ. In *IEEE Vehicular Technology Conference (VTC2022-Spring)*. Accepted in VTC2022-Spring.

Qiao Jin, Bhuwan Dhingra, Zhengping Liu, William Cohen, and Xinghua Lu. 2019. Pubmedqa: A dataset for biomedical research question answering. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2567–2577, Hong Kong, China. Association for Computational Linguistics.

Donghyun Lee and Mo Tiwari. 2024. Prompt infection: Llm-to-llm prompt injection within multi-agent systems. *arXiv preprint arXiv:2410.0.07283*. Submitted on 9 Oct 2024 (v1), last revised 11 Oct 2024 (v4).

Chin-Yew Lin and Eduard Hovy. 2003. Automatic evaluation of summaries using n-gram co-occurrence statistics. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, volume N03-1020, pages 150–157, NAACL. Association for Computational Linguistics. DOI: [insert DOI here].

Xiao-Yang Liu, Guoxuan Wang, Hongyang Yang, and Daochen Zha. 2023. Fingpt: Democratizing internet-scale data for financial large language models. arXiv preprint arXiv:2307.10485. 43 pages, 8 tables, and 2 figures.

Hangyu Mao, Zhengchao Zhang, Zhen Xiao, Zhibo Gong, and Yan Ni. 2020. Learning agent communication under limited bandwidth by message pruning. In *AAAI Conference on Artificial Intelligence (AAAI 2020)*. Accepted as a regular paper with poster presentation at AAAI 2020.

Joshua McClellan, Naveed Haghani, John Winder, Furong Huang, and Pratap Tokekar. 2024. Boosting sample efficiency and generalization in multi-agent reinforcement learning via equivariance. In *NeurIPS 2024*. Accepted as a poster at NeurIPS 2024.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, volume P02, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics. NAACL 2018 Test-of-Time Award.

Arkil Patel, Satwik Bhattamishra, and Navin Goyal. 2021. Are nlp models really able to solve simple math word problems? *arXiv preprint arXiv:2103.07191*. ArXiv:2103.07191v2 [cs.CL] for this version; presented at NAACL 2021.

Pierre Peigne-Lefebvre, Mikolaj Kniejski, Filip Sondej, Matthieu David, Jason Hoelscher-Obermaier, Christian Schroeder de Witt, and Esben Kran. 2025. Multi-agent security tax: Trading off security and collaboration capabilities in multi-agent systems. In *AAAI Conference on Artificial Intelligence (AAAI 2025)*. Accepted to AAAI 2025 Conference.

Qiwei Peng, Yekun Chai, and Xuhong Li. 2024. Humaneval-xl: A multilingual code generation benchmark for cross-lingual natural language generalization. *arXiv preprint arXiv:2402.16694*. To appear in LREC-COLING 2024.

Pouya Pezeshkpour, Eser Kandogan, Nikita Bhutani, Sajjadur Rahman, Tom Mitchell, and Estevam Hruschka. 2024. Reasoning capacity in multi-agent systems: Limitations, challenges and human-centered solutions. *arXiv preprint arXiv:2402.01108*. ArXiv.org perpetual non-exclusive license.

Yangjun Ruan, Honghua Dong, Andrew Wang, Silviu Pitis, Yongchao Zhou, Jimmy Ba, Yann Dubois, Chris J. Maddison, and Tatsunori Hashimoto. 2024. Identifying the risks of lm agents with an lm-emulated sandbox. *arXiv preprint arXiv:2309.15817*. Submitted on 25 Sep 2023 (v1), last revised 17 May 2024 (v2).

Manish Sanwal. 2025. Layered chain-of-thought prompting for multi-agent llm systems: A comprehensive approach to explainable large language models. *arXiv preprint arXiv:2501.18645*. Submitted on 29 Jan 2025 (v1), last revised 3 Feb 2025 (v2).

Soumyendu Sarkar, Vineet Gundecha, Alexander Shmakov, Sahand Ghorbanpour, Ashwin Ramesh Babu, and 1 others. 2022. Multi-agent reinforcement learning controller to maximize energy efficiency for multi-generator industrial wave energy converter. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 12135–12144, June 2022. AAAI Press.

Qihe Shan, Fei Teng, Tieshan Li, and C.L. Philip Chen. 2020a. Containment control of multi-agent systems with nonvanishing disturbance via topology reconfiguration. *Science China Information Sciences*, 64(7):179203:1–179203:3.

