清華大學
Tsinghua University

Scaling Large Language Model-Based Multi-Agent Collaboration

Qian C, Xie Z, et al.
Tsinghua University & Peng Cheng Laboratory

Contents

Introduction

LLM-Powered Multi-Agent Collaboration

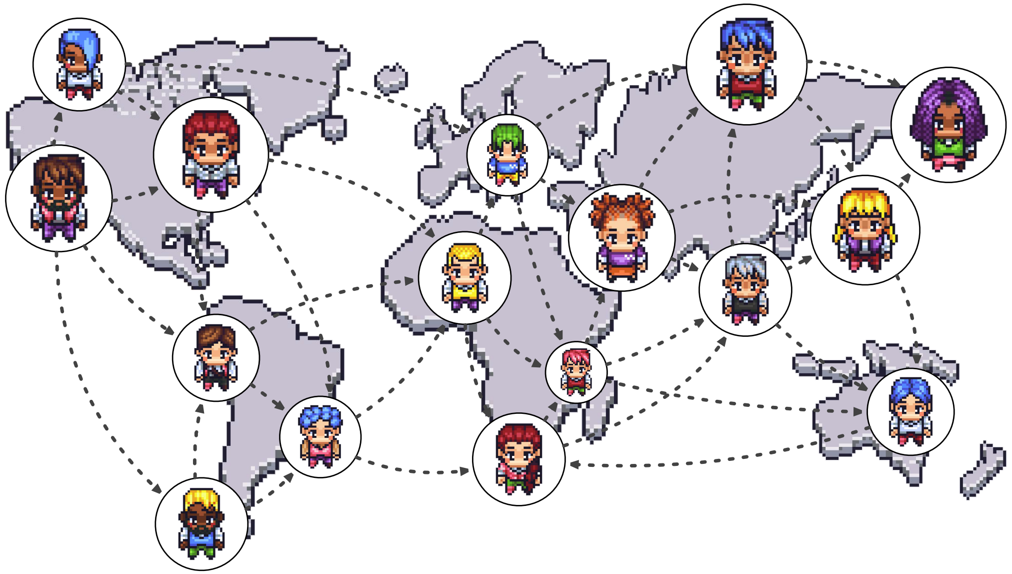
- **Single Agent**: Context Explosion & Lack of Diversity ☹
- **Multi-Agent Collaboration:** facilitated by interactive behaviors, often surpasses standalone intelligence

+

- **Emergent Abilities:** ① Nature: sardine schools, ant swarms ② large language models

=

- **Scaling** the number of agents in Multi-Agent Systems (MAS) -> Emergent behaviors? Collaborative Scaling Law?



Multi-Agent Collaboration

Qian C, Xie Z, et al. Scaling Large Language Model-Based Multi-Agent Collaboration. In International Conference on Learning Representations.

Introduction

LLM-Powered Multi-Agent Collaboration

- **Single Agent**: Context Explosion & Enclosed Reasoning ☹
- **Multi-Agent Collaboration:** facilitated by interactive behaviors, often surpasses standalone intelligence

＋

- **Emergent Abilities:** ① Nature: sardine schools, ant swarms ② large language models

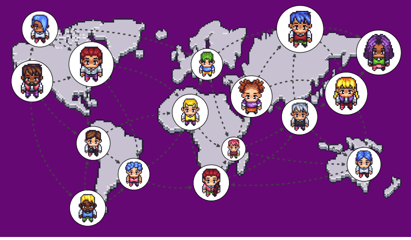
＝

- **Scaling** the number of agents in Multi-Agent Systems (MAS) -> Emergent behaviors? Collaborative Scaling Law?

Emergent behavior of sardine schools

Qian C, Xie Z, et al. Scaling Large Language Model-Based Multi-Agent Collaboration. In International Conference on Learning Representations.

Introduction



LLM-Powered Multi-Agent Collaboration

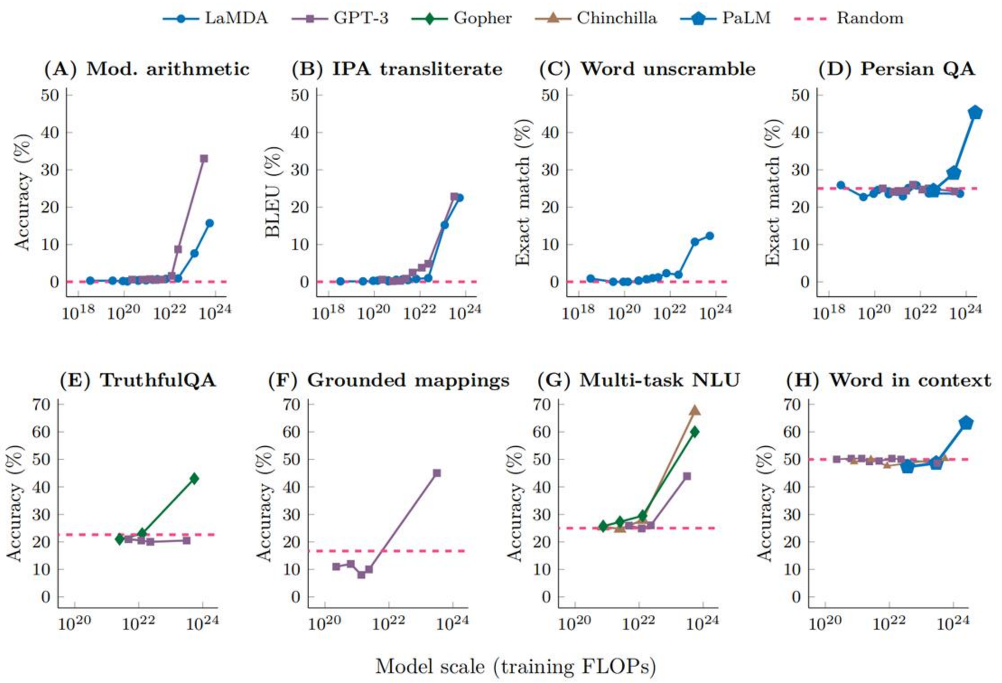
- **Single Agent**: Context Explosion & Enclosed Reasoning ☹
- **Multi-Agent Collaboration:** facilitated by interactive behaviors, often surpasses standalone intelligence

+

- **Emergent Abilities:** ① Nature: sardine schools, ant swarms ② large language models

=

- **Scaling** the number of agents in Multi-Agent Systems (MAS) -> Emergent behaviors? Collaborative Scaling Law?



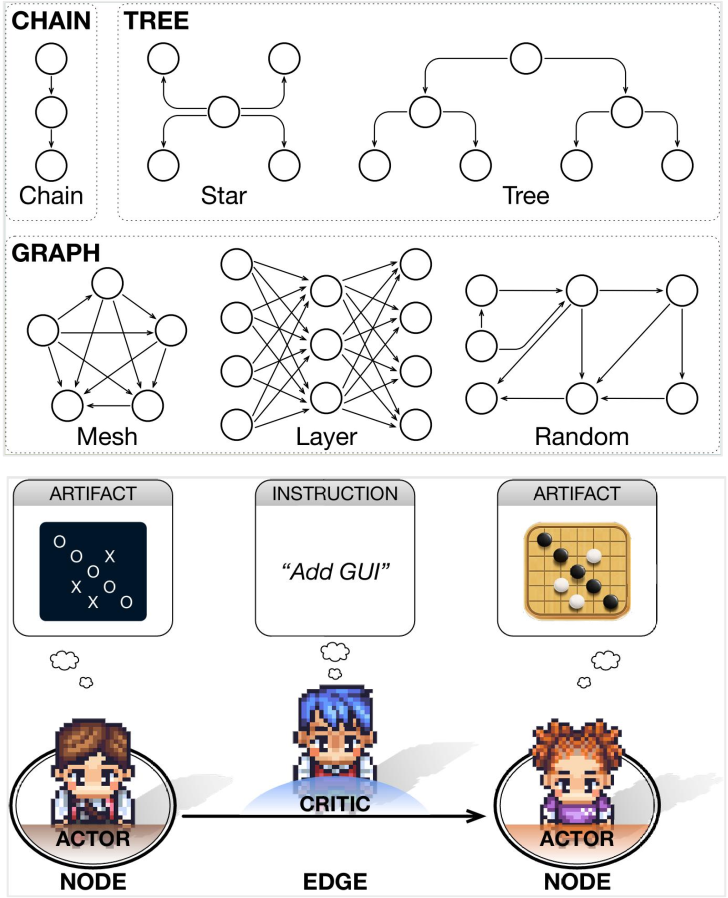
Emergent abilities of large language models