Qihe Shan, Fei Teng, Tieshan Li, and C.L. Philip Chen. 2020b. Containment control of multi-agent systems with nonvanishing disturbance via topology reconfiguration. *Science China Information Sciences*, 64(7):179203:1–179203:3.

Min Yee Teh, Shizhen Zhao, Peirui Cao, and Keren Bergman. 2022. Enabling quasi-static reconfigurable networks with robust topology engineering. *IEEE/ACM Transactions on Networking*, PP(99):1–15.

Cong Tran, Won-Yong Shin, and Andreas Spitz. 2021. Community detection in partially observable social networks. *ACM Transactions on Knowledge Discovery from Data*. 24 pages, 8 figures, 5 tables; to appear in an upcoming issue.

Khanh-Tung Tran, Dung Dao, Minh-Duong Nguyen, Quoc-Viet Pham, Barry O'Sullivan, and Hoang D. Nguyen. 2025. Multi-agent collaboration mechanisms: A survey of llms. *arXiv preprint arXiv:2501.06322*. Submitted on 10 Jan 2025.

Sharlin Utke, Jeremie Houssineau, and Giovanni Montana. 2025. Investigating relational state abstraction in collaborative marl. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 20947–20955, Philadelphia, Pennsylvania, USA. AAAI.

Ling Wang, Dani Yogatama, Chris Dyer, and Phil Blunsom. 2017. Program induction by rationale generation: Learning to solve and explain algebraic word problems. *arXiv preprint arXiv:1705.04146*. ArXiv:1705.04146v3 [cs.AI] for this version.

Neng Wang, Hongyang Yang, and Christina Dan Wang. 2023. Fingpt: Instruction tuning benchmark for open-source large language models in financial datasets. arXiv preprint arXiv:2310.04793. Workshop on Instruction Tuning and Instruction Following at NeurIPS 2023.

Zhexuan Wang, Yutong Wang, Xuebo Liu, Liang Ding, Miao Zhang, Jie Liu, and Min Zhang. 2025. Agentdropout: Dynamic agent elimination for token-efficient and high-performance llm-based multi-agent collaboration. *arXiv preprint arXiv:2503.18891*. Submitted on 24 Mar 2025.

Di Wu, Xian Wei, Guang Chen, Hao Shen, Xiangfeng Wang, Wenhao Li, and Bo Jin. 2025. Generative multi-agent collaboration in embodied ai: A systematic review. *arXiv preprint arXiv:2502.11518*. 18 pages.

Hongyang Yang, Xiao-Yang Liu, and Christina Dan Wang. 2023. Fingpt: Open-source financial large language models. arXiv preprint arXiv:2306.06031.

Yijun Yang, Tianyi Zhou, Kanxue Li, Dapeng Tao, Lusong Li, Li Shen, Xiaodong He, Jing Jiang, and Yuhui Shi. 2024a. Embodied multi-modal agent trained by an llm from a parallel textworld. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 26265–26275.

Yingxuan Yang, Qiuying Peng, Jun Wang, Ying Wen, and Weinan Zhang. 2024b. Llm-based multi-agent systems: Techniques and business perspectives. *arXiv preprint arXiv:2411.14033*. Submitted on 21 Nov 2024 (v1), last revised 28 Dec 2024 (this version, v2).

John Yen, Jianwen Yin, Thomas R. Ioerger, Michael S. Miller, Dianxiang Xu, and Richard A. Volz. CAST: Collaborative Agents for Simulating Teamwork. Department of Computer Science, Texas A&M University.

Jiawei You, Youlong Wu, Dingzhu Wen, Yong Zhou, Yuning Jiang, and Yuanming Shi. 2024. Dynamic communication in multi-agent reinforcement learning via information bottleneck. In *GLOBECOM 2024 - 2024 IEEE Global Communications Conference*, Shanghai, China. IEEE. To appear.

Yangyang Yu, Zhiyuan Yao, Haohang Li, Zhiyang Deng, Yupeng Cao, Zhi Chen, Jordan W. Suchow, Rong Liu, Zhenyu Cui, Zhaozhuo Xu, Denghui Zhang, Koduvayur Subbalakshmi, Guojun Xiong, Yueru He, Jimin Huang, Dong Li, and Qianqian Xie. 2024. Fincon: A synthesized llm multi-agent system with conceptual verbal reinforcement for enhanced financial decision making. *arXiv preprint arXiv:2407.06567*. Submitted on 9 Jul 2024 (v1), last revised 7 Nov 2024 (v3).