Schaeffer R., Miranda B., Koyejo S. (2023). Are Emergent Abilities of Large Language Models a Mirage? arXiv preprint arXiv:2304.15004.

Methodology

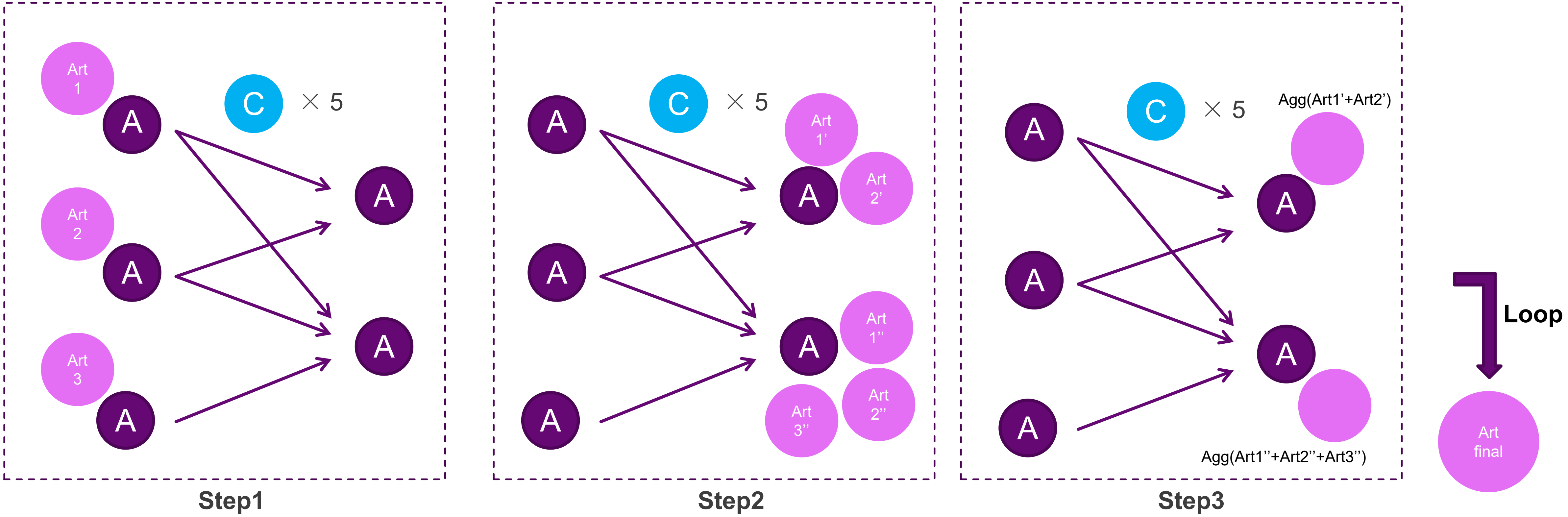
Network Construction

- Describe **agents** and their **interrelations** as a **directed acyclic graph (DAG)** - prevents backflow and enhances generalizability
- Agentization: structurally assign a **critic** to each edge and an **actor** to each node
- The actor's outputs (artifacts) are **refined** under the critic's instructions, eventually **aggregated** into one artifact - information propagation



Qian C, Xie Z, et al. Scaling Large Language Model-Based Multi-Agent Collaboration. In International Conference on Learning Representations.

Methodology



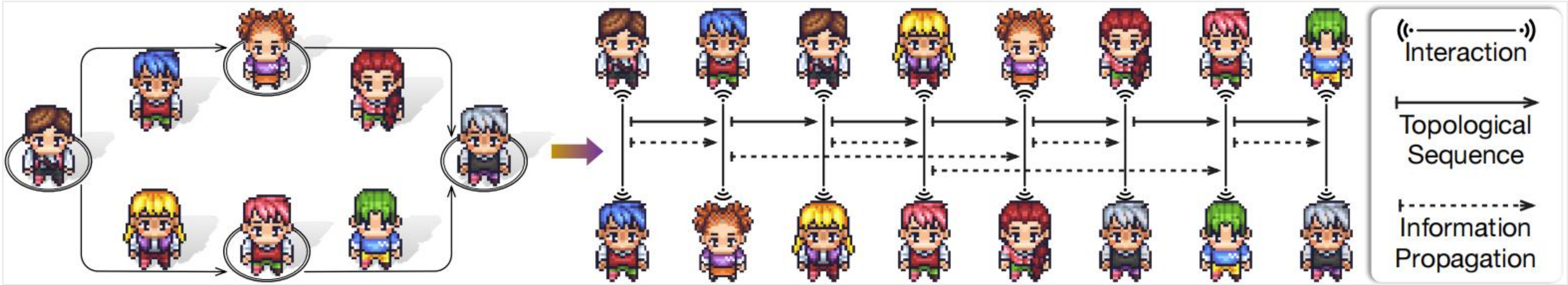
Qian C, Xie Z, et al. Scaling Large Language Model-Based Multi-Agent Collaboration. In International Conference on Learning Representations.

Methodology

Interactive Reasoning

- Traversal strategy: **topological ordering**
  - Orderly information transmission
  - High scalability & generalizability
- The interaction pattern: multi-turn **instruction-response** sequence -> **mitigates hallucinations**

- Memory control: **long-term** (global) & **short-term** (interaction-level) memory decouples context length from **quadratic** to **linear** growth



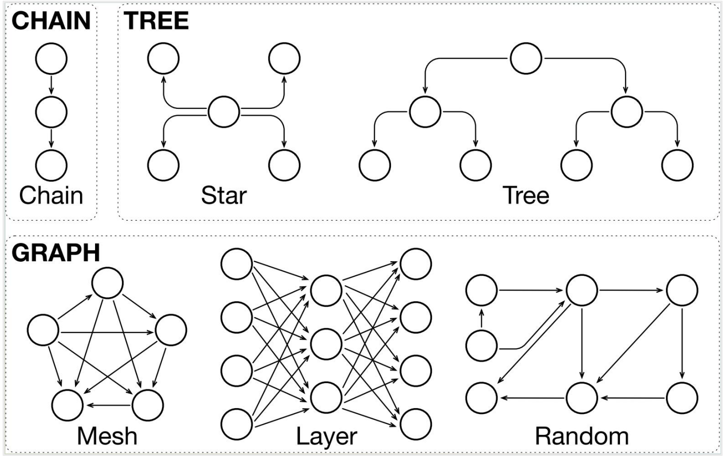
Qian C, Xie Z, et al. Scaling Large Language Model-Based Multi-Agent Collaboration. In International Conference on Learning Representations.