Boyu Zhang, Hongyang Yang, and Xiao-Yang Liu. 2023a. Instruct-fingpt: Financial sentiment analysis by instruction tuning of general-purpose large language models. arXiv preprint arXiv:2306.12659. FinLLM Symposium at IJCAI 2023.

Boyu Zhang, Hongyang Yang, Tianyu Zhou, Ali Babar, and Xiao-Yang Liu. 2023b. Enhancing financial sentiment analysis via retrieval augmented large language models. arXiv preprint arXiv:2310.04027. ACM International Conference on AI in Finance (ICAIF) 2023.

Boyu Zhang, Hongyang Yang, Tianyu Zhou, Ali Babar, and Xiao-Yang Liu. 2023c. Enhancing financial sentiment analysis via retrieval augmented large language models. *arXiv preprint arXiv:2310.04027*. Presented at ACM International Conference on AI in Finance (ICAIF) 2023.

Cuijuan Zhang, Lianghao Ji, Shasha Yang, Xing Guo, and . 2024a. Distributed optimal consensus control for multi-agent systems based on event-triggered and prioritized experience replay strategies. *Science China Information Sciences*, 68(1).

Guibin Zhang, Yanwei Yue, Zhixun Li, Sukwon Yun, Guancheng Wan, Kun Wang, Dawei Cheng, Jeffrey Xu Yu, and Tianlong Chen. 2024b. Cut the crap: An economical communication pipeline for llm-based multi-agent systems. *arXiv preprint arXiv:2410.02506*. Submitted on 3 Oct 2024.

Zhuohui Zhang, Bin He, Bin Cheng, and Gang Li. 2025. Bridging training and execution via dynamic directed graph-based communication in cooperative multi-agent systems. In *The 39th AAAI Conference on Artificial Intelligence (AAAI 2025)*. 9 pages, 7 figures.

Yihe Zhou, Yuxuan Zheng, Yue Hu, Kaixuan Chen, Tongya Zheng, Jie Song, Mingli Song, and Shunyu Liu. 2025. Cooperative policy agreement: Learning diverse policy for offline marl. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 23018–23026, Philadelphia, Pennsylvania, USA.

A Appendix

A.1 Graph Sampling for inference

The inference stage of our proposed Adaptive Bayesian Optimization (ABO) framework requires efficient sampling of communication graph topologies from the learned probabilistic model. This appendix details the implementation of the sampling module used during inference.

A.1.1 Probabilistic Graph Sampling

After training, our model has learned parameters for the adjacency matrix, represented as distributions with mean $\mu_{ij}^{(t)}$ and variance $(\sigma_{ij}^{(t)})^2$ for each potential edge $e_{ij}^{(t)}$. Algorithm 1 illustrates our complete sampling procedure, which follows these key steps:

1. **Graph Sampling**: For each potential edge, we sample from the learned normal distribution $\mathcal{N}(\mu_{ij}^{(t)}, (\sigma_{ij}^{(t)})^2)$ and compare with a threshold to determine edge existence. The algorithm generates multiple graph samples to capture the uncertainty in the learned model.

2. **MCMC-Adjustment**: As all valid communication graphs must be directed acyclic graphs (DAGs), we employ a verification step after initial sampling. When cycles are detected, our MCMC adjustment algorithm (Algorithm 1, Phase 2) resolves constraint violations by selectively removing edges with the lowest probability until the DAG property is satisfied.

3. **Inference with Sampled Graphs**: The sampling process balances between exploration of the graph space and exploitation of high-probability edges. The variance term $(\sigma_{ij}^{(t)})^2$ naturally captures the uncertainty in edge existence, allowing more diverse sampling for edges with high uncertainty.

A.1.2 Implementation Considerations

When implementing the sampling module, several practical considerations are addressed:

- **Computational Efficiency**: For large-scale multi-agent systems, efficient cycle detection and DAG verification algorithms are crucial. We implement an optimized topological sorting algorithm with $O(V+E)$ complexity as part of our MCMC-Adjustment function.