Evaluation

Primary Results

- **Significant improvement**
  - **Chain-structure** method consistently outperforms most baseline models
  - **Random-structure** method achieves the highest performance
- **Ablation study**
  - Setting temperature to **0.0** and ablating profile causes an average performance degradation of **3.67%** across all topologies

Method	Paradigm	MMLU	HumanEval	SRDD	CommonGen	Quality
CoT		0.3544†	<u>0.6098</u>†	0.7222†	0.6165†	0.5757†
AutoGPT		0.4485†	0.4809†	0.7353†	0.5972	0.5655†
GPTSwarm		0.2368†	0.4969†	0.7096†	0.6222†	0.5163†
AgentVerse		0.2977†	**0.7256**†	0.7587†	0.5399†	0.5805
MacNet-Chain		0.6632	0.3720	**0.8056**	0.5903	0.6078
MacNet-Star		0.4456†	0.5549†	0.7679†	<u>0.7382</u>†	0.6267
MacNet-Tree		0.3421†	0.4878†	0.8044	**0.7718**†	0.6015
MacNet-Mesh		<u>0.6825</u>	0.5122†	0.7792†	0.5525†	<u>0.6316</u>†
MacNet-Layer		0.2780†	0.4939†	0.7623†	0.7176†	0.5629†
MacNet-Random		**0.6877**	0.5244†	<u>0.8054</u>	0.5912	**0.6522**†



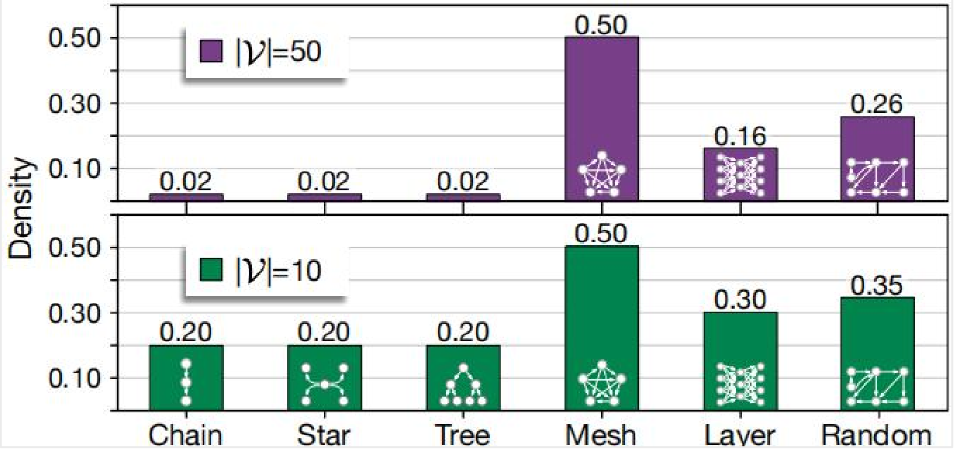
Qian C, Xie Z, et al. Scaling Large Language Model-Based Multi-Agent Collaboration. In International Conference on Learning Representations.

Evaluation

Topological Perspective Analysis

- **No one-fits-all topology**
- **Density**
  - Positively correlated with performance
- **Shape**
  - **High clustering coefficient** tend to exhibit superior performance —— "small-world collaboration phenomenon"
  - **Wider** topologies generally outperform **deeper** topologies
  - Performance of each topology varies across task domains

Method	Paradigm	MMLU	HumanEval	SRDD	CommonGen	Quality
CoT		0.3544†	<u>0.6098</u>†	0.7222†	0.6165†	0.5757†
AutoGPT		0.4485†	0.4809†	0.7353†	0.5972	0.5655†
GPTSwarm		0.2368†	0.4969†	0.7096†	0.6222†	0.5163†
AgentVerse		0.2977†	**0.7256**†	0.7587†	0.5399†	0.5805
MacNet-Chain		0.6632	0.3720	**0.8056**	0.5903	0.6078
MacNet-Star		0.4456†	0.5549†	0.7679†	<u>0.7382</u>†	0.6267
MacNet-Tree		0.3421†	0.4878†	0.8044	**0.7718**†	0.6015
MacNet-Mesh		<u>0.6825</u>	0.5122†	0.7792†	0.5525†	<u>0.6316</u>†
MacNet-Layer		0.2780†	0.4939†	0.7623†	0.7176†	0.5629†
MacNet-Random		**0.6877**	0.5244†	<u>0.8054</u>	0.5912	**0.6522**†