- **Threshold Selection**: The threshold for edge existence can be adaptively set based on the distribution of sampled probabilities or fixed at a predetermined value (e.g., 0.5). Similarly, the consensus threshold for the final graph integration is carefully selected.

- **Sample Diversity**: To ensure diverse graph samples, the sampling process leverages the probability distributions learned during training, particularly the variance parameters that indicate model uncertainty.

A.1.3 Inference with Sampled Graphs

When multiple graph samples are generated ($N_s > 1$), our algorithm employs a consensus-based approach to aggregate results:

- The algorithm computes the edge occurrence frequency f_{ij} across all sampled graphs.

- Edges appearing in a majority of samples (above the consensus threshold) are included in the final graph.

- The final consensus graph is verified to maintain the DAG property through the same MCMC-Adjustment process if needed.

- For classification tasks, this consensus approach effectively implements majority voting, while for regression tasks, it provides a stable foundation for prediction averaging.

The complete graph sampling algorithm (Algorithm 1) enables our framework to leverage the full probabilistic information learned during training, resulting in more robust and adaptable multi-agent communication structures during deployment.

A.2 Experimental Details

A.2.1 Evaluation Protocol

- **Benchmark Datasets**: MMLU (general reasoning), HumanEval (code generation), GSM8K, AQUA, and SVAMP (mathematical reasoning)

- **Comparison Methods**: Vanilla (baseline), CoT (Chain of Thought), MAS (standard multi-agent system), AgentPrune, and Agent-Dropout

- **Consistency Measures**: Same agent configuration and prompts across all experiments, with agent prompts inherited from Agent-Dropout; consistent random seeds

A.2.2 Implementation Details

- **Hardware**: All experiments conducted on 2× RTX 3090 GPUs

- **ABO Parameters**:

  - Node deactivation rate: $\alpha_N = 0.15$
  - Edge deactivation rate: $\beta_E = 0.15$
  - MCMC sampling times: $M = 10$
  - Initial learning rate: $\eta = 0.001$
  - Training: 20 epochs using Adam optimizer
  - Bayesian edge weight prior distribution: $\mathcal{N}(0.5, 1.0)$
  - Weight coefficients: $\lambda_1 = \lambda_2 = 0.01$, $\lambda_3 = 0.005$, and $\lambda_4 = 0.1$

A.3 Adjacency Matrix Structure Analysis

We analyzed the communication structure learned by ABO through adjacency matrix heatmaps. As shown in Figure 5, the trained adjacency matrix exhibits distinct modular characteristics compared to the initial state, with more concentrated edge weight distributions forming tightly connected substructures. In the initial state, the communication structure exhibits a random connection pattern: Node 0 connects to Nodes 2-4, Node 1 connects to Node 0, and Node 2 connects to Nodes 0-1. After training, the ABO-optimized topology shows a selective connection pattern, particularly with Node 3 forming strong connections with Nodes 1, 2, and 4, and Node 2 forming bidirectional connections with Node 3, while other regions remain sparse. This modular structure implies more efficient information transfer and specialized division of labor in multi-agent systems.

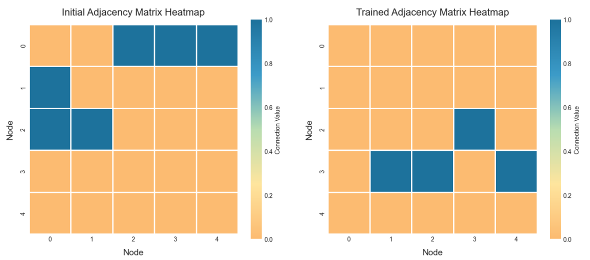

Figure 5: Initial adjacency matrix (left) vs. trained adjacency matrix (right). Blue indicates strong connections, yellow indicates weak or no connections. Node 3 emerges as a key information hub after ABO optimization.

A.4 Case Study

- **General Reasoning (MMLU)**: ABO outperforms AgentDropout by +0.97%, +3.48%, +1.37% across three model scales. It enhances cross-domain reasoning via Bayesian uncertainty quantification and dynamic communication topology adaptation.

- **Code Generation (HumanEval)**: ABO achieves +1.4%, +1.53%, +2.16% gains in programming tasks. It resolves algorithmic uncertainty through probabilistic code pattern evaluation and MCMC-driven knowledge integration.