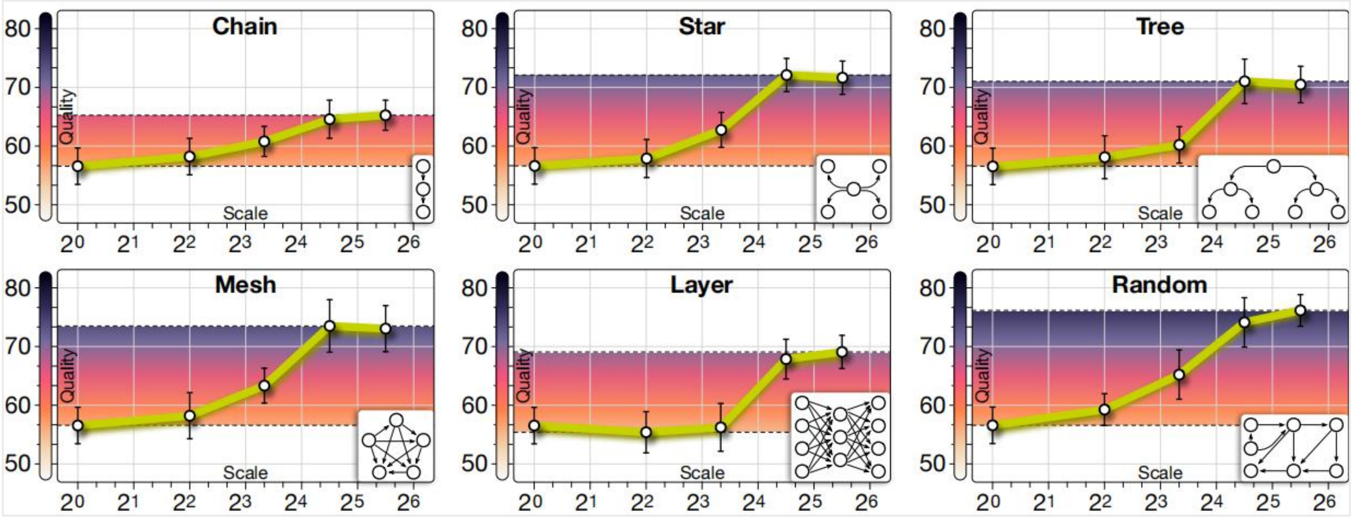


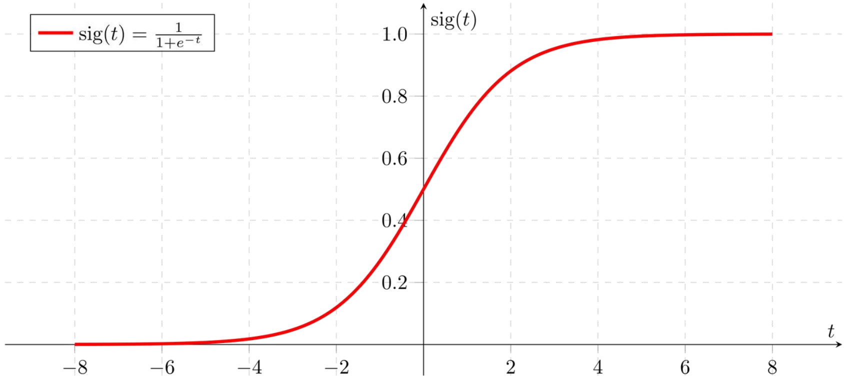
Qian C, Xie Z, et al. Scaling Large Language Model-Based Multi-Agent Collaboration. In International Conference on Learning Representations.