- **Mathematical Reasoning (AQUA/GSM8K/SVAMP)**: ABO shows +4.77% gains on AQUA, highlighting its multi-step problem-solving ability. Bayesian agents verify solution reliability through uncertainty propagation and MCMC-based collaborative refinement.

- **Scaling Characteristics**: ABO exhibits +2.588%, +1.828%, +1.182% improvements from 8B to 70B models. Its uncertainty-aware mechanism optimizes knowledge routing, reducing error accumulation in smaller models.

A.5 Communication Graph Structure Analysis

To intuitively demonstrate how AgentBayesianOut optimizes communication topology, we analyzed the communication graph structure learned on the GSM8K task, as shown in Figure 6.

Comparing the edge distribution and node in-degree/out-degree changes between the initial state (top) and trained state (bottom), we can observe that ABO forms more meaningful communication

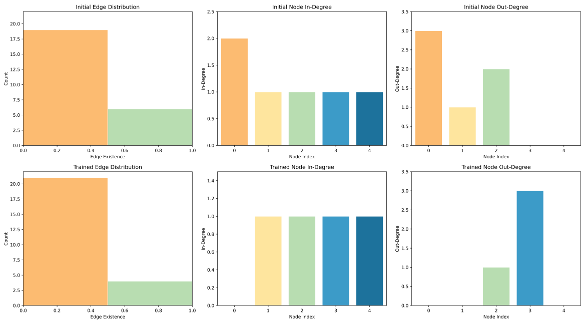

Figure 6: Initial communication graph structure (top) vs. trained structure (bottom). Left: edge existence distribution; Center: node in-degree; Right: node out-degree. Node 3 emerges as the primary information output node after training, while connections are redistributed more efficiently.

patterns. In the initial graph, Node 0 exhibits high in-degree and out-degree, indicating imbalanced communication load. After training, the system retains key bidirectional connections between mathematical experts and problem decomposers while reducing direct connections from commentators to executors, instead using mathematical experts as information intermediaries to form clearer information flow paths. Particularly noteworthy is

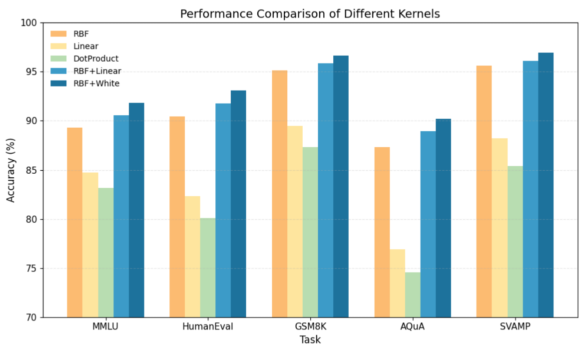

Figure 7: Performance Comparison of Different Kernels across Tasks. The RBF+White kernel shows the highest average accuracy, highlighted in sky blue.

that Node 3 becomes the primary information output node after training, while Node 0's out-degree is completely eliminated, and the roles of Nodes 1 and 2 also undergo significant changes. This demonstrates that ABO can intelligently identify and retain the critical participation of specific nodes (such as result verifiers) rather than mechanically applying uniform deactivation patterns. This case clearly illustrates how ABO achieves more flexible and efficient communication topology optimization.

A.6 Detailed Further Analysis

Figure 7 shows that the RBF+White kernel combination performs optimally (93.76%), outperforming single RBF kernels (91.57%) and other choices such as linear kernels.

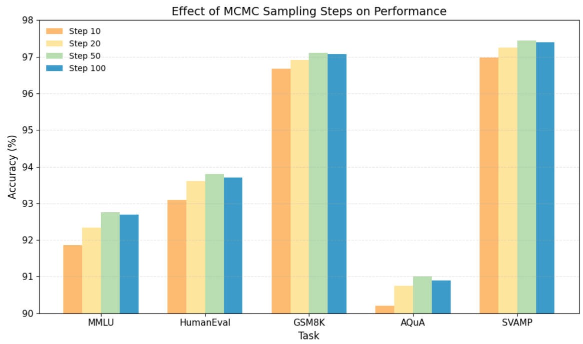

Figure 8: Effect of MCMC Sampling Steps on Model Performance across Tasks. Increasing the number of MCMC steps improves accuracy up to 50 steps, after which performance saturates or slightly declines.