Evaluation

Collaborative Scaling Law

- **Trend Perspective**
  - Performance & number of nodes: resembles a **sigmoid-variant** function
  - Necessitates appropriate collboration strategies
- **Timing Perspective**
  - **Much smaller** scales to manifest emergence, compared to **neural scaling law**
  - Agent collaboration may serve as a **"shortcut"** to enhance intelligence levels



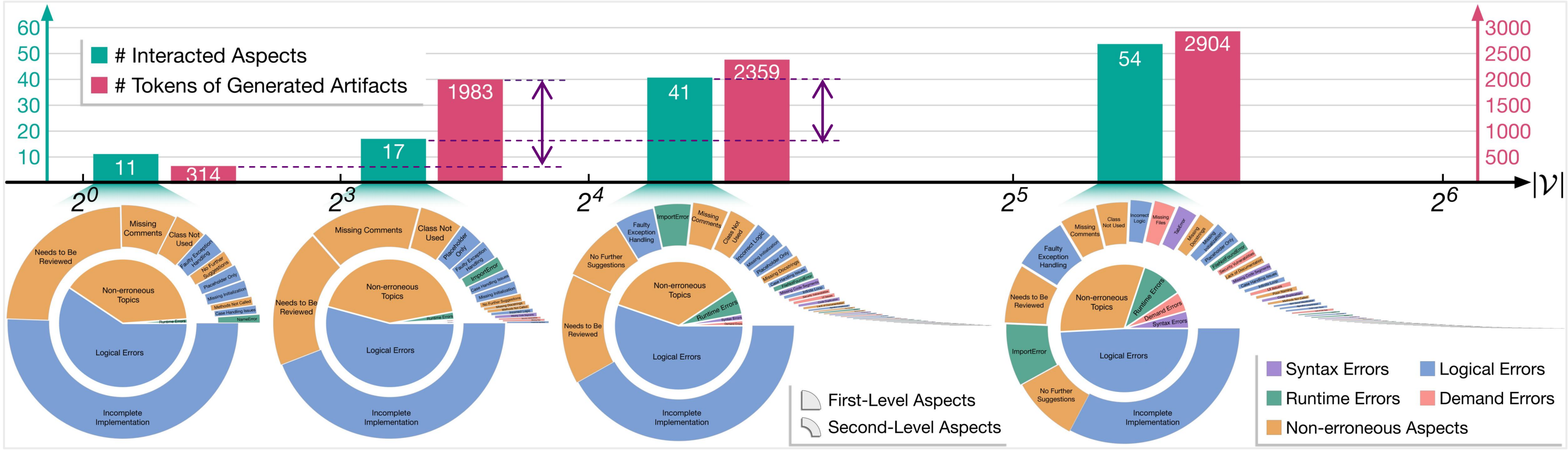


Qian C, Xie Z, et al. Scaling Large Language Model-Based Multi-Agent Collaboration. In International Conference on Learning Representations.

Evaluation

Collaborative Emergence

- **Emergent Behaviors:**
- **# Interacted Aspects** and **# Tokens of Generated Artifacts** exhibits rapid growth when the node scale reaches **2^3** or **2^4**, which aligns with the collaborative scaling law observed in performance
- **Token Distribution:**
- The token distribution from underlying models typically follows a **long-tail** pattern, necessitating larger-scale sampling to likely capture these tail tokens -> **Emergence**



Qian C, Xie Z, et al. Scaling Large Language Model-Based Multi-Agent Collaboration. In International Conference on Learning Representations.

清華大學
Tsinghua University

THANKS!

Qian C, Xie Z, et al.

Tsinghua University & Peng Cheng Laboratory

Contact: qianc62@gmail.com xzh18610221075@gmail.com