Figure 8 demonstrates that model performance reaches optimum between 20-50 MCMC sampling steps, with performance stabilizing or slightly declining after 50 steps.

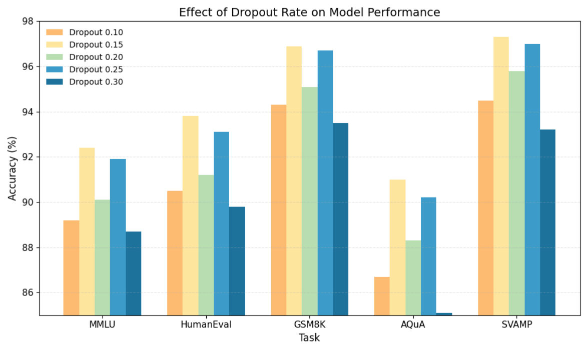

Figure 9: Effect of Dropout Rate on Model Performance across Tasks. The performance peaks at a dropout rate of 0.15, after which over-regularization degrades accuracy.

Dropout rate analysis presents a bimodal distribution, as shown in Figure 9. Performance reaches two peaks at 0.15 92.68% and 0.25 93.78%, corresponding to different optimal topological structures. Performance significantly decreases beyond 0.30.

Algorithm 1: Communication Graph Sampling for Inference

Data: Trained parameters $\tilde{A}_{\text{intra}}$ and $\tilde{A}_{\text{inter}}$ (mean $\mu_{ij}^{(t)}$ and variance $(\sigma_{ij}^{(t)})^2$), number of samples N_s
Result: Set of sampled communication graph(s) $\{G_1, G_2, ..., G_{N_s}\}$

Phase 1: Graph Sampling;
Initialize empty set of graph samples $\mathcal{G} \leftarrow \emptyset$;
for $s = 1$ **to** N_s **do**
  Initialize empty adjacency matrix $A^{(s)} \leftarrow \mathbf{0}_{n \times n}$;
  for *each time step* t **do**
    for *each potential edge* $e_{ij}^{(t)}$ **do**
      Sample edge probability $p_{ij}^{(t)} \sim \mathcal{N}(\mu_{ij}^{(t)}, (\sigma_{ij}^{(t)})^2)$;
      if $p_{ij}^{(t)} >$ *threshold* **then**
        Set $A_{ij}^{(s)} \leftarrow 1$; `/* Tentatively add edge */`
      end
    end
    Check if current $A^{(s)}$ maintains DAG property;
    if $A^{(s)}$ *violates DAG constraint* **then**
      $A^{(s)} \leftarrow$ MCMC-Adjustment$(A^{(s)}, \{p_{ij}^{(t)}\})$; `/* Fix DAG constraint */`
    end
  end
  Add graph to sample set: $\mathcal{G} \leftarrow \mathcal{G} \cup \{A^{(s)}\}$;
end
Phase 2: MCMC-Adjustment;
Function `MCMC-Adjustment(`A, $P = \{p_{ij}^{(t)}\}$`)`**:**
  Initialize $A' \leftarrow A$;
  while A' *violates DAG constraint* **do**
    Identify set of edges $E_{\text{violation}}$ that cause cycles;
    Select edge $e_{ij}^{(t)} \in E_{\text{violation}}$ with lowest probability $p_{ij}^{(t)}$;
    Remove edge: $A'_{ij} \leftarrow 0$; `/* Remove least probable edge in cycle */`
    if *DAG check(*A'*) = TRUE* **then**
      break;
    end
  end
  return A';
Phase 3: Inference with Sampled Graph;
if $N_s = 1$ **then**
  Use the single sampled graph G_1 directly for inference;
else
  Perform MCMC-based graph integration;
  Initialize final graph $G_{\text{final}} \leftarrow$ empty graph;
  for *each edge position* (i, j) *across time steps* **do**
    Compute edge occurrence frequency $f_{ij} = \frac{1}{N_s} \sum_{s=1}^{N_s} \mathbb{I}[A_{ij}^{(s)} = 1]$;
    if $f_{ij} >$ *consensus_threshold* **then**
      $G_{\text{final}}(i, j) \leftarrow 1$;
    end
  end
  Ensure G_{final} maintains DAG property using MCMC-Adjustment if needed;
  Use G_{final} for inference;
end
return $\mathcal{G}$